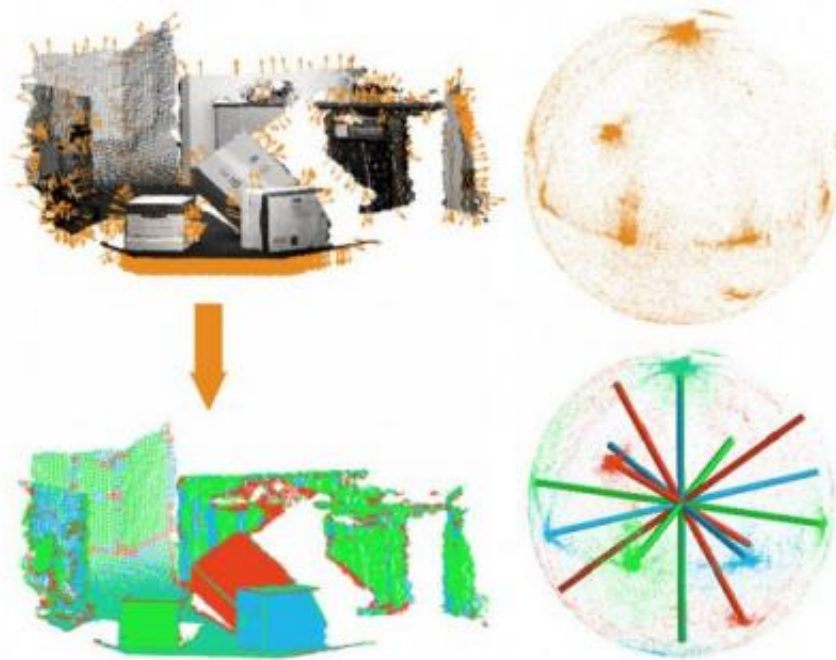


Algorithm for determining orientation of objects could aid robots in navigation, scene understanding

April 4 2014, by Larry Hardesty



The algorithm's first step is to estimate the orientations of individual points in the scene (orange arrows), which it maps onto the surface of a sphere (orange clusters). Through an iterative process, it finds the set of axes that best fit the point clusters (red, blue, and green columns), which it re-identifies with the points in the scene.

Suppose you're trying to navigate an unfamiliar section of a big city, and

you're using a particular cluster of skyscrapers as a reference point. Traffic and one-way streets force you to take some odd turns, and for a while you lose sight of your landmarks. When they reappear, in order to use them for navigation, you have to be able to identify them as the same buildings you were tracking before—as well as your orientation relative to them.

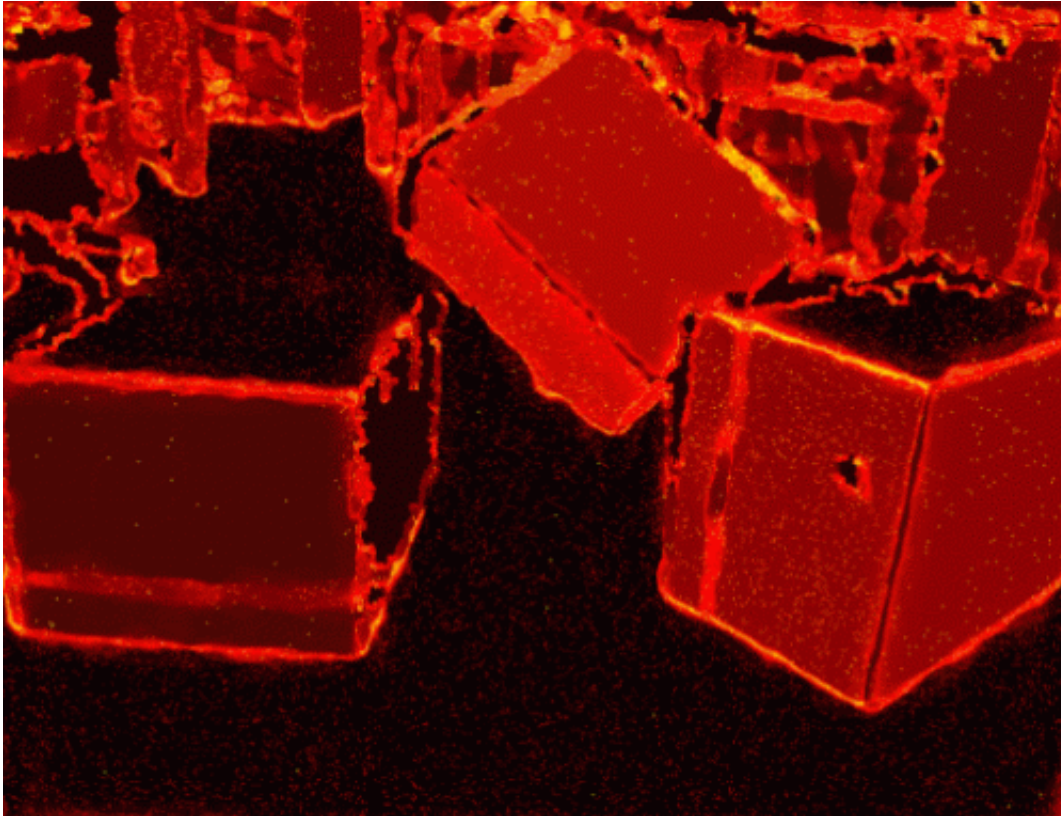
That type of re-identification is second nature for humans, but it's difficult for computers. At the IEEE Conference on Computer Vision and Pattern Recognition in June, MIT researchers will present a new [algorithm](#) that could make it much easier, by identifying the major orientations in 3-D scenes. The same algorithm could also simplify the problem of scene understanding, one of the central challenges in computer vision research.

The algorithm is primarily intended to aid robots navigating unfamiliar buildings, not motorists navigating unfamiliar cities, but the principle is the same. It works by identifying the dominant orientations in a given scene, which it represents as sets of axes—called "Manhattan frames"—embedded in a sphere. As a [robot](#) moved, it would, in effect, observe the sphere rotating in the opposite direction, and could gauge its orientation relative to the axes. Whenever it wanted to reorient itself, it would know which of its landmarks' faces should be toward it, making them much easier to identify.

As it turns out, the same algorithm also drastically simplifies the problem of plane segmentation, or deciding which elements of a visual scene lie in which planes, at what depth. Plane segmentation allows a computer to build boxy 3-D models of the objects in the scene—which it could, in turn, match to stored 3-D models of known objects.

Julian Straub, a graduate student in electrical engineering and computer science at MIT, is lead author on the paper. He's joined by his advisors,

John Fisher, a senior research scientist in MIT's Computer Science and Artificial Intelligence Laboratory, and John Leonard, a professor of mechanical and ocean engineering, as well as Oren Freifeld and Guy Rosman, both postdocs in Fisher's Sensing, Learning, and Inference Group.



The researchers' new algorithm takes depth information (red) about a visual scene and determines the orientation of the objects depicted (red, blue, and green). That makes the problem of plane segmentation — deciding which elements of the scene lie in which planes, at what depth — much simpler (multiple colors).

The researchers' new algorithm works on 3-D data of the type captured by the Microsoft Kinect or laser rangefinders. First, using established

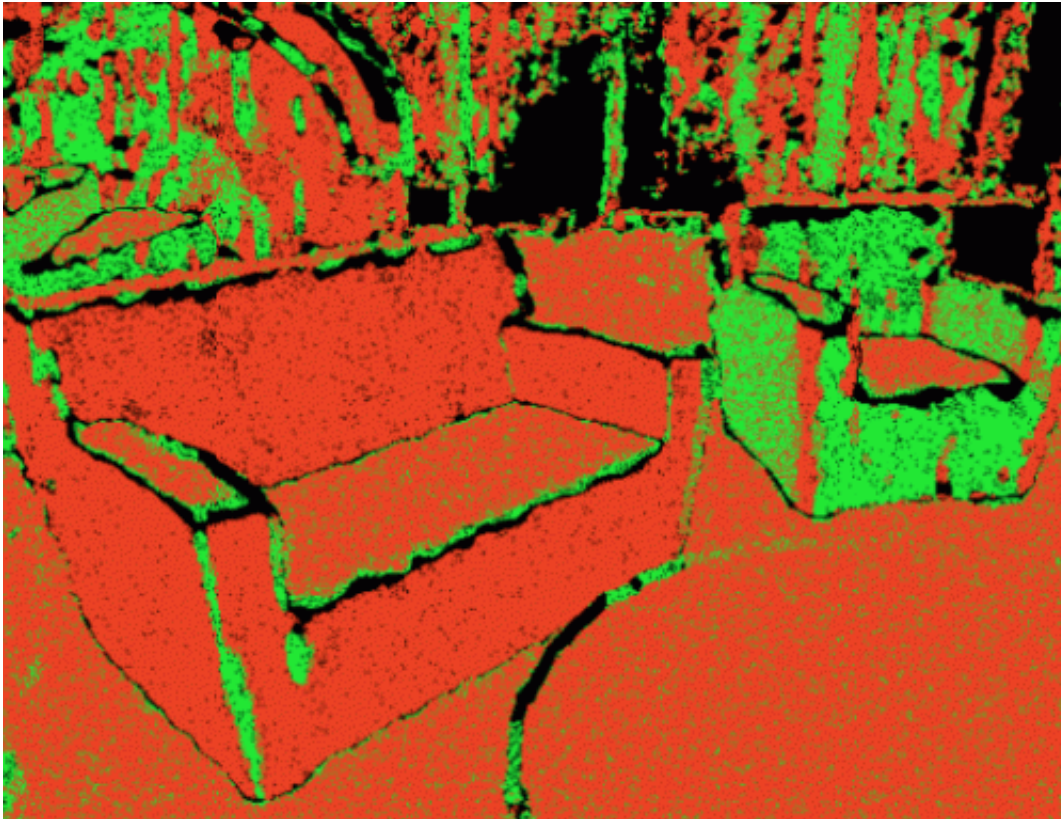
procedures, the algorithm estimates the orientations of a large number of individual points in the scene. Those orientations are then represented as points on the surface of a sphere, with each point defining a unique angle relative to the sphere's center.

Since the initial orientation estimate is rough, the points on the sphere form loose clusters that can be difficult to distinguish. Using statistical information about the uncertainty of the initial orientation estimates, the algorithm then tries to fit Manhattan frames to the points on the sphere.

The basic idea is similar to that of regression analysis—finding lines that best approximate scatters of points. But it's complicated by the geometry of the sphere. "Most of classical statistics is based on linearity and Euclidean distances, so you can take two points, you can sum them, divide by two, and this will give you the average," Freifeld says. "But once you are working in spaces that are nonlinear, when you do this averaging, you can fall outside the space."

Consider, for instance, the example of measuring geographical distances. "Say that you're in Tokyo and I'm in New York," Freifeld says. "We don't want our average to be in the middle of the Earth; we want it to be on the surface." One of the keys to the new algorithm is the fact it incorporates these geometries into the statistical reasoning about the scene.

In principle, it would be possible to approximate the point data very accurately by using hundreds of different Manhattan frames, but that would yield a model that's much too complex to be useful. So another aspect of the algorithm is a cost function that weighs accuracy of approximation against number of frames. The algorithm starts with a fixed number of frames—somewhere between three and 10, depending on the expected complexity of the scene—and then tries to pare that number down without compromising the overall cost function.



The researchers' new algorithm takes depth information (red) about a visual scene and determines the orientation of the objects depicted (red, blue, and green). That makes the problem of plane segmentation — deciding which elements of the scene lie in which planes, at what depth — much simpler (multiple colors).

The resulting set of Manhattan frames may not represent subtle distinctions between objects that are slightly misaligned with each other, but those distinctions aren't terribly useful to a navigation system. "Think about how you navigate a room," Fisher says. "You're not building a precise model of your environment. You're sort of capturing loose statistics that allow you to complete your task in a way that you don't stumble over a chair or something like that."

Once a set of Manhattan frames has been determined, the problem of plane segmentation becomes much easier. Objects that don't take up much of the visual field—because they're small, distant, or occluded—make trouble for existing plane segmentation algorithms, because they yield so little depth information that their orientations can't be reliably inferred. But if the problem is one of selecting among just a handful of possible orientations, rather than a potential infinitude, it becomes much more tractable.

Frank Dellaert, a professor of interactive computing at Georgia Tech who was not involved in this research, calls the work "interesting," adding that it "generalizes to non-vertical frames, which is important in a manipulation context, and it works with depth images, which have become very popular with the rise of Kinect and other depth sensors."

"I believe that techniques such as these should be applied," Dellaert says. "Whether they will or not depends on how much perceived added value is seen by companies that will eventually implement and deploy these techniques in mass-produced autonomous systems, be they household robots or self-driving cars. There is always a trade-off between algorithm generality and a certain amount of complexity to exploit constraints such as manmade structures. In the present case, I believe the advantages of using these constraints are significant and will be exploited in future autonomous systems."

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: Algorithm for determining orientation of objects could aid robots in navigation, scene understanding (2014, April 4) retrieved 25 April 2024 from

<https://phys.org/news/2014-04-algorithm-aid-robots-scene.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.