EIC's Message

Clicking on Things

John R. Smith
IBM Research

The world is becoming clickable. I don't mean buttons are being attached to things—at least not actual physical buttons. Rather, physical objects are being linked to the digital world using multimedia technologies such as audiovisual content recognition and large-scale multimedia content-based search.

It all started simply enough with books, CDs, and DVDs. For example, in a typical scenario, a person with a mobile device takes a picture of the cover of a friend's book. Then a service matches the image to known books to find the right one and connects the user to more information or allows him or her to buy the book. Relying on this kind of content-based visual search for books is nice, although not really critical. Of course, the same person could just as easily type in the book title and search for it that way. But it's definitely more fun with a mobile device camera.

Similar technologies for visual recognition are being applied to other content you wouldn't necessarily want to type in—such as advertisements in magazines, newspaper pages, scientific articles, slides on a screen, TV programs, posters, and paintings—to link these physical artifacts to digital content, services, or related online information. And multiple commercial services and products are being developed that expand on the use of audio, music, voice, and visual matching to make different facets of the real world clickable (see the "Things To Click On" sidebar for some examples).

To make these kinds of applications work in practice, content-based matching technologies need to be robust under a wide range of conditions such as noise, lighting variation, perspective transformation, rotation, cropping, occlusion, blurring, and zooming. And they need to work on a large scale, which means a huge database of objects needs to be quickly and accurately searched to find the correct matches. Take books as an example. It's estimated that there are approximately 130 million books in the world.[1] The number is higher if you consider unique book covers. In the case of other media types, at least one commercial solution has as many as eight million CDs, 100 million music tracks, and 400,000 DVDs.[2] Making content-based matching work at these scales requires design of compact descriptors that effectively capture the salient features of the objects as well as large-scale indexing techniques that allow highly efficient matching.

This gets really technically challenging and even more fun when applied to the real world beyond planar 2D surfaces (such as books, covers, pages, and screens) to everyday 3D objects.

Things To Click On

Examples of commercial products and services for clicking on things include:

- ClusterMedia Labs (semantic audiovisual analysis),
- Digimarc Mobile (linking print and digital using watermarking),
- Google Goggles (mobile picture search),
- Gracenote (content search and music recognition),
- IQ Engines (image recognition engine),
- Kooaba (smart visuals of newspapers and magazines, buildings),
- Like.com (visual search and shopping),
- LinkMe Mobile (visual recognition),
- Mobile Acuity (mobile visual product search),
- Mobot (mobile visual search),
- Nokia Point and Find (automated object recognition),
- Shazam (content-based music search),
- Snapnow (mobile visual search service),
- Snaptell (visual product search), and
- TinEye (reverse image search).

1070-986X/10/$26.00 © 2010 IEEE

Published by the IEEE Computer Society

New Additions

The magazine welcomes Hari Sundaram, Rong Yan, and Winston Hsu to the *IEEE MultiMedia* editorial board.

Sundaram is an associate professor of media arts and computing in the School of Arts Media and Engineering and in the computer science department at Arizona State University. He received a PhD from the Department of Electrical Engineering at Columbia University in 2002. His research focuses on designing intelligent media environments that exist as part of our physical world (for example, mediated environments that help stroke patients recover and information search), and developing new algorithms and systems to understand online human activity. His research has won several awards—including the best student paper award at Joint Conference on Digital Libraries (JCDL) in 2007 and the best ACM Multimedia demo award in 2006. He received the best student paper award at ACM Multimedia 2002, and the Eliahu I. Jury Award for best PhD dissertation in 2002. He also received in 2000 a best paper award on video retrieval from *IEEE Transactions on Circuits and Systems for Video Technology.*

Yan is a research scientist at Facebook. He was a research staff member in the IBM T.J. Watson Research Center from 2006 to 2009. He received an MS in 2004 and a PhD in 2006 from Carnegie Mellon University's School of Computer Science. Yan's research interests include large-scale machine learning, data mining, advertisement optimization, multimedia information retrieval, and computer vision. He received the best paper runner-up awards at ACM Multimedia in 2004 and ACM Conference on Image and Video Retrieval in 2007. He received the IBM Research External Recognition Award in 2007.

Hsu is an assistant professor in the Graduate Institute of Networking and Multimedia and the Department of Computer Science and Information Engineering, National Taiwan University, and the founder of Multimedia Indexing, Retrieval, and Analysis (MiRA) Research Group. He received a PhD in 2007 from Columbia University. Before that, he was devoted to a multimedia software startup, CyberLink, as engineer, project leader, and R&D manager. Hsu's current research interests are to enable next-generation multimedia retrieval and generally include content analysis, mining, retrieval, and machine learning over large-scale multimedia databases. His research proposal in image and video reranking received the best paper runner-up award at ACM Multimedia 2006.

The potential augmented-reality applications are mind-boggling and span a wide range of settings in e-commerce, travel and tourism, education, and product and service reviews. For example, imagine snapping a photo of your friend's new shoes and immediately finding them online. The same idea applies to other kinds of clothing, such as ties and t-shirts, or objects such as cars, bikes, toys, and so on. Real-time visual search can be applied for travel to aid in navigation or help find and recognize landmarks, signs, and buildings. Or it can enhance sightseeing by recognizing and automatically retrieving information about monuments, museums, and art work. It can also be used to provide services that deliver reviews on-demand. Consider the case when you are walking down Main St. and want to know which restaurant to go to. Simply snap your picture of a candidate restaurant and get your ratings, reviews, recommended dishes, and other relevant information right on the spot. It can work similarly for stores, theaters and other venues.

The logical conclusion of this multimedia content-based approach of clicking on things to link the physical and digital world is its intersection with the Internet of things (see http://en.wikipedia.org/wiki/Internet_of_Things), which has the goal of networked interconnection of everyday objects using technologies such as RFID, barcodes, tag readers, and other sensors. Mobile cameras, microphones, and audiovisual content-recognition and multimedia search technologies will become additional mechanisms for realizing the ultimate goal of automatic identification and tracking of up to 100 trillion everyday objects.

Now that's a lot of things to click on. **MM**

References

1. L. Taycher, ''Books of the World, Stand Up and Be Counted! All 129,864,880 of You,'' blog 5 Aug. 2010; http://booksearch.blogspot.com/2010/08/books-of-world-stand-up-and-be-counted.html.
2. Gracenote, ''Gracenote Reaches Global Media Database Milestone,'' 9 Sept. 2010; http://www.gracenote.com/press/09/09/2010/.

Contact John R. Smith at jsmith@us.ibm.com.

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