Researchers teach computers how to name images by 'thinking'

November 1 2006

Penn State researchers have "taught" computers how to interpret images using a vocabulary of up to 330 English words, so that a computer can describe a photograph of two polo players, for instance, as "sport," "people," "horse," "polo."

The new system, which can automatically annotate entire online collections of photographs as they are uploaded, means significant time-savings for the millions of Internet users who now manually tag or identify their images. It also facilitates retrieval of images through the use of search terms, said James Wang, associate professor in the Penn State College of Information Sciences and Technology, and one of the technology's two inventors.

The system is described in a paper, "Real-Time Computerized Annotation of Pictures," given at the recent ACM Multimedia 2006 conference in Santa Barbara, Calif., and authored by Jia Li, associate professor, Department of Statistics, and Wang. Penn State has filed a provisional patent application on the invention. Major search engines currently rely upon uploaded tags of text to describe images. While many collections are annotated, many are not. The result: Images without text tags are not accessible to Web searchers. Because it provides text tags, the ALIPR system-Automatic Linguistic Indexing of Pictures-Real Time-makes those images visible to Web users.

ALIPR does this by analyzing the pixel content of images and comparing that against a stored knowledge base of the pixel content of tens of

thousands of image examples. The computer then suggests a list of 15 possible annotations or words for the image.

"By inputting tens of thousands of images, we have trained computers to recognize certain objects and concepts and automatically annotate those new or unseen images," Wang said. "More than half the time, the computer's first tag out of the top 15 tags is correct."

In addition, for 98 percent of images tested, the system has provided at least one correct annotation in the top 15 selected words. The system, which completes the annotation in about 1.4 seconds, also can be applied to other domains such as art collections, satellite imaging and pathology slides, Wang said. The new system builds on the authors' previous invention, ALIP, which also analyzes image content. But unlike ALIP which characterized images by incorporating computational-intensive spatial modeling, ALIPR characterizes images by modeling distributions of color and texture.

The researchers acknowledge computers trained with their algorithms have difficulties when photos are fuzzy or have low contrast or resolution; when objects are shown only partially; and when the angle used by the photographer presents an image in a way that is different than how the computer was trained on the object. Adding more training images as well as improving the training process may reduce these limitations-future areas of research.

Source: Penn State