Academia Meets Industry at the Multimedia Grand Challenge

Cees G.M. Snoek
University of Amsterdam

Malcolm Slaney
Yahoo! Research

For many years, people have consumed more multimedia content than written information. Yet the tools consumers use to find the content they want to see are often based on text. Hence, despite years of vibrant research, the multimedia field has been criticized for a lack of real-world applications.[1] A recent panel discussion on multimedia search had the revealing title: Multimedia Information Retrieval: What Is It, and Why Isn't Anyone Using It?[2] With the exception of face recognition and optical character recognition, little of the wonderful technology created by the multimedia community has met with commercial success. Search engines look for multimedia content on the basis of the text around the object. Recommendation engines do better by ignoring the content completely and correlating user's rating data instead.

The ACM Multimedia Grand Challenge was designed to bring such commercial needs to the attention of researchers. Seven industrial partners led the Multimedia Grand Challenge by identifying issues they think are important to their business and worth further study.

Key challenges

The industrial partners pointed out that addressing these challenges would open up new business opportunities and create a richer experience for their users. In addition, they said they hoped that describing real-world challenges would allow researchers to focus on projects that have a better chance of success in the marketplace.

Thus, the following 10 challenges were identified in the 2010 edition of the Multimedia Grand Challenge:

- *Automatic theme-identification of photo sets*. Provide a user with a selection of photo-book styles accommodating the user's preferences for the images in the set, and their preferred structure.

- *Genre classification for video*. Automatically classify user-generated videos and their metadata into genres.

- *High-impact visual communication*. Create a high-impact collage (and its description) that uses a set of photos to convey information across cultural boundaries.

- *Interactive searching in personal diaries*. Develop good schema, algorithms, and user interfaces that make multimedia diaries accessible.

- *Multimedia-content adaptation*. Adapt, in real-time, the same multimedia content for different receiving devices, in a way that is perceptually optimal for users.

- *Novel image-understanding*. Develop methods that move beyond simple image classification.

- *Photo location and orientation*. Derive the exact camera poses (location and orientation) for photos that are lacking location annotation.

- *Robust video segmentation*. Develop algorithms to automatically create narrative themes for a given video and present the content to end users in a search-engine experience.

Editor's Note

This column is about last year's ACM Multimedia Grand Challenge in Florence, Italy, an event that endeavors to connect (academic) researchers more effectively with the realities of the business world. The authors describe the 10 challenges and present the three winning applications.

—Frank Nack

Multimedia Grand Challenge 2010 Winners

First prize went to Jana Machajdik, Allan Hanbury, and Julian Stöttinger who are affiliated with Vienna University of Technology and the Information Retrieval Facility, both in Austria. Machajdik presented their solution to the challenge on novel image-understanding methods. The winning idea is a system able to search for images with a requested emotional effect—such as images being cute, fearful, or sad. Figure A shows examples of the work. The system computes image features deemed relevant for emotions on the basis of art theory and psychological studies. With the help of crowd-sourced training data and supervised machine learning the system can then predict emotional states of images.[1] They went home with US$1,500 dollars.

The second prize went to Wei Song, Dian Tjondronegoro, and Ivan Himawan who are all affiliated with Queensland University of Technology, Australia. Tjondronegoro presented their solution to the challenge on multimedia content adaptation. The researchers proposed a scheme consisting of three parts: in the first part, the system performs region-of-interest detection; in the second part, the video codec adaptively uses different strategies to compress the input video into low bit-rate video streams;[2] in the final part, the system transmits the streams to the mobile-conferencing users. This scheme is simple but effective in optimizing mobile users' viewing experience. The researchers went home with US$1,000 dollars.

The third prize went to Julien Law-To, Gregory Grefenstette, Jean-Luc Gauvain, Guillaume Gravier, Lori Lamel, and Julien Despres who are affiliated with Exalead, LIMSI CNRS, and Vecsys Research, all in France. Law-To presented their solution to the robust video-segmentation challenge. The researchers presented Voxalead-News (see http://vox.labs. exalead.com/voxalead), which allows for segment-based retrieval of broadcast news using speech recognition and natural-language processing. An elaborate interface provides additional annotations for each segment, including named entities, query-term timestamps, a timeline showing mentions, and a list of terms mentioned per segment. They went home with US$500 dollars.

References

1. J. Machajdik and A. Hanbury, ''Affective Image Classification Using Features Inspired by Psychology and Art Theory,'' *Proc. ACM Multimedia,* ACM Press, 2010.
2. W. Song et al., ''Impact of Zooming and Enhancing Region of Interests for Optimizing User Experience on Mobile Sports Video,'' *Proc. ACM Multimedia,* ACM Press, 2010.
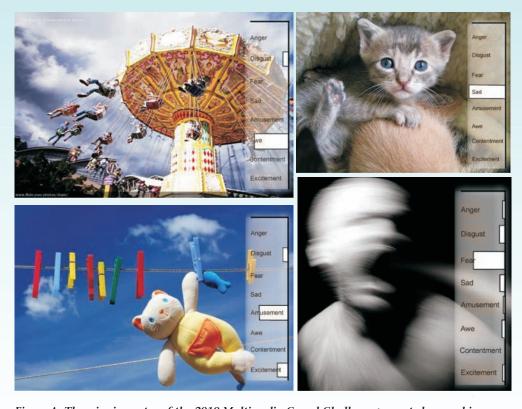
Figure A. The winning entry of the 2010 Multimedia Grand Challenge presented a novel image-understanding method for image search using emotional effects, such as amusement, disgust, or contentment.[1] (Images courtesy of these Flickr users: alphadesigner, Immagina, fofurasfelinas, and Christopher Chan.)

- *Sports-activity analysis in camera networks.* Identify the limits in terms of 2D and 3D data extraction from a low-cost camera network for sports, such as tennis matches.

- *Videoconferencing experience.* Develop new technologies and ideas to surpass the in-person meeting experience.

Consider the challenge raised by HP to make it easier to create high-impact visual communication. Because images serve as a powerful communications vehicle, HP challenged researchers to create a high-impact multimedia summary from a large collection of photos. The ideal solution would be to create a collage of images and words to tell a story. As another example, consider a scenario explored by Nokia to address the question of where a photo is taken and how. Here, the ideal solution would derive the exact camera pose from photos that are lacking any location annotation. The ultimate goal is to add metadata to existing or newly captured photos. Together, the problems defined for the Multimedia Grand Challenge cover the full range of multimedia research, including systems, applications, content analysis, and human-centered aspects.

Product experience

The partners contributed their problems, prize money, and in some cases data sets to the Multimedia Grand Challenge. Most importantly, they contributed judges with product experience, many of whom are directly responsible for multimedia products. The concerns of these judges were different from those of conference judges determining the merits of scientific papers. Will this approach lead to more spam? What does a better precision recall curve mean to a soccer mom? Will users understand the technology? Can the technology do something magical without any user input? Most importantly, does the research do something for which people will be willing to pay money, or to which they will devote their attention?

Papers were selected for the final round of judging in Florence on the basis of whether they

- significantly address one of the industrial challenges;

- depict working, presentable systems or demos; and

- describe why the system presents a novel and interesting solution.

In the end, 18 contributions were accepted that together covered all 10 industry problems.

Multimedia idol

All finalists for the Multimedia Grand Challenge presented their work to the panel of judges in a conference-wide event. Figure 1 shows an overview of their approaches. Each finalist had three minutes to present the idea in a format that mixed an elevator pitch with a show like *American Idol*. An elevator pitch is important because researchers have to distill the essence of their idea for a quick presentation. The term comes from the venture-capital world and refers to what you would say if you were lucky enough to run into a venture capitalist on a short elevator ride. *American Idol* and similar shows are important because the presenter's energy and entertainment value excite the audience—and in the end you want the users to be excited about your work.

After an energetic session led by master of ceremony David Ayman Shamma, the industry partners decided on the three winners. Some participants went home with prize money, but everyone went home with the message that solving real-world industry problems can be as challenging as research.
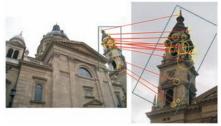
Looking back and ahead

The Multimedia Grand Challenge would not have been possible without the help of many. As the event's organizers, we thank the industry partners for sponsoring and supporting the Multimedia Grand Challenge. We also thank the jury members and researchers for their participation. We are grateful to Alberto Del Bimbo and Marco Bertini for the local arrangements in Florence. And finally, we appreciate the effort that Mor Naaman made to start the Multimedia Grand Challenge in 2009 and share his experiences.

In the end, the Multimedia Grand Challenge offered a meeting place where academia interacts with industry. The biggest value in the Multimedia Grand Challenge was to bring industry expertise to an event at an academic conference, allowing experienced product

managers to give feedback, ask the right questions, and guide the eager researchers toward even better technology for future multimedia products. There will be another Multimedia Grand Challenge at ACM Multimedia 2011 in Scottsdale, Arizona. Hope to meet you there. **MM**

References

1. M. Kankanhalli and Y. Rui, ''Application Potential of Multimedia Information Retrieval,'' *Proceedings IEEE,* vol. 96, no. 4, 2008, pp. 712-720.

2. A. Jaimes et al., ''Multimedia Information Retrieval: What Is It, And Why Isn't Anyone Using It?'' *Proc. ACM SIGMM Int'l Workshop Multimedia Information Retrieval,* ACM Press, 2005, pp. 3-8.

Contact author Cees G.M. Snoek at cgmsnoek@ uva.nl.

Contact editor Frank Nack at F.M.Nack@uva.nl.

cn *Selected CS articles and columns are also available for free at http://ComputingNow. computer.org.*

Figure 1. Snapshot taken from several demos, videos and slide shows as presented at the 2010 Multimedia Grand Challenge. Note the diversity in multimedia research solutions for the 10 practical problems raised by industry. (Images used with permission.)