Kirk L. Kroeker

Face Recognition Breakthrough

By using sparse representation and compressed sensing, researchers have been able to demonstrate significant improvements in accuracy over traditional face-recognition techniques.

THE THEORIES OF sparse representation and compressed sensing have emerged in recent years as powerful methods for efficiently processing data in unorthodox ways. One of the areas where these theories are having a major impact today is in computer vision. In particular, the theories have given new life to the field of face recognition, which has seen only incremental increases in accuracy and efficiency in the past few decades. Now, thanks to the application of these theories to classic face-recognition problems, researchers at the University of Illinois at Urbana-Champaign (UIUC) have been able to demonstrate significant improvements in accuracy over traditional techniques.

The idea of applying sparse representation and compressed sensing to face recognition was so novel in 2007 that two papers outlining the method were rejected by mainstream vision conferences. "Just like compressed sensing, our approach to face recognition is completely unorthodox," says Yi Ma, a professor of electrical and computer engineering at UIUC. "The reviewers simply did not believe such good results were possible, or that sparsity was even relevant."

Ma had been studying the sparse representation and compressed sensing theories of Emmanuel Candes and David Donoho while on sabbatical at the University of California, Berkeley in early 2007. It was then, he says, that he connected the theories to computer vision by applying the ideas to problems associated with one of the most readily available sources of raw data for vision research: face images. "The results were far beyond what I had ever expected or imagined from the beginning," Ma says. "What happened next was the most exciting period of research I have ever had."
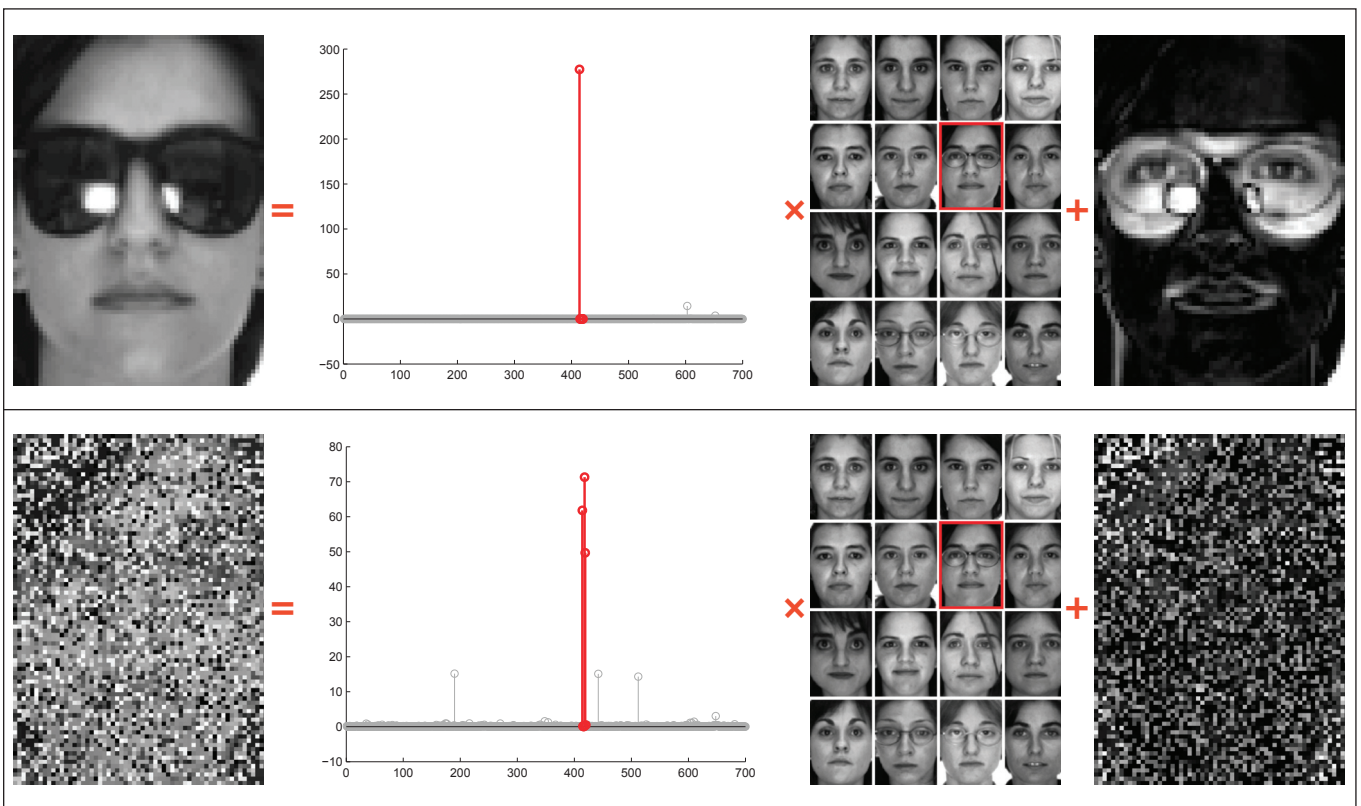
The face-recognition method developed at the University of Illinois at Urbana-Champaign. On the left, the test face is partially covered with sunglasses (top) and corrupted (bottom). The face is represented in the middle as a sparse linear combination of the training images plus sparse errors due to occlusion (right top) and corruption (right bottom), with the red coefficients corresponding to the training images of the correctly identified individual.

While traditional face-recognition methods record certain facial measurements, such as the distance between the eyes and the width of the mouth, the method developed by Ma turns the conventional wisdom about such strategies on its head by representing faces as the sparsest linear combination of images in a database. Unlike the traditional methods that focus on low-dimensional features, Ma's technique works directly with high-dimensional images as a whole, focusing on their sparse structures.

In a way, says Ma, the idea is simple. "Given a database of face images, our algorithm simply tries to use the fewest possible images to interpret a query image," he says. The idea is based on the notion of treating the given training data as a large dictionary for representing test images. For each test image, the algorithm seeks the sparsest representation from the main dictionary and from an auxiliary dictionary consisting of the individual pixels.

"The use of the data itself as dictionary is a very interesting concept for certain types of data and applications," says Guillermo Sapiro, whose recent work as a professor in the department of electrical and computer engineering at the University of Minnesota has focused on sparse representation and dictionary learning. "The face-recognition work being carried out by Ma and his colleagues represents a novel take on this problem and a very refreshing one that we are all looking at with a lot of excitement."

Accurate Identification

The aspect of Ma's approach that is arguably generating the most excitement is its ability to identify faces despite the images being corrupted or the faces being partially covered. Traditional face-recognition techniques are susceptible to noise and their accuracy is significantly reduced when parts of a face are covered or obscured, such as with a mask or a disguise. Ma says the algorithm can handle up to 80% random corruption or occlusion of the face image and still reliably recognize a person, making it more capable than even the human brain in its ability to identify a face.

"Our theorem even suggests that the percentage of corruption can ap-

Yi Ma envisions a future in which a user can capture a person's picture with a mobile phone's camera, submit a query, and receive a match moments later.

proach 100% as the dimension goes to infinity," says Ma. In contrast, the accuracy of traditional face-recognition techniques declines to less than 70% when part of a face is covered or if the query image suffers from noise or poor resolution. Ma's algorithm can withstand such occlusions and corruptions because it does not focus on specific facial details, such as the size of a nose. In Ma's algorithm, the sparsest representation accounts for the parts of the face that are not occluded or corrupted. This capability alone has drawn the attention not only of the media, but also of more than a dozen companies that are interested in licensing the technology.

"We've been pleasantly surprised by the breadth of interest we've received, both from companies working in traditional application areas of face recognition, such as security and access control, as well as in many less-traditional areas," says John Wright, a UIUC graduate student who won the $30,000 Lemelson-Illinois Student Prize this year for his work creating a prototype application using Ma's algorithm.

However, despite the commercial interest, several challenges remain. "Although the initial idea seemed very natural, it has since taken a lot of mathematical work to justify why it should work so well," says Wright. "It also has taken a lot of engineering work to bring it into the real world." One of the challenges, as with other face-recognition systems, is how to achieve high performance with massive data sets.

The UIUC method is based on a scalable algorithm, but Ma says that if

the number of subjects becomes large, running the algorithm in real time on current hardware is challenging. Currently, he says, the algorithm can run in real time on a database consisting of up to 1,000 subjects, making it suitable for use in an access-control system for a small company. However, the UIUC team is working on parallel and graphics processing unit implementations so the system can scale to much larger numbers, possibly even to millions of subjects.

Given the algorithm's accuracy, the implications of scaling it to millions of subjects could be significant, leading to new ways of searching for images, annotating multimedia, and even monitoring crowds. For example, Ma envisions a Web-based service that would allow users to capture a person's picture with a mobile phone's camera, remotely submit a query, and have a match returned moments later. "In my view," says Ma, "this would be a killer app for the emerging parallel, distributed, and cloud initiatives."

Another challenge the UIUC team faces is to extend the method to deal with less controlled training data, particularly for applications such as online photo tagging or image searching. For these applications, the researchers say, it is crucial to understand the minimum amount of training data needed for the algorithm to succeed. But for Ma, this challenge, in particular, is not specific to face recognition. "One of the fundamental problems that we need to address in computer vision," he says, "is how to obtain or learn a good dictionary for the problem at hand."

Still, Ma says he and his team remain confident that for applications where the acquisition of training images can be controlled, an obtainable goal for their technology is real-time, highly accurate face recognition that far exceeds the capabilities of current systems. "I am confident that with sufficient support from the industry and possibly from the government," he says, "we could start to see working face-recognition systems based on this research in the next five to 10 years." ▣

Kirk L. Kroeker works in communications and has written extensively about the impact of emerging technologies.