Cees Snoek
University of Amsterdam

The Future of Multimedia Analysis and Mining: Visions from the Shonan Meeting

Nozha Boujemaa
INRIA, France

Alexander G. Hauptmann
Carnegie Mellon University

Shin'ichi Satoh
National Institute of Informatics, Japan

The National Institute of Informatics (NII) Shonan Meeting, following the style of the well-known Dagstuhl seminars, gathered world-class researchers at a venue in Shonan, a suburb of Tokyo, for focused and in-depth discussion on topics related to informatics. This meeting offered senior researchers an ideal occasion for informal scientific gatherings.[1] The ACM International Multimedia Conference was held in Nara, Japan, in 2012, where researchers gathered from all over the world. Thus, we decided to organize our Shonan Meeting right after ACM Multimedia.

Thanks to this collocation and succession, we were able to attract many top-class multimedia researchers to the meeting. The meeting's topic was "The Future of Multimedia Analysis and Mining." Several such previous meetings set a successful precedence, including the ACM SIGMM retreat in 2003[2] and the Dagstuhl seminar "Multimedia Research: Where Do We Need to Go Tomorrow?" in 2005.[3] Although these previous meetings were comprehensive and well organized, we designed the gatherings to be more informal. However, we are very much confident that the discussions were both exciting and timely. This article summarizes the outcome of the four days of intensive discussion.

Shonan Meeting Participation

The meeting began with seven focused talks designed to share the ideas of cutting-edge multimedia research and to stimulate a discussion about the future multimedia research. The remaining time was mostly dedicated to free discussion. (The full program is available at www.nii.ac.jp/shonan/seminar002.) The meeting's participants were as follows:

- Benoit Huet, Eurecom
- Ichiro Ide, Nagoya University
- Kevin Jing, Google
- Kunio Kashino, NTT
- Akisato Kimura, NTT
- Rainer Lienhart, University of Augsburg
- Tao Mei, MSRA
- Frank Nack, University of Amsterdam
- Yuichi Nakamura, Kyoto University
- Chong-Wah Ngo, City University of Hong Kong
- Vincent Oria, New Jersey Institute of Technology
- Masanori Sano, NHK Science and Technology Research Labs
- Alan Smeaton, Dublin City University
- Hari Sundaram, Arizona State University
- Marcel Worring, University of Amsterdam
- Xiaomeng Wu, NII
- Cai-Zhi Zhu, NII

Editor's Note

After ACM Multimedia 2012 in Nara, Japan, several senior multimedia researchers gathered in the Shonan Village Center to discuss the future of our discipline. This article summarizes their main findings.

1070-986X/13/$31.00 © 2013 IEEE

Published by the IEEE Computer Society

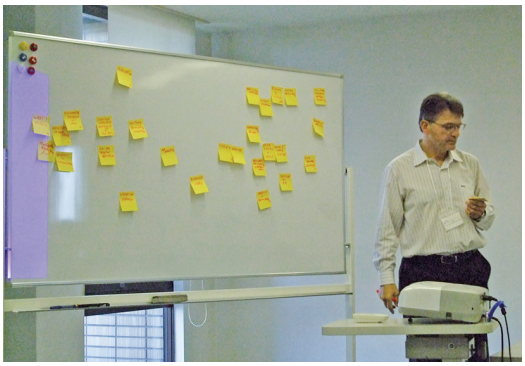
Figure 1 shows a group photo of all the participants.

The discussion sessions consisted of both plenary and breakout sessions. As the program was informal, in the first plenary session, we discussed what to discuss during the meeting. The agreed on topics were

- fundamental multimedia science,
- socially motivated multimedia applications, and
- education.

After the first plenary session, we divided into two to three breakout session groups, and each group discussed one of these topics (see Figure 2). After the breakout sessions, we reported and summarized the results in plenary sessions. Finally, we concluded the meeting with a free discussion on possible research collaborations and sharable research resources among participants.

Figure 1. Group photo of participants at the NII Shonan Meeting.



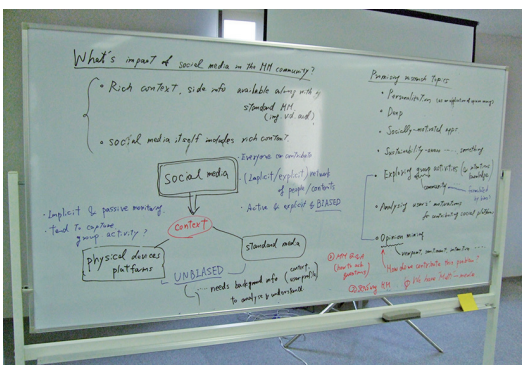


Figure 2. NII Shonan Meeting brainstorming and breakout sessions.

Fundamental Multimedia Science

By its nature, multimedia is multidisciplinary, relating to many scientific fields including signal processing, computer vision, database, network, middleware, human-computer interaction, social science, and the humanities. We first discussed fundamental research topics in multimedia as a scientific field.

We began with this question: What is missing from current multimedia research? We then discussed and identified six areas in which further research efforts are necessary:

- *Data representation*. Although there are already numerous feature-extraction approaches leading to many different attributes, current research results clearly show their limitations. Additionally, there is a need to handle heterogeneous data effectively and efficiently.
- *Multimodal fusion*. The intrinsic heterogeneity of multimedia data requires specific research to be undertaken combining multiple data-sensing sources (such as GPS and depth), high dimensionality, and temporality.
- *Machine learning*. Current approaches suffer from a number of limitations and flaws.

```
[Define] (vague/concrete) application around media
Do for several years
{
   If not (ill posed)
         use the appropriate tools
   else
         search for a (possibly applicable) toolset
}
   Let other communities pick up the problem to provide more
         fundamental/accurate solutions
   Move on, i.e. go back to the first step
```

Figure 3. Algorithmic formulation of the problems facing the field of multimedia. As an empirical science, multimedia is based on a fundamental strategy of exploring novel ill-posed problems around media.

Research should take into account the need for learning from a few examples and/or incomplete data as well as dealing with multidimensionality and large-scale issues.

- *Context*. Context is an important information source that can be observed from many sensors (such as location and time) but also from social networks (community).

- *Events as structuring elements*. It is convenient and natural for humans to structure their memory in terms of events (such as a wedding, vacation, or conference). Because events are generally described in terms of the four Ws (who, what, where, when), they constitute the basis for representing and organizing life observations (whether human centric such as life logging or data centric such as environmental sensing).

- *Human factors*. Human factors and abilities should be addressed more thoroughly and be included and considered throughout various research steps, whether in the form of user interaction or in addressing the user's intent and preferences (personalization).

While discussing these points, we identified a number of research topics and directions for which progress still needs to be made. Example areas include browsing and navigating through media/multimedia, summarizing media, indexing and retrieving media, multimedia mining and analytics, media recommendation, and the long-term research objective of multimedia understanding.

Because of multimedia's multidisciplinary nature, we sometimes regard all the core research problems as belonging to other scientific fields and get lost in the quest of core multimedia research problems. Based on this view, the meeting also had a controversial discussion on the unique structure of multimedia as a science. The field of multimedia differs from other computer science fields (such as operating systems, algorithms, compilers, and networking) in that it consists less of fundamental laws and findings in a strict mathematical and algorithmic sense. Instead, as an empirical science, it is based on a fundamental strategy of exploring novel ill-posed problems around media. Figure 3 formulates this problem algorithmically.

Thus, multimedia research influences the field of computer science by exploring what questions and problems should be asked. Many key components of multimedia research are expressed by this algorithm. For example, most multimedia research is about the integration of existing and newly developed methods to achieve the desired result. Integration is a challenging and interesting engineering task.

In addition, most multimedia research explores the limits of the current toolset (including experimentation). We address questions other communities do not yet dare to ask, because these questions are regularly ill-posed, before they are correctly understood and changed to well-defined problems. One example of this is content-based image retrieval. While the computer vision community was still exclusively focusing on images taken in controlled environments, the multimedia community has been working on image retrieval of images in the wild (the Internet).

Multimedia research mostly (not always) extends tools in an ad hoc manner. However, they rely on other communities to pick up the questions and quest once the problem definition and first ad hoc solutions have reached certain thoroughness. In that sense, multimedia research is important for the other related computer science research fields (such as computer vision and audio processing) because we experiment with novel problems and their definition, which other related fields can take as inspiration. This exploration of ad hoc solutions for novel problems produces a butterfly effect.

Lastly, although there are new things to explore within the current applications, we must impose ''time-limited persistence.'' That is, we should repeat the steps in Figure 3 for a limited time before moving on.

In summary, our ''ironic'' conclusion to this discussion was this: The fundamental quest in our field is to explore around media.

Socially Motivated Multimedia Applications

No one can be independent from society. One way to observe society is the use of social media, which is one of the hottest research topics in multimedia right now.

Social media is a relatively new type of media for the multimedia community and has a great potential to capture human activities and standard media contents. Social media platforms contain rich contexts, user networks, and geolocations. Also, they are often associated with standard media content through social media platforms such as simple notification services (Facebook and Twitter), media uploading platforms (Flickr, Instagram, and YouTube) and social curation services (Storify, Scoop.it, and Togetter).

Overview of the Current Multimedia

The observations we are dealing with can be roughly classified into three types.

- *standard multimedia content*, such as images, video, speech, and music;
- *context*, often obtained from various physical devices and platforms, including geolocations (through GPS, Wi-Fi, or RFID tags), accelerometers, and browsing logs; and
- *social media*, user-contributed data such as images, videos, and text.

A significant characteristic of social media compared to other content is that it often contains user tags, comments, and ''likes.'' This implies that users actively publish their social media content, as opposed to the other two types of observations. Therefore, social media content is highly biased by nature.

Another point we have to note is the existence of networks of people and content, which might be implicitly or explicitly but naturally constructed. These networks organize groups of people or content, and the groups reflect interests, opinions, and real-world relationships.

Passively Versus Actively Contributed Data

Passively contributed data are information or signals that are captured regardless of the users' intentions. Such data may reveal personal and/or group activities and opinions. This implies that, especially when dealing with this type of data, we have to consider the trade-off between benefits (capturing personal and group activities and opinions) and drawbacks (privacy issues).

On the other hand, actively contributed data are voluntarily provided by users, usually with an intention. However, even within actively contributed contents, especially in case of image and video contents, information may exist that a user did not intend to provide.

Explicitly Versus Implicitly Contributed Data

In the case of actively contributed data, some information are explicitly represented (for example, the content itself such as an image or video clip). However, implicitly represented information, such as the time, the frequency of contributions, and small or blurred objects in the background, can also be valuable information to analyze a user or group's activities and opinions. In many cases, such information is not intended to be provided by the user, so privacy issues may arise depending on how they are computed and used.

Promising research topics in this area include the following.

- exploring group activities (intentions and knowledge),
- personalization,
- sustainability through social awareness,
- analyzing users' motivations for contributing social platforms,
- opinion mining (viewpoints, sentiments, intentions); and
- handling bias in media.

Multimedia Community Contributions

All the research topics we have introduced thus far are interesting and challenging.

Emergencies and natural disasters are ideal scenarios that might benefit from multimedia technologies.

However, there seems to be many important issues originating from data mining, social media analysis, and machine learning rather than traditional multimedia processing. Therefore, the multimedia community should keep in mind what its original contribution can be with respect to such research topics.

The Shonan Meeting participates highlighted some promising directions for the multimedia community when tackling those research topics:

- The multimedia community possesses and exploits ''multi''-media, including conventional multimedia content (such as text, images, audio, and video). Standard multimedia content has a better chance of containing implicit and passively provided information than the content other communities are handling and thus privacy concerns tend to exist.
- Sensing multimedia might include active sensing with additional sensors or platforms.
- How we ask questions in multimedia Q&A scenario is a significant issue. Still, filtering potential answers available on the Web is difficult because they are biased in nature.

During the meeting, participants had strong opinions about how multimedia technologies should contribute to our society. For instance, Japan was deeply impacted by the earthquake and tsunami in March 2011. During this disastrous event, proper use of technologies, especially multimedia technologies, might have helped human lives.

As yet another aspect of the relation between society and multimedia, we discussed possible or desirable social impact that multimedia might have, society's expectations of the multimedia research community, and technologies that multimedia can and/or is expected to contribute.

Social Impact

Several breakout sessions explored the social impact of multimedia technologies. The topic discussions covered a wide range of issues. For example, can we reveal the bias in news media? That is, participants discussed whether it was possible to reveal the subtle coordination that may exist between different players, including bloggers, corporations, news organizations, and politicians to promote or suppress news. Ensuring transparency in news should be a critical goal because this will promote civic engagement with the public.

Another topic discussed was corruption, a major concern in developing countries, including India, China, and many other parts of the world. For instance, how might we reveal major donors to politicians? On a smaller scale, some corruption deals with petty issues such as paying policemen to avoid fines, which may be arbitrary. How do we encourage the public to report instances of corruption? The I Paid a Bribe website (www.ipaidabribe.com) is a wonderful example of citizen reports of corruption in India. Corruption is a cultural issue, and the people who pay bribes are in some sense also complicit. We need to provide alternatives to people do not want to pay a bribe but find themselves unable to find alternatives.

Emergencies and natural disasters are ideal scenarios that might benefit from multimedia technologies. In the Great East Japan Earthquake, much of the coastal infrastructure was wiped out. Maps as well the knowledge of services were immediately rendered obsolete. To ensure relevant information, we need to crowdsource this data, such as which roads are accessible and the locations of doctors and nurses. How else can we better deal with emergency situations or rapidly developing events, such as riots or violent mobs?

Here's a list of additional topics explored:

- *Vaccinations*. If enough people take the recommended vaccinations for influenza, it is highly probable we can prevent large-scale epidemics. How do we encourage people to get vaccinated?

- *Voting*. There is significant voter apathy at the national and state levels. For example, presidential voting rates in the US rarely exceed 60 percent.

- *Disease/health*. How we encourage people to lead healthier lifestyles? For example, can we persuade teens not to consume fast food? How do we take care of an aging population? Can we encourage communities to help the elderly? How do we create environments, including applications and networks of people, that enable people to better cope with depression (or loneliness)?

- *Sustainability*. How can we encourage more sustainable behaviors, such as recycling, reducing energy consumption, and participating in energy generation?

- *Security*. How do we increase the sense of security for the public? Make it safe for young children to travel and safer for women?

- *Crime prevention*. How can we reduce the spread of hate speech and child pornography?

Needs

Based on the previously discussed topics, the meeting participants brainstormed some of the related technological needs in the field. For example, we need application-appropriate time sensitivity. Some applications require near-real-time support (such as emergencies), while others including vaccinations and voting require coarse temporal resolution. We also need social awareness. That is, we need to create technologies that connect people. We need to identify relevant groups within large networks, and within these groups, we need to identify exemplars, including people and processes.

We need high-quality information about the activities that affect outcomes. We need to identify people and information repositories that engender trust within the network.

Information repositories need to be reactive. That is, they need to be updated with high-quality information in near real time, especially in applications that deal with disaster relief and crime.

In the broadest sense, sensing has to be multimodal. With input from the public, government agencies, and nonprofits, sources can be diverse. They can include images, SMS text, audio, videos, blog entries, micro-blogs including Facebook, news aggregators including Digg, and social blogs such as Tumblr.

The focus on preserving privacy is paramount in socially aware applications. In all cases, we should err on the side of caution and ensure that no unauthorized disclosure of information takes place. We need to be concerned about the consequences of information aggregation; mining activity patterns of citizens can lead to discoveries that violate privacy.

Socially aware applications need to go beyond information presentation and reflection but still be cognizant of constraints on time and other resources that prevent people from acting for the common good. We need to empower users to act and support the application's goals. This includes providing resources, peer support, monetary and nonmonetary incentives, and information that supports people's ability to act.

Socially motivated applications require working with all the different stakeholders in the outcome. We should take advantage of domain experts as well as participants with significant experience. These people should be identified so as to provide feedback to not only application developers but to act as resources to participants in the socially motivated multimedia applications.

Lastly, we need to ensure that information infrastructures are robust against attacks. For example, we need to be able to rapidly filter out spam and prevent attacks. Unlike the traditional denial of service attacks, in socially aware applications, denial of service may take the form of deliberate information overload that prevents effective decision making.

Technologies

The technologies that might help meet these needs were also explored. For instance, scalable infrastructure and information processing are necessary. Such technologies can help reactively connect people and allow a high level of interaction and personalization.

Methods that allow for faceted data mining through a subset of multiple dimensions are also important. In this regard, scalable data mining methods are needed. Visual analytics can provide the opportunity for such faceted content exploration and facilitate user uptake of complex information spaces.

Multimedia research can ultimately produce practical improvements for kindergarten through university instruction.

In data mining approaches, we also need to focus on outlier detection. Outliers can be rare but informative (offering dissenting opinions). The diversity of opinions should be well depicted and represented to take into account the majority as well as minorities.

Technology could help improve trust between individuals (encouraging the formation of small social groups) but also can help aggregate coherent or dissonant diverse information sources to improve the truthfulness.

Lastly, technology should help reveal information sources in a transparent way for the information consumer. For example, we need to make connections between media and corporations (such as between bloggers and corporations) and between corporations and politicians (who is sponsoring whom). This can prevent information bias that could impact and influence opinions.

Education

Multimedia research can ultimately produce practical improvements for kindergarten through university instruction. The foundations here have the potential to impact lifelong learning for whole generations. Ever since multimedia started to coalesce as a field, there has been the vision that it would be useful for education. However, the reality is that progress had been glacial compared to other advances in the field.

More recently, game-changing developments have come through the emergence of massive, open, online courses (MOOC), where universities and educational organizations are providing course materials online, usually in video and/or slide form. Such resources can also be combined with other related text materials.

The result has been that courses can now reach hundreds of thousands of students who could not previously afford the costs associated with higher education or who lived in remote areas without access to quality educational services.

This new opportunity has provided testbeds to conduct large-scale research evaluations to improve learning through better teaching methods, discovering and applying principles of human learning, and validating research through empirical tests with statistical validity. This is a problem ideally suited for multimedia research because it will require the exploration of all types of visual, audio, graphic, language, and social information to understand what is likely to result in effective teaching approaches.

Multimedia-Based Education: Research Questions

Of course, all this cannot be accomplished merely by the multimedia community. It will require a multidisciplinary approach, bringing together educators with a deep understanding of pedagogy, evaluation experts, psychologists, graphic designers, and of course, multimedia researchers to deal with content acquisition, analysis, structuring, search, scalability, and reusability.

With these new opportunities, the effectiveness of new approaches, which might have originated from hypotheses generated by educators, can now be measured at scale.

Additional advances that we believe are achievable include improvements in the time necessary to create courses, lesson plans, lectures, or units. These improvements could arise from better methodology and more advanced tools for content browsing, generation, and rendering. This goal implies that we work toward building systems for multimedia integration to support multiple subjects. This requires progress on different levels:

- *Lectures, slides, homework, and exercises*. At the structural level, it is a different task to design a course based on state-of-the art research insights than to create a specific lecture on a given topic from a plethora of available materials such as slides, videos, and research papers. It is even unclear that lectures are the best mechanism in many circumstances. Homework and exercises must be created to fit the lecture content, with examples widely available on the Web, although understanding their quality is difficult without collaborative ranking tools.

- *Content creation*. Because it is unlikely that all content is already available, we also want to enable easy creation of multimedia content. The content needs to fit the topic and the educational intent. For example, if we try to merely communicate conceptual knowledge, a simple slide might be sufficient. On the other hand, if the goal is critical reflection, it would be appropriate to enable a social tool that allows discussion. To teach particular skills, an interactive animation might be best suited to demonstrate action and reaction. Authoring tools are necessary to support all these different educational purposes.

- *Measurement tools*. To validate that our educational outcomes are actually improving, we need tools that can test if the approaches are effective. This evaluation has two purposes: It can answer the fundamental question, ''Are students learning?'' and, at the same time, discover whether specific methods used to present the material are more effective than others. In addition, our evaluations must deal with large numbers of students, where manual grading and commenting may no longer be feasible.

Benefits

The potential innovations would benefit both instructors and students. They also have the potential to vastly advance education science through efficient metrics-based evaluation of the effectiveness of learning methods for skills (how to do something), knowledge (what is this?), and critical as well as creative thinking.

The multimedia field should particularly relish the challenge in the intelligent creation and integration of diverse educational materials—slides, lectures recorded in low- or high-quality video, Wikipedia, textbooks, research papers, exercise worksheets, homework templates, and even standalone instructional videos found on the Internet.

A developed system could test the effectiveness of approaches for asynchronous learning. For example, students might be in different times zones and learning at their own pace using telepresence where the (perhaps virtual) instructor is directly interacting with the student. Direct physical live instruction techniques augmented by peer-group mentoring, teaching assistants, and (synthetic) question answering might also be used.

The multimedia field should particularly relish the challenge in the intelligent creation and integration of diverse educational materials.

We believe there are no shortcuts to learning. Education will certainly remain a difficult process, requiring work and dedication by the student, even with better tools and materials. A suggested approach would be to first develop and evaluate the effectiveness of new and existing methods on statistically representative student populations. Once the fundamentals have been established, there are opportunities to move on to personalized learning.

We recognize that no one method is optimal for everyone, and many students have different learning styles. A new paradigm could be enabled that allows us to specialize approaches for different student subpopulations based on their knowledge, background, and skills. Individualized consideration can also be given to students with special needs. At times, resources may be constrained, such as when a student only has cell phone access from a remote location, so high-definition video may not be appropriate. Finally, we can also take into account that the student or educator's time is constrained. Thus, a different approach may be required if the student only has two hours to devote to the material as opposed to a whole day.

Encouragement of Multimedia-Based Education

Clearly, this is an exciting area for multimedia researchers with the potential for significantly impacting education throughout the world. But it is also worth pointing out that this research is not quick or easy. It will require years of forging interdisciplinary communities of experts, engagement with student populations,

(a)

(b)

Figure 4. NII Shonan Meeting (a) Kamakura excursion and (b) Mt. Fuji view from the Shonan Village.

and iterating design options. In addition, the progress is likely to be slow at times, resulting in a potentially low rate of publications. Nevertheless, we feel this is an opportunity at an inflection point that should not be ignored.

One concrete step that the multimedia community can take to encourage creative thinking about improving educational presentation is to support a competition with a prize at a conference such as ACM Multimedia. One thought is to encourage different groups to create alternative presentations using the previous year's best paper. Then, some form of judging by the conference participants would determine the approach deemed most effective to explain the paper's content.

Conclusion

This article summarizes the outcome of the 2012 Shonan Meeting ''Future of Multimedia Analysis and Mining.'' The meeting was really interesting, and the participants had a fun time with an Kamakura excursion and fine dinners, in addition to in-depth discussions on ready-to-go hot research topics (see Figure 4). We have enjoyed sharing even part of our experiences with readers here. **MM**

References

1. M.Y. Vardi, ''Where Have All the Workshops Gone?'' *Comm. ACM,* vol. 54, no. 1, 2011, p. 5.
2. L.A. Rowe and R. Jain, ''ACM SIGMM Retreat Report on Future Directions in Multimedia Research,'' *ACM Trans. Multimedia Computer Comm.,* vol. 1, no. 1, 2005, pp. 3–13.
3. ''Multimedia Research: Where Do We Need to Go Tomorrow?'' Dagstuhl Seminar, 2005; www.dagstuhl.de/de/programm/kalender/semhp/?semnr=05091.

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. Contact her at nozha.boujemaa@inria.fr.

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. Contact him at alex+@cs.cmu.edu.

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. Contact him at satoh@nii.ac.jp.