Extreme makeover: computer science edition

November 12 2008

Siddharth Batra, graduate student in computer science, places the Stanford logo onto a wall of the Hewlett Teaching Center using software that can embed graphics into a video—images, photos or even another video.

(PhysOrg.com) -- Suppose you have a cherished home video, taken at your birthday party. You're fond of the video, but your viewing experience is marred by one small, troubling detail. There in the video, framed and hanging on the living room wall amidst the celebration, is a color photograph of your former significant other.

Bummer.

But what if you could somehow reach inside the video and swap the offending photo for a snapshot of your current love? How perfect would that be?

A group of Stanford University researchers specializing in artificial intelligence have developed software that makes such a switch relatively simple. The researchers, computer science graduate students Ashutosh Saxena and Siddharth Batra, and Assistant Professor Andrew Ng, see interesting potential for the technology they call ZunaVision.

They say a user of the software can easily plunk an image on almost any planar surface in a video, whether wall, floor or ceiling. And the embedded images don't have to be still photos—you can insert a video inside a video.

Here's the opportunity to sing karaoke side-by-side with your favorite American Idol celebrity and post the video to YouTube. Or preview a virtual copy of a painting on your wall before you buy. Or liven up those dull vacation videos.

There is also a potential financial aspect to the technology. The researchers suggest that anyone with a video camera might earn some spending money by agreeing to have unobtrusive corporate logos placed inside their videos before they are posted online. The person who shot the video, and the company handling the business arrangements, would be paid per view, in a fashion analogous to Google AdSense, which pays websites to run small ads.

The embedding technology is driven by an algorithm that first analyzes the video, with special attention paid to the section of the scene where the new image will be placed. The color, texture and lighting of the new image are subtly altered to blend in with the surroundings. Shadows seen in the original video will be seen in the added image as well. The result is a photo or video that appears to be an integral part of the original scene, rather than a sticker pasted artificially on the video.

For the algorithm ("3D Surface Tracker Technology") to produce these

realistic results, it also must deal with what researchers call "occluding objects" in the video. In our birthday video, an "occluding object" might be a partygoer walking in front of the newly hung photo. The algorithm can handle most such objects by keeping track of which pixels belong to the photo and which belong to the person walking in the foreground; the photo disappears behind the person walking by and then reappears, just as in the original video.

Camera motion gives the algorithm another item to digest. As the camera pans and zooms, the portion of the wall containing the embedded object moves and changes shape. The embedded image must keep up with this shape-shifting geometry, or the video may go one direction while the embedded image goes another.

To prevent such mishaps, the algorithm builds a model, pixel by pixel, of the area of interest in the video. "If the lighting begins to change with the motion of the video or the sun or the shadows, we keep a belief of what it will look like in the next frame. This is how we track with very high sub-pixel accuracy," Batra said. It's as if the embedded image makes an educated guess of where the wall is going next, and hurries to keep up.

Other technologies can perform these tricks—witness the spectacular special effects in movies and the virtual first-down lines on televised football games—but the Stanford researchers say the existing systems are expensive, time consuming and require considerable expertise.

Some of the recent Stanford work grew out of an earlier project, Make3D, a website that converts a single still photograph into a brief 3D video. It works by finding planes in the photo and computing their distance from the camera, relative to each other.

"That means, given a single image, our algorithm can figure out which parts are in the front and which parts are in the background," said

Saxena. "Now we have extended this technology to videos."

The researchers realize that their technology will be used in unpredictable ways, but they have some guesses. "Suppose you're a student living in a dorm and suppose you want to show it to your parents [in a video]. You can put a nice poster there of Albert Einstein," Batra said. "But if you want to show it to your friends, you can have a Playboy poster there."

A hands-on demonstration of the technology can be seen at zunavision.stanford.edu .

Provided by Stanford University

Citation: Extreme makeover: computer science edition (2008, November 12) retrieved 23 April 2024 from https://phys.org/news/2008-11-extreme-makeover-science.html