Report from the MMM 2018 Special Session on Multimedia Analytics: Perspectives, Techniques and Applications (MAPTA 2018)

Report by Björn Þór Jónsson, Cathal Gurrin, Stevan Rudinac, and Laurent Amsaleg.

Björn Þór Jónsson (http://www.itu.dk/people/bjth/) is an assistant professor in the Computer Science Department at the IT University of Copenhagen. His research focuses on interactivity and scalability of multimedia analytics and multimedia retrieval applications.

Cathal Gurrin (http://www.computing.dcu.ie/~cgurrin/) is an associate professor at the School of Computing, at Dublin City University, Ireland and a principal co-investigator at the Insight Centre for Data Analytics. His research focuses on multimedia information retrieval and personal data analytics, with special emphasis on lifelogging applications.

Stevan Rudinac (http://www.uva.nl/profiel/r/u/s.rudinac/s.rudinac.html) is an assistant professor at the Amsterdam Business School, University of Amsterdam. His research focus is on multimedia analytics applications in urban computing, business and security.

Laurent Amsaleg (http://people.rennes.inria.fr/Laurent.Amsaleg/) is a CNRS researcher with the IRISA laboratory in Rennes, France. His research focuses on large-scale high-dimensional indexing, multimedia analytics, and adversarial machine learning.

Photo: http://www.itu.dk/people/bjth/MAPTA/Media/IMG_0371.jpg
Caption: Werner Bailer replies to a question during the panel discussions at MAPTA 2018. From the left: Björn Þór Jónsson (moderator), Stevan Rudinac, Werner Bailer, Masoud Mazloom, Feiyan Hu, and Cathal Gurrin.

Photo: http://www.itu.dk/people/bjth/MAPTA/Media/IMG_0372.jpg
Caption: The special session was well attended and attendees engaged in a lively discussion with the panel.

This report summarizes the presentations and discussions of the special session titled "Multimedia Analytics: Perspectives, Techniques and Applications" (http://www.itu.dk/people/bjth/MAPTA/) held during the 24th International Conference on

MultiMedia Modeling (MMM 2018), in Bangkok, Thailand on February 6, 2018. The special session consisted of five brief paper presentations, followed by a panel discussion with questions from the audience moderated by Björn Þór Jónsson. The goal of this report is to record the conclusions of the special session, in the hope that it may serve members of our community who are interested in Multimedia Analytics.

Overall, the discussions indicated that there are many potential applications of multimedia analytics, that we must work more closely with users on solving their application needs, that we should also consider working with experts from other fields to understand user-centered evaluation methodologies and application requirements, and that the multimedia community needs to consider how to evaluate and reward work of high quality in this field, in order to give it a more equal footing in its publication venues.

Presentations

First, Stevan Rudinac presented “Rethinking Summarization and Storytelling for Modern Social Multimedia” [1]. This position paper resulted from a working group discussion at the Schloss Dagstuhl seminar “User-Generated Content in Social Media” in July 2017. The working group argues that traditional summarization initiatives have been too narrow in scope for today’s user-generated content in different modalities, formats and languages. In the paper they therefore proposed a framework for flexible multimedia storytelling. At the end of his presentation, Stevan discussed new research avenues in the field of storytelling.

Second, Werner Bailer presented his vision paper “On the Traceability of Results from Deep Learning-based Cloud Services” [2]. Werner explained how the use of deep neural networks for media content analysis raises issues of traceability, reproducibility, and ability to interpret results. The issues are caused by the dependency on training data sets and their possible bias, the change of training data sets over time, and the lack of transparent and interoperable representations of models. Werner analysed these problems in detail and provided some compelling examples, and then proposed six recommendations to address these issues.

Third, Masoud Mazloom presented “Category Specific Post Popularity Prediction” [3]. As social media platforms have become ubiquitous in recent years, understanding what makes postings to such platforms popular has become a field of study. Masoud presented an approach to popularity prediction which takes into account the category of the posting (action, scene, people, or animal) and, by focusing on different attributes for different categories, manages to improve prediction quality over previous approaches.

Fourth, Feiyan Hu presented “Image Aesthetics and Content in Selecting Memorable Keyframes from Lifelogs” [4]. Events captured using wearable cameras can be represented as a visual storyboard, a collection of chronologically ordered images which summarise the day’s happenings. How each event is represented is an important issue, and Feiyan presented an

approach using image aesthetics, in combination with content analysis and temporal offsets, and showed several examples from a publicly-available lifelog dataset.

Finally, Cathal Gurrin presented “Approaches for Event Segmentation of Visual Lifelog Data” [5]. A personal visual lifelog consists of streams of multimodal data with thousands of camera images and other sensor readings per day, which can be used to augment human memory. Cathal presented two novel approaches to event segmentation, based on occurence of visual concepts and image categorization. Results from a user experiment with ten users showed that their approaches performed better than the state of the art. Cathal then presented a set of suggestions for next steps for the research community.

Discussions

Following the presentations, Björn Þór Jónsson moderated a panel discussion. He started with a definition of multimedia analytics, drawn from the discussion of a corresponding special session at MMM 2016 (http://records.mlab.no/2016/02/08/report-from-the-mmm-special-session-perspectives-on-multimedia-analytics/). Björn then asked the panelists how their work relates to multimedia analytics, how it can benefit from multimedia analytics and how it can contribute to the research field? The following discussion, involving all speakers, pointed out three converging trends: (a) modern data analysis techniques (DNN, etc.) give the feeling that more accurate and elaborated information can be extracted from multimedia/multimodal datasets, giving rise to sophisticated applications; (b) most such extraction methods are black boxes, with parameters that are complicated to set; and (c) the users, who are looking for new ways to extract meaning out of multimedia data are not necessarily media-savvy. Together, these trends indicate that the community should work on facilitating accountability and traceability of the processes analysing multimedia material, to help non-expert users better understand the capacity and limitations of the tools. There was agreement that today’s multimedia analytics applications are coming from the people, such as city administrators, business people, lifeloggers, and more, who want to understand and analyse a variety of data.

Next it was the audience’s turn. Tat-Seng Chua described a potential future where, one day, people may own their own datasets, rather than the large companies. In this setting, analytics tasks must be able to gather data from individuals, with appropriate privacy concerns, to get some results. He asked what the panelists see as challenges in this kind of future. The panelists agreed that this is a potential future, and that parts of that future are already happening with sensors everywhere and the associated privacy challenges, concerns, and opportunities. Cathal proposed that distributed access to personal data is likely to become normal, where access to data happens through individuals rather than companies. He noted, however, that in this context the new General Data Protection Regulation (GDPR) will be a big issue, as often researchers wish to study applications and algorithms to do things that may be illegal under GDPR.

Björn then asked how researchers should evaluate success in multimedia analytics, when the goal is insight and knowledge? As educators, he explained, we are told that we cannot evaluate knowledge, as it is lodged firmly inside students' heads. Instead, all we can measure is their ability to solve tasks, describe concepts, apply algorithms, and so on. And if measuring knowledge is difficult, insight may be even trickier. So, how should we proceed?

The panelists agreed that multimedia analytics researchers need to get closer to the users, or as Stevan put it: the place for multimedia analytics is at the user side. We have new metrics, as discussed in some of the presentations, including memorability and aesthetic. Masoud suggested that turning to crowdsourcing as a means to obtain user feedback is one way forward. Feiyan suggested turning to other disciplines, such as psychology, which may already have some answers to this question. It was noted that the morning's keynote ("Perception of Visual Sentiment: From Experimental Psychology to Computational Modeling" by Mohan Kankanhalli) had indeed adopted methodologies from psychology.
A comment was made, however, that the key research venues for the multimedia community might dismiss such work as applications work rather than research work. But if our end-result should be solving societal problems, we may need to rethink evaluation of research work.

Following the last comment, Björn posed a question to the audience in general, and to the attending MMM steering committee members in particular: what can MMM do to support this transition? Are we putting too much emphasis on algorithms and (sometimes incremental) benchmark improvements? Tat-Seng Chua responded that working with users is very tedious work, and not all researchers know how to do that well. He proposed that those who are adept at user-based evaluations could perhaps create and open up relevant benchmarks and datasets, so that the multimedia community at large can contribute. He cautioned against abandoning algorithms completely, but advocated including interesting and relevant problems. It was noted that TRECvid this year has two tasks that are relevant to multimedia analytics: storytelling and video summarisation. David A. Shamma then proposed that as we go from numerical measures, such as recall, f-scores, and response time, and into research domains were these legacy metrics do not apply, we should consider "interdisciplinary" work with experts in the field of human-computer interaction (HCI). Cathal agreed that this does not happen nearly enough and Stevan added that we should also work more often with other disciplines, for example political scientists. His observation was that some problems that we spend a lot of time on are not important to them, while sometimes their important issues can be readily solved. There was consensus, however, that some of the concerns of other disciplines translate to really interesting and relevant research problems, and those are the problems we need to identify and work on. In the end, it was left with the MMM steering committee to consider how to evolve the review and evaluation process to invite more user-centric and application-based contributions into our research areas.

Acknowledgments

The session was organized by the authors of the report. The panel format of the special session made the discussion much more lively and interactive than that of a traditional technical session. We would like to thank the presenters and their co-authors for their excellent contributions, as well as the members of the audience who contributed greatly to the success of the session.

References

The following papers were presented at the special session and published in the proceedings of MMM 2018:

[1] Stevan Rudinac, Tat-Seng Chua, Nicolas Diaz-Ferreyra, Gerald Friedland, Tatjana Gornostaja, Benoit Huet, Rianne Kaptein, Krister Lindén, Marie-Francine Moens, Jaakko Peltonen, Miriam Redi, Markus Schedl, David A. Shamma, Alan Smeaton and Lexing Xie: *Rethinking Summarization and Storytelling for Modern Social Multimedia*.

[2] Werner Bailer: *On the Traceability of Results from Deep Learning-based Cloud Services*.

[3] Masoud Mazloom, Iliana Pappi and Marcel Worring: *Category Specific Post Popularity Prediction*.

[4] Feiyan Hu and Alan Smeaton: *Image Aesthetics and Content in Selecting Memorable Keyframes from Lifelogs*.

[5] Rashmi Gupta and Cathal Gurrin: *Approaches for Event Segmentation of Visual Lifelog Data.*