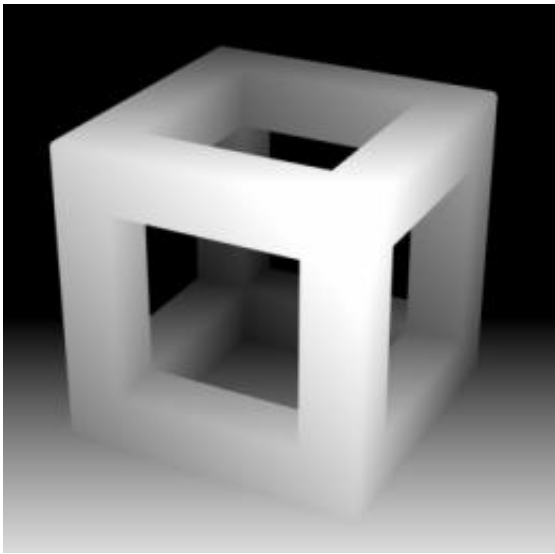


Clever math could enable a high-quality 3-D camera for cellphones

January 6 2012, by Larry Hardesty



Depth-sensing cameras can produce 'depth maps' like this one, in which distances are depicted as shades on a gray-scale spectrum (lighter objects are closer, darker ones farther away). Image: flickr/Dominic

When Microsoft's Kinect -- a device that lets Xbox users control games with physical gestures -- hit the market, computer scientists immediately began hacking it. A black plastic bar about 11 inches wide with an infrared rangefinder and a camera built in, the Kinect produces a visual map of the scene before it, with information about the distance to individual objects. At MIT alone, researchers have used the Kinect to create a "Minority Report"-style computer interface, a navigation system for miniature robotic helicopters and a holographic-video transmitter,

among other things.

Now imagine a device that provides more-accurate depth information than the Kinect, has a greater range and works under all lighting conditions — but is so small, cheap and power-efficient that it could be incorporated into a cellphone at very little extra cost. That's the promise of recent work by Vivek Goyal, the Esther and Harold E. Edgerton Associate Professor of Electrical Engineering, and his group at MIT's Research Lab of Electronics.

"3-D acquisition has become a really hot topic," Goyal says. "In consumer electronics, people are very interested in 3-D for immersive communication, but then they're also interested in 3-D for human-computer interaction."

Andrea Colaco, a graduate student at MIT's Media Lab and one of Goyal's co-authors on a paper that will be presented at the IEEE's International Conference on Acoustics, Speech, and Signal Processing in March, points out that gestural interfaces make it much easier for multiple people to interact with a computer at once — as in the dance games the Kinect has popularized.

"When you're talking about a single person and a machine, we've sort of optimized the way we do it," Colaco says. "But when it's a group, there's less flexibility."

Ahmed Kirmani, a graduate student in the Department of Electrical Engineering and Computer Science and another of the paper's authors, adds, "3-D displays are way ahead in terms of technology as compared to 3-D cameras. You have these very high-resolution 3-D displays that are available that run at real-time frame rates.

"Sensing is always hard," he says, "and rendering it is easy."

Clocking in

Like other sophisticated depth-sensing devices, the MIT researchers' system uses the “time of flight” of light particles to gauge depth: A pulse of infrared laser light is fired at a scene, and the camera measures the time it takes the light to return from objects at different distances.

Traditional time-of-flight systems use one of two approaches to build up a “depth map” of a scene. LIDAR (for light detection and ranging) uses a scanning laser beam that fires a series of pulses, each corresponding to a point in a grid, and separately measures their time of return. But that makes data acquisition slower, and it requires a mechanical system to continually redirect the laser. The alternative, employed by so-called time-of-flight cameras, is to illuminate the whole scene with laser pulses and use a bank of sensors to register the returned light. But sensors able to distinguish small groups of light particles — photons — are expensive: A typical time-of-flight camera costs thousands of dollars.

The MIT researchers' system, by contrast, uses only a single light detector — a one-pixel camera. But by using some clever mathematical tricks, it can get away with firing the laser a limited number of times.

The first trick is a common one in the field of compressed sensing: The light emitted by the laser passes through a series of randomly generated patterns of light and dark squares, like irregular checkerboards. Remarkably, this provides enough information that algorithms can reconstruct a two-dimensional visual image from the light intensities measured by a single pixel.

In experiments, the researchers found that the number of laser flashes — and, roughly, the number of checkerboard patterns — that they needed to build an adequate depth map was about 5 percent of the number of pixels in the final image. A LIDAR system, by contrast, would need to

send out a separate laser pulse for every pixel.

To add the crucial third dimension to the depth map, the researchers use another technique, called parametric signal processing. Essentially, they assume that all of the surfaces in the scene, however they're oriented toward the camera, are flat planes. Although that's not strictly true, the mathematics of light bouncing off flat planes is much simpler than that of light bouncing off curved surfaces. The researchers' parametric algorithm fits the information about returning light to the flat-plane model that best fits it, creating a very accurate depth map from a minimum of visual information.

On the cheap

Indeed, the algorithm lets the researchers get away with relatively crude hardware. Their system measures the time of flight of photons using a cheap photodetector and an ordinary analog-to-digital converter — an off-the-shelf component already found in all cellphones. The sensor takes about 0.7 nanoseconds to register a change to its input.

That's enough time for light to travel 21 centimeters, Goyal says. "So for an interval of depth of 10 and a half centimeters — I'm dividing by two because light has to go back and forth — all the information is getting blurred together," he says. Because of the parametric algorithm, however, the researchers' system can distinguish objects that are only two millimeters apart in depth. "It doesn't look like you could possibly get so much information out of this signal when it's blurred together," Goyal says.

The researchers' algorithm is also simple enough to run on the type of processor ordinarily found in a smartphone. To interpret the data provided by the Kinect, by contrast, the Xbox requires the extra processing power of a graphics-processing unit, or GPU, a powerful

special-purpose piece of hardware.

“This is a brand-new way of acquiring depth information,” says Yue M. Lu, an assistant professor of electrical engineering at Harvard University. “It’s a very clever way of getting this information.” One obstacle to deployment of the system in a handheld device, Lu speculates, could be the difficulty of emitting light pulses of adequate intensity without draining the battery.

But the light intensity required to get accurate depth readings is proportional to the distance of the objects in the scene, Goyal explains, and the applications most likely to be useful on a portable device — such as gestural interfaces — deal with nearby objects. Moreover, he explains, the researchers’ system makes an initial estimate of objects’ distance and adjusts the intensity of subsequent light pulses accordingly.

The telecom giant Qualcomm, at any rate, sees enough promise in the technology that it selected a team consisting of Kirmani and Colaco as one of eight winners — out of 146 applicants from a select group of universities — of a \$100,000 grant through its 2011 Innovation Fellowship program.

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: Clever math could enable a high-quality 3-D camera for cellphones (2012, January 6) retrieved 9 April 2024 from <https://phys.org/news/2012-01-clever-math-enable-high-quality-d.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private

study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.