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. 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. 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.
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. Or it can 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.

References


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

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