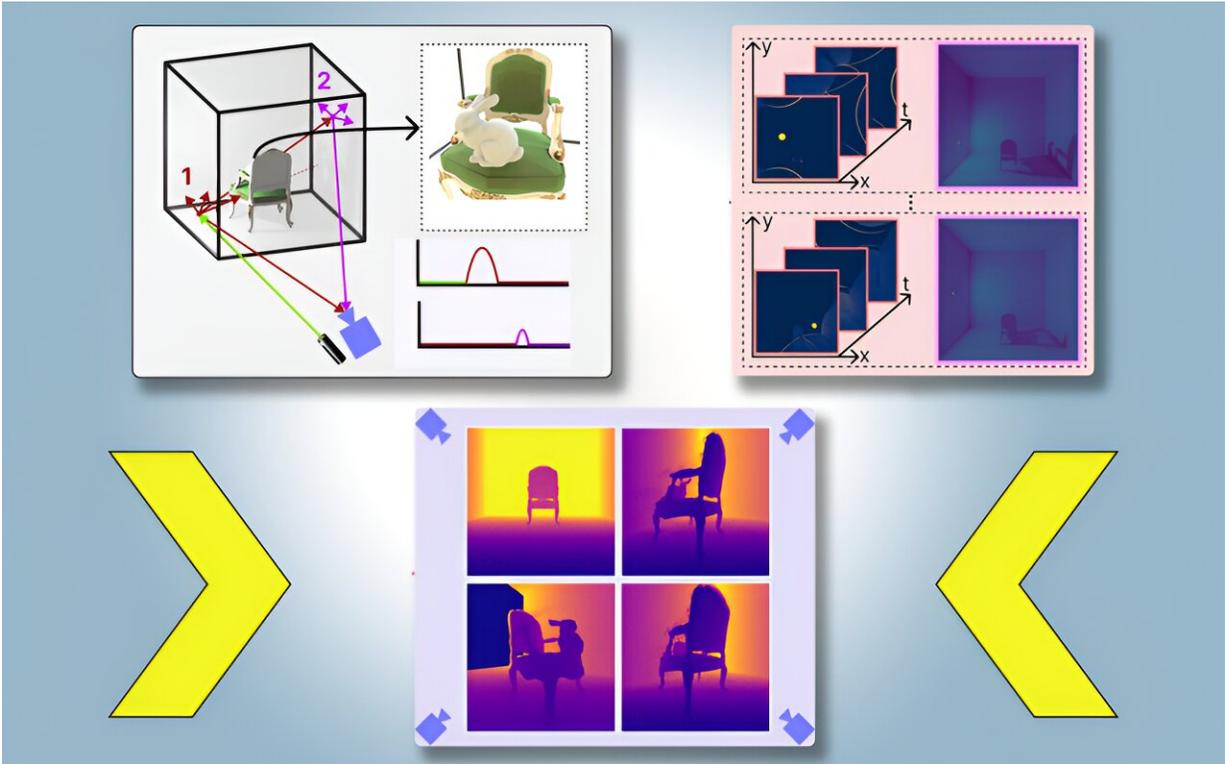


Researchers leverage shadows to model 3D scenes, including objects blocked from view

June 18 2024, by Adam Zewe



Plato-NeRF is a computer vision system that combines lidar measurements with machine learning to reconstruct a 3D scene, including hidden objects, from only one camera view by exploiting shadows. Here, the system accurately models the rabbit in the chair, even though that rabbit is blocked from view. Credit: Massachusetts Institute of Technology

Imagine driving through a tunnel in an autonomous vehicle, but

unbeknownst to you, a crash has stopped traffic up ahead. Normally, you'd need to rely on the car in front of you to know you should start braking. But what if your vehicle could see around the car ahead and apply the brakes even sooner?

Researchers from MIT and Meta have developed a computer vision technique that could someday enable an [autonomous vehicle](#) to do just that.

They have introduced a [method](#) that creates physically accurate, 3D models of an entire scene, including areas blocked from view, using images from a single camera position. Their technique uses shadows to determine what lies in obstructed portions of the scene.

They call their approach PlatoNeRF, based on Plato's allegory of the cave, a passage from the Greek philosopher's "Republic" in which prisoners chained in a cave discern the reality of the outside world based on shadows cast on the cave wall.

By combining [lidar](#) ([light detection](#) and ranging) technology with [machine learning](#), PlatoNeRF can generate more accurate reconstructions of 3D geometry than some existing AI techniques. Additionally, PlatoNeRF is better at smoothly reconstructing scenes where shadows are hard to see, such as those with high ambient light or dark backgrounds.

In addition to improving the safety of autonomous vehicles, PlatoNeRF could make AR/VR headsets more efficient by enabling a user to model the geometry of a room without the need to walk around taking measurements. It could also help warehouse robots find items in cluttered environments faster.

"Our key idea was taking these two things that have been done in

different disciplines before and pulling them together—multibounce lidar and machine learning. It turns out that when you bring these two together, that is when you find a lot of new opportunities to explore and get the best of both worlds," says Tzofi Klinghoffer, an MIT graduate student in media arts and sciences, affiliate of the MIT Media Lab, and lead author of the paper on PlatoNeRF.

Klinghoffer wrote the paper with his advisor, Ramesh Raskar, associate professor of media arts and sciences and leader of the Camera Culture Group at MIT; senior author Rakesh Ranjan, a director of AI research at Meta Reality Labs; as well as Siddharth Somasundaram at MIT, and Xiaoyu Xiang, Yuchen Fan, and Christian Richardt at Meta. The research is being presented at the [Conference on Computer Vision and Pattern Recognition](#), held 17–21 June.

Shedding light on the problem

Reconstructing a full 3D scene from one camera viewpoint is a complex problem.

Some machine-learning approaches employ generative AI models that try to guess what lies in the occluded regions, but these models can hallucinate objects that aren't really there. Other approaches attempt to infer the shapes of hidden objects using shadows in a color image, but these methods can struggle when shadows are hard to see.

For PlatoNeRF, the MIT researchers built off these approaches using a new sensing modality called single-photon lidar. Lidars map a 3D scene by emitting pulses of light and measuring the time it takes that light to bounce back to the sensor. Because single-photon lidars can detect individual photons, they provide higher-resolution data.

The researchers use a single-photon lidar to illuminate a target point in

the scene. Some light bounces off that point and returns directly to the sensor. However, most of the light scatters and bounces off other objects before returning to the sensor. PlatoNeRF relies on these second bounces of light.

By calculating how long it takes light to bounce twice and then return to the lidar sensor, PlatoNeRF captures additional information about the scene, including depth. The second bounce of light also contains information about shadows.

The system traces the secondary rays of light—those that bounce off the target point to other points in the scene—to determine which points lie in shadow (due to an absence of light). Based on the location of these shadows, PlatoNeRF can infer the geometry of hidden objects.

The lidar sequentially illuminates 16 points, capturing multiple images that are used to reconstruct the entire 3D scene.

"Every time we illuminate a point in the scene, we are creating new shadows. Because we have all these different illumination sources, we have a lot of light rays shooting around, so we are carving out the region that is occluded and lies beyond the visible eye," Klinghoffer says.

A winning combination

Key to PlatoNeRF is the combination of multibounce lidar with a special type of machine-learning model known as a neural radiance field (NeRF). A NeRF encodes the geometry of a scene into the weights of a neural network, which gives the model a strong ability to interpolate, or estimate, novel views of a scene.

This ability to interpolate also leads to highly accurate scene reconstructions when combined with multibounce lidar, Klinghoffer

says.

"The biggest challenge was figuring out how to combine these two things. We really had to think about the physics of how light is transporting with multibounce lidar and how to model that with machine learning," he says.

They compared PlatoNeRF to two common alternative methods, one that only uses lidar and the other that only uses a NeRF with a color image.

They found that their method was able to outperform both techniques, especially when the lidar sensor had lower resolution. This would make their approach more practical to deploy in the real world, where lower resolution sensors are common in commercial devices.

"About 15 years ago, our group invented the first camera to 'see' around corners, that works by exploiting multiple bounces of light, or 'echoes of light.' Those techniques used special lasers and sensors, and used three bounces of light. Since then, lidar technology has become more mainstream, that led to our research on cameras that can see through fog," Raskar says.

"This new work uses only two bounces of light, which means the signal to noise ratio is very high, and 3D reconstruction quality is impressive."

In the future, the researchers want to try tracking more than two bounces of light to see how that could improve [scene](#) reconstructions. In addition, they are interested in applying more deep learning techniques and combining PlatoNeRF with [color image](#) measurements to capture texture information.

"While camera images of shadows have long been studied as a means to

3D reconstruction, this work revisits the problem in the context of lidar, demonstrating significant improvements in the accuracy of reconstructed hidden geometry. The work shows how clever algorithms can enable extraordinary capabilities when combined with ordinary sensors—including the lidar systems that many of us now carry in our pocket," says David Lindell, an assistant professor in the Department of Computer Science at the University of Toronto, who was not involved with this work.

More information: PlatoNeRF: [3D Reconstruction in Plato's Cave via Single-View Two-Bounce Lidar](#)

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: Researchers leverage shadows to model 3D scenes, including objects blocked from view (2024, June 18) retrieved 29 June 2024 from <https://techxplore.com/news/2024-06-leverage-shadows-3d-scenes-blocked.html>

<p>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.</p>
--