Researchers Find Way To Make Internet Video More Appealing

December 21 2004

Jay Leno's comedy routines are helping to advance technologies for distance learning on the Internet.
Ohio State University engineers are using video recordings of Leno and other TV personalities to test software that transmits more information in an Internet video using less bandwidth.

One of the obstacles to distance learning on the Internet is the difficulty with viewing lectures, explained James Davis, professor of computer science and engineering at Ohio State. A high-resolution video of a speaker takes too long to download, but a low-resolution video makes fine details such as the speaker's face and hands appear fuzzy.

"When we communicate, we say a lot with our face and hands," Davis said. "Our voice, gestures, and facial expressions are all intertwined. If I'm watching a lecture and I'm trying to learn something, I need to be able to see the speaker's face and hands."

He and his students have created software that zeroes in on a speaker's face and gesturing hands and sharpens the image in just those spots, while slightly lowering the resolution of the rest of the image. In that way, the final video communicates more information without increasing bandwidth.

In a recent issue of the journal Computer Vision and Image Understanding, Davis and former undergraduate student Robin Tan reported that the software worked successfully in initial tests. The

engineers were able to enhance a speaker's face and gesturing hands without increasing bandwidth, and -- not surprisingly -- users greatly preferred viewing the enhanced video stream.

Davis and Tan inserted their algorithms into a publicly available MPEG encoder -- the software that compresses video for efficient digital transmission on the Internet -- and an MPEG decoder that converts the digital signal back to video so a user can view it on their computer.

For the test, five volunteers watched three short video sequences: one taken from an educational lecture on public television; one from the opening monologue of "Saturday Night Live," featuring actor Leslie Nielsen; and one taken from a Jay Leno monologue on "The Tonight Show."

Using a typical computer monitor, they watched the enhanced video and the unenhanced video side-by-side.

The unenhanced version was typical of what most people see when they try to watch video over the Internet -- pixelated and blurry. The enhanced version featured clearer visuals of the speaker's face and gesturing hands, with a slightly lower-quality background than in the unenhanced video. For both versions, the number of kilobytes of data per frame was kept constant, so they both required the same amount of bandwidth for transmission.

All five volunteers immediately picked the enhanced video as their favorite.

Davis was motivated to pursue this work several years ago, when he was a research assistant in the Media Laboratory at the Massachusetts Institute of Technology. There, he interacted with researchers that studied video of Leno and others in order to better understand how

people use speech and gestures in communication.

He took note of patterns in gestures, such as Leno's technique of opening his hands to the audience when he delivers a punch line. A "pounding" gesture is also common, when speakers are trying to make a point, he said.

The hard part was enabling the software to recognize these hand gestures. He and Tan wrote those algorithms at Ohio State, and combined them with other algorithms to track a speaker's face and hands as they moved.

In the future, Davis plans to incorporate algorithms that can track the movements of multiple speakers at once. He also wants to selectively enhance background objects, such as whiteboards, when they are used in a lecture.

Source: Ohio State University