EIC's Message

Herding Cats

John R. Smith
IBM T. J. Watson Research Center

The more things change, the more they stay the same. Over the last decade, digital video has gone from nowhere to everywhere. Today video makes up more than half of mobile and peak Internet traffic, and more video is being captured in a second than anyone can view in a lifetime. According to YouTube, one of today's most popular video sites, 100 hours of video are uploaded to its site every minute and more than 6 billion hours of video are watched each month (see www.youtube.com/yt/press/statistics.html).

But therein lies the problem. Searching for video is still difficult, and techniques for browsing and visualizing video content are primitive at best. The predominant method for video search today is based on user tags and limited metadata. If the search corpus is large enough and meta information is available, video search can often find some relevant videos. However, there are problems with both precision and recall in video search today. This means greater emphasis is needed on interactive tools that allow fast and effective triage of video search results.

As an illustrative example, the video searches in Table 1 were performed on YouTube. The table shows the number of results found for each of seven video searches. Although the sheer number of cat videos that can be searched on YouTube is truly amazing, there are problems, as the table shows. Searching for "kittens" provides more than 3 million results, which is way beyond what any user can scan. Attempts to make the search more specific by searching for "calico kittens playing" or "calico kittens playing outdoors with dog and children" reduce the number of results significantly. Still, the result size is still in the tens of thousands.

Finally, searching for "calico kittens playing outdoors with dog and children and cat toy" reduces the number of results to 2,570. However, these results look sketchy and more precise queries seem to reduce both precision and recall. If users are looking for specific content, there's really nothing they can do to find the relevant videos except click and watch each one.

Table 1. Video searches on YouTube.

Search	Number of results
Kittens	3,340,000
Calico kittens	55,600
Calico kittens playing	42,400
Calico kittens playing outdoors	34,100
Calico kittens playing outdoors with dog	37,000
Calico kittens playing outdoors with dog and children	10,400
Calico kittens playing outdoors with dog and children and cat toy	2,570

This is a situation where compact visual summaries are needed to make video triage easier. The most basic video summary is a thumbnail, which is what YouTube provides. Thumbnail images have been around since the beginning of the Web and are commonly used for image search. Providing a single frame thumbnail is obviously an impoverished visual summary for video given it's a temporal medium. Some sites go further and provide an animated thumbnail that displays a sequence of frames from the video. More advanced techniques base the animated sequence on automatically detected keyframes from the video. Although animations can provide a better summary, they display their information over time, which can slow down triage. Alternatively, techniques like video filmstrips and mosaics can display the keyframe information in 2D space, making it easier to see more of the video content at a glance. However, if videos are long, these representations can be dense or take up too much screen real estate.

1070-986X/13/$31.00 © 2013 IEEE

Published by the IEEE Computer Society

New Editorial Board Members

Benoit Huet is an assistant professor in the multimedia information processing group at Eurecom, France. In 1993, he received an MSc in artificial intelligence from the University of Westminster, UK, where he then spent two years working as a research and teaching assistant. Huet received his DPhil in computer science from the University of York, UK, for his research on object recognition from large databases. He was awarded the HDR (Habilitation to Direct Research) from the University of Nice Sophia Antipolis, France, in October 2012 on the topic of "Multimedia Content Understanding: Bringing Context to Content." He is an associate editor for *Multimedia Tools and Application* and *Multimedia Systems* and has been guest editor for special issues of the *EURASIP Journal on Image and Video Processing* and *IEEE MultiMedia*. He chaired the Multimedia Modeling Conference in 2009 and has cochaired several other workshops and special conference sessions. He regularly serves on the technical program committee of the top conferences in the field, including ACM Multimedia, ACM International Conference on Multimedia Retrieval (ICMR), and IEEE International Conference on Multimedia and Expo (ICME). He is chairing the IEEE MMTC Interest Group on Visual Analysis, Interaction, and Content Management (VAIG) and is vice chair of the International Association for Pattern Recognition (IAPR) Technical Committee 14 on Signal Analysis for Machine Intelligence. For more information, visit www.eurecom.fr/en/people/huet-benoit.

Gerald Friedland is the director of audio and multimedia research at the International Computer Science Institute, a private lab affiliated with the University of California, Berkeley, where he is currently leading a group of multimedia researchers supported by NSF, IARPA, and industry grants. The group mostly specializes in audio methods for video content analysis, applying speech, music, and environmental sound analysis to tasks like automatic geolocation of consumer-produced videos, video concept detection, and privacy related topics. Friedland has recently finished authoring a new textbook on multimedia computing together with Ramesh Jain (to appear spring 2014). Friedland cofounded the IEEE International Conference on Semantic Computing and is the founder and program director of the IEEE International Summer School on Semantic Computing at UC Berkeley. He is an associate editor of *ACM Transactions on Multimedia Computing, Communications, and Applications* and has served on the organization committee of ACM Multimedia 2011, 2012, and 2014 as well as the TPC cochair of IEEE ICME 2012. He received the European Academic Software Award and the Multimedia Entrepreneur Award by the German Federal Department of Economics and lead the team that won the ACM Multimedia Grand Challenge in 2009. Friedland received an MS and PhD in computer science from Freie Universitaet Berlin, Germany, in 2002 and 2006, respectively. He is a senior member of ACM and IEEE. For more information, visit www.icsi.berkeley.edu/~fractor.

There may not be a single, best video summary for all situations. Depending on the context of the user's search and other factors such as screen size, bandwidth, and number of search results, there should be different tradeoffs in space, time, and content for making effective and efficient video summaries.

Video has indeed come a long way. But video search has not. Video summarization and search go hand in hand, and more effective techniques are needed for both. Until then, finding relevant videos will remain akin to herding cats. **MM**

John R. Smith is a senior manager of Intelligent Information Management at IBM T.J. Watson Research Center. Contact him at jsmith@us.ibm.com.

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