Multimedia Grand Challenge 2012

Yushi Jing
Google Research

Go Irie
Nippon Telegraph & Telephone

Marcel Worring
University of Amsterdam

The Multimedia Grand Challenge is a recurring event at the ACM Multimedia Conference series. During this event, delegates from various industries define a number of challenges that they consider of interest from both a business and scientific perspective, giving the multimedia research community an opportunity to solve relevant, interesting, and challenging questions in the multimedia industry's two to five year horizon.

This year, three industry challenges were adaptations from challenges contributed in 2011. Another three new challenges, mostly from Japan-based companies or subsidiaries, were also introduced for 2012. The combination of global multinationals and local partnerships provided the event with a diverse set of perspectives and experiences. The proposed challenges touched various aspects of multimedia, including content creation, media interaction in virtual environments, and content analysis for images, videos, and social networks.

For the 2012 Grand Challenge competition, we received 36 quality submissions from seven countries. By pairing with our industry reviewers, we reviewed each submission and selected 17 finalists. Each finalist submitted a two-page short paper, which was published in the ACM Multimedia 2012 proceedings. Figure 1 illustrates the composition of the finalists by country and challenge topic. China (mainland), Taiwan, France, and Singapore contributed multiple finalist teams to the event. Also, each challenge accepted an average of three finalist teams, with Google and HP accepting the most (four) teams, which reflects the initial distribution of papers over the challenges. The overall acceptance rate was 47 percent.

During the two-hour plenary session at the ACM Multimedia Conference in Nara, Japan, the 17 finalists gave 240-second presentations in front of an audience of more than 300 people in quick succession. Each finalist team also had 120 seconds of questioning from the audience and the jury members: Brad Ellis (Google), Christophe Diot (Technicolor), Qian Lin (HP), Zhangqing Qing (Huawei), Simon Clippingdale (HNK), and Takeshi Yoshimura (NTT-docomo). Both quantitative scores and qualitative comments from the jury were used to determine the winners of the Grand Challenge prizes.

Editor's Note

The Fourth Annual Multimedia Grand Challenge was held at the 2012 ACM Multimedia Conference in Nara, Japan. The Grand Challenge event invites researchers from academia to face multimedia challenges inspired by industry. This article describes the 2012 challenges and highlights the winning entries.

The 2012 Challenges

Six industry partners participated in this year's Grand Challenge program: Google Japan, HP, Huawei/3DLife, NTT Docomo, NHK, and Technicolor. The six challenges for 2012 were as follows:

Google: Automatic Generation of Music Videos

Google challenged the multimedia community to come up with methods for automatically generating music videos, either by recommending appropriate music to a user video or recommending suitable video clips to accompany a soundtrack. If done right, such "personalized" music videos can add significant entertainment value, making video sharing a lot more fun for the users.

HP: Understanding the Emotional Impact of Image and Videos

As a powerful medium of communication, multimedia can invoke an emotional response from the viewers. The HP challenge solicits works that investigate the connection between

1070-986X/13/$31.00 © 2013 IEEE

multimedia content and its effect on viewer emotions and then leverage that understanding to improve multimedia presentation and interaction.

Huawei/3DLife

As a continuation from last year, Huawei and 3DLife provided the multimedia community with a sensor-rich dataset of dance instruction sessions. The dataset consists of both music excerpts from dance sessions, inertial sensors (such as an accelerometer or gyroscope) capturing data from multiple locations on the student's body, depth maps of student performance, and a set of metadata including ratings of the student performance and annotation of choreographies. The question posed by the challenge is whether, given such a rich set of data, we can realistically render the student's avatar in the virtual environment in real time, while subject to the modes of interaction available from each capturing mechanism.

NTT Docomo: Event Understanding through Social Media and its Text-Visual Summarization

The NTT Docomo challenge focuses on summarization of social media content (such as Twitter). Specifically, the challenge calls for methods that convert real-time text and image information into topic-specific magazines. In addition to collecting sufficient images and text from social media, it is necessary to tackle problems such as location identification, events summarization, representative image selection, and layout planning. The final output, presented in the form of a ''dynamic'' magazine, was judged by the relevance of the information, quality of summarization, and the quality of the magazine design.

NHK: Where Is Beauty? Video Segment Extraction Based on Aesthetic Quality Assessment

NHK's challenge, titled ''Where Is Beauty?'' was aimed at automatic recognition of beautiful scenes in broadcast programs. The implicit assumption was that some video snippets would be perceived as more beautiful than others by an average human user. The community was encouraged to test this assumption and propose solutions that can improve the process of broadcast video clip selection. NHK provided 10 video programs, each 25 minutes long, as a testing dataset.

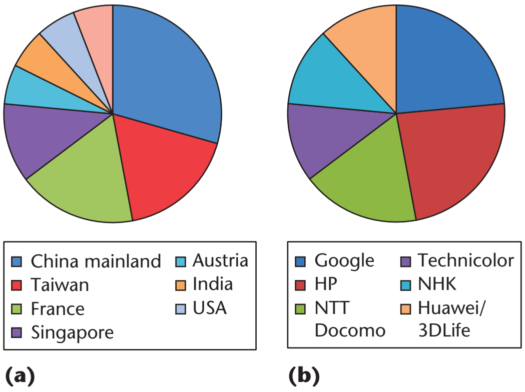


Figure 1. 2012 Multimedia Grand Challenge finalists. Breakdown (a) by country and (b) challenge.

Technicolor: Audio-Visual Recognition of Specific Events

The Technicolor challenge asked the community to search unstructured multimedia datasets based on audio-visual cues. The goal was to develop methods to automatically extract as much information as possible from the audio-visual query (such as compact low-level audio-visual signatures or text present in the images) and to use it to search the intertwined textual, audio, and visual components of the database. In particular, Technicolor asked, given a short video sequence of a public event, can a system automatically produce semantic information, such as an event's location and participants?

2012 Award Selection Criteria and Winners

This year, the Grand Challenge recognized the 17 finalists and awarded first- and second-place prizes as well as a multimodal prize. Following the tradition of past Grand Challenge programs, the first- and second-place winners were selected based on the following criteria:

- effectiveness of the solution in addressing the challenge,
- quality of the presentation and Q&A session, and
- reproducibility of the solution in real-world systems.

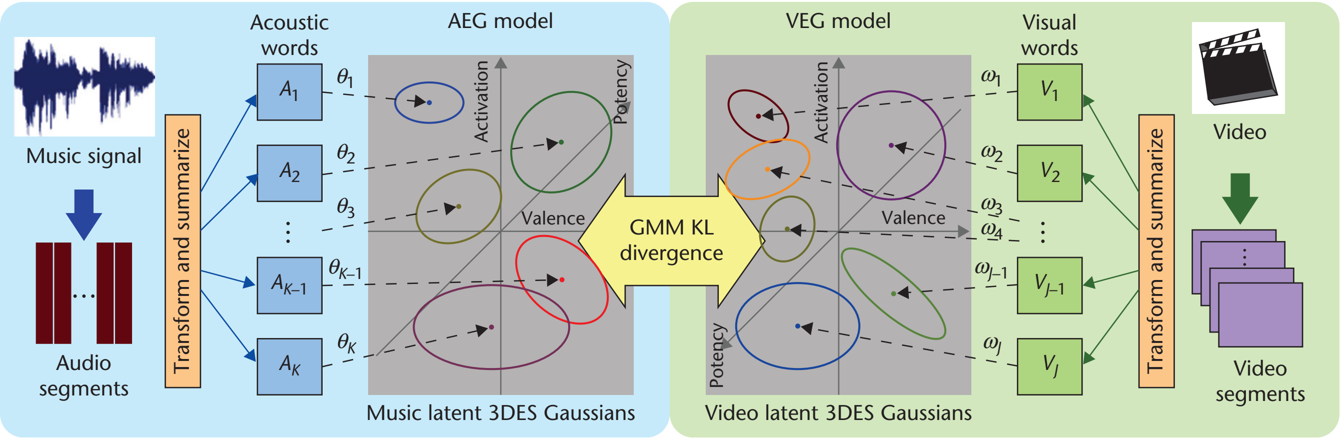


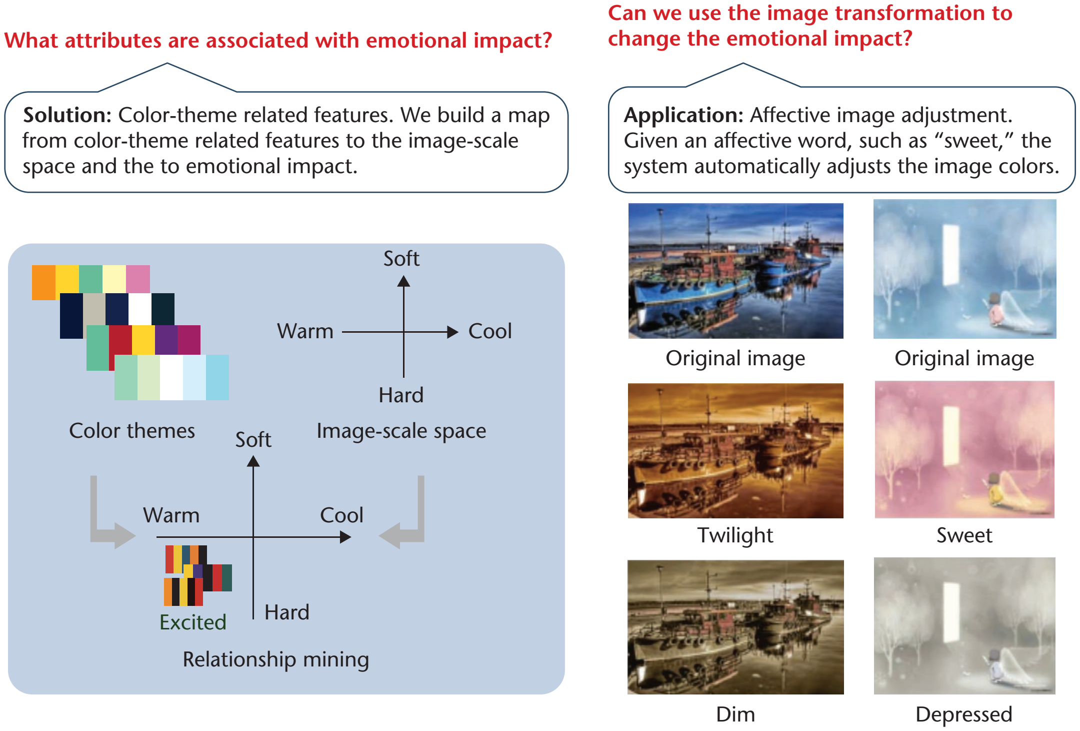
Figure 2. The 2012 Multimedia Grand Challenge first-place prize winner from Academia Sinica and National Taiwan University. The presenters addressed the Google Japan challenge of automatic music video generation.

The first-place prize was awarded to ''Acoustic-Visual Emotion Gaussians Model for Automatic Generation of Music Video'' by Ju-Chiang Wang, Yi-Hsuan Yang, I-Hong Jhuo, Yen-Yu Lin and Hsin-Min Wang from Academia Sinica and the National Taiwan University, Taiwan.[1] Addressing the Google Japan challenge, this work presented a novel machine-learning framework called Acoustic-Visual Emotion Gaussians to utilize the perceived emotion of multimedia content as a bridge to connect music and videos. For a music piece of a video sequence, they applied the AVEG model to predict the emotion distribution from the corresponding low-level acoustic and visual features (see Figure 2). The resulting ''emotion distributions'' help determine the match between audio and video.

This year, there was a tie for second place. The first winner was ''Understanding the Emotional Impact of Images'' by Xiaohui Wang, Jia Jia, Peiyun Hu, Sen Wu, Jie Tang, and Lianghong Cai from multiple institutions including the Key Laboratory of Pervasive Computing, Ministry of Education, TNList, and Tsinghua University, China.[2] The presenters proposed a content-based framework to estimate the likely emotional impact on the viewers of Web images (the HP challenge). Specifically, they used aesthetic categories (such as elegant, classic, and romantic) as the intermediate layer to reduce the gap between the visual features and high-level semantics, and their aesthetic models learned from Web images and online social networks. Figure 3 shows their resulting model, which they applied to affective image adjustment and emotion prediction from images.

The other second-place winner was ''Emotion-Based Sequence of Family Photos'' by Vassilios Vonikakis and Stefan Winkler from the Advanced Digital Science Center (ADSC), a unit of University of Illinois at Urbana-Champaign in Singapore.[3] Focusing on HP challenge, the presenters proposed a method for automatically creating slideshows from family photo collections based on the emotions of a given group of people. The user specifies the desired person(s) to be included in the slideshow. From there, as Figure 4 shows, a natural image sequence and meaningful slideshow transitions are formed based on people's emotions and several other user-defined image similarity attributes. This process uses a new image dissimilarity function that can integrate various attribute combinations and preferences.

The Multimodal Prize was given to ''Analyzing Social Media via Event Facets'' by Zhiyu Wang, Peng Cui, Lexing Xie, Hao Chen, Wenwu Zhu, and Shiqiang Yang from multiple institutions including the Tsinghua University, Beijing Key Laboratory of Networked Multimedia, and Australian National University and NICTA.[4] The presenters designed a rich-media analysis system to address the challenge of sensing and exploring events from social media in real time (the NTT Docomo challenge). The system includes a novel bilateral correspondence topic model for extracting



What attributes are associated with emotional impact?

Solution: Color-theme related features. We build a map from color-theme related features to the image-scale space and the to emotional impact.

Can we use the image transformation to change the emotional impact?

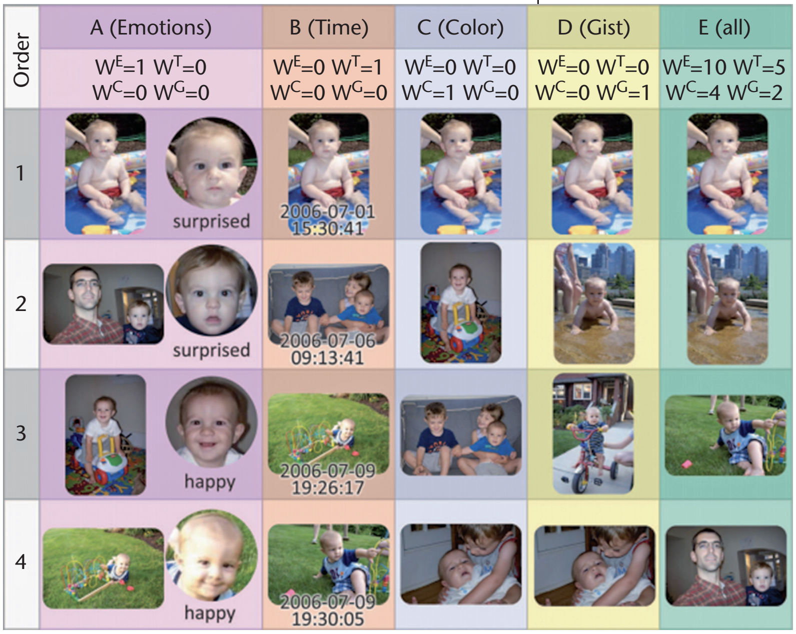
Application: Affective image adjustment. Given an affective word, such as "sweet," the system automatically adjusts the image colors.

Figure 3. 2012 Multimedia Grand Challenge second-place prize co-winner from the Key Laboratory of Pervasive Computing, TNList, and Tsinghua University. "Understanding the Emotional Impact of Images" addressed the HP challenge, the problem of understanding the emotional impact of images, and applied that understanding to improve photo selection and representation.

representative content and meaningful facets about events over time. It also includes a digital magazine that anchors user interaction with event facets. Figure 5 shows an example generated from more than 4 million rich media microblogs.

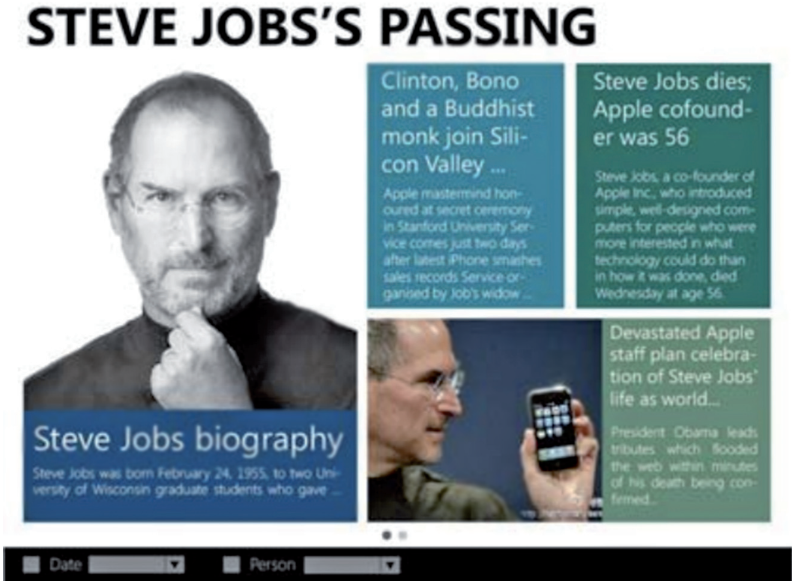
Looking Forward

The plenary session in Nara was enjoyed by the presenters, judges, and audience alike and was one of the highlights of the ACM Multimedia Conference. The community should keep this tradition alive with continuous innovation with new and even more challenging problems and further advancement of solutions. The Multimedia Grand Challenge once again showed that high science and business opportunities can be an excellent match. Via this exciting channel, multimedia solutions developed in our community might more easily find their way to multimedia search engines and products of the future. See you all again in Barcelona. **MM**



Order	A (Emotions)	B (Time)	C (Color)	D (Gist)	E (all)
	$W^E=1$ $W^T=0$ $W^C=0$ $W^G=0$	$W^E=0$ $W^T=1$ $W^C=0$ $W^G=0$	$W^E=0$ $W^T=0$ $W^C=1$ $W^G=0$	$W^E=0$ $W^T=0$ $W^C=0$ $W^G=1$	$W^E=10$ $W^T=5$ $W^C=4$ $W^G=2$
1	surprised	2006-07-01 15:30:41			
2	surprised	2006-07-06 09:13:41			
3	happy	2006-07-09 19:26:17			
4	happy	2006-07-09 19:30:05			

Figure 4. 2012 Multimedia Grand Challenge second-place prize co-winner from University of Illinois at Urbana-Champaign. "Emotion-Based Sequence of Family Photos" presented a method for automatically creating family-photo slideshows based on group emotions, addressing the HP challenge.



STEVE JOBS'S PASSING

Steve Jobs biography

Steve Jobs was born February 24, 1955, to two University of Wisconsin graduate students who gave ...

Clinton, Bono and a Buddhist monk join Silicon Valley ...

Apple mastermind honoured at secret ceremony in Stanford University Service comes just two days after latest iPhone smashes sales records Service organised by Job's widow ...

Steve Jobs dies; Apple cofounder was 56

Steve Jobs, a co-founder of Apple Inc., who introduced simple, well-designed computers for people who were more interested in what technology could do than in how it was done, died Wednesday at age 56.

Devastated Apple staff plan celebration of Steve Jobs' life as world...

President Obama leads tributes which flooded the web within minutes of his death being confirmed...

Date ▼ Person ▼

Figure 5. 2012 Multimedia Grand Challenge Multimodal Prize. The authors of ''Analyzing Social Media via Event Facets'' addressed the NTT Docomo challenge by analyzing social media using text and image information.

References

1. J.-C. Wang et al., ''The Acousticvisual Emotion Gaussians Model for Automatic Generation of Music Video,'' *Proc. 20th ACM Int'l Conf. Multimedia,* ACM, 2012, pp. 1379–1380.
2. X. Wang et al., ''Understanding the Emotional Impact of Images,'' *Proc. 20th ACM Int'l Conf. Multimedia,* ACM, 2012, pp. 1369–1370.
3. V. Vonikakis et al., ''Emotion-Based Sequence of Family Photos,'' *Proc. 20th ACM Int'l Conf. Multimedia,* ACM, 2012, pp. 1371–1372.
4. Z. Wang et al., ''Analyzing Social Media via Event Facets,'' *Proc. 20th ACM Int'l Conf. Multimedia,* ACM, 2012, pp. 1359–1360.

Yushi Jing is a senior scientist at Google Research. Contact him at jing@google.com.

Go Irie is a researcher at Nippon Telegraph & Telephone. Contact him at irie.go@lab.ntt.co.jp.

Marcel Worring is an associate professor at the University of Amsterdam. Contact him at m.worring@uva.nl.

PARALLEL AND DISTRIBUTED SYSTEMS
SERVICES COMPUTING
COMPUTATIONAL BIOLOGY AND BIOINFORMATICS
VISUALIZATION AND COMPUTER GRAPHICS
SOFTWARE ENGINEERING
PATTERN ANALYSIS AND MACHINE INTELLIGENCE
HAPTICS