

# Tool transforms world landmark photos into 4-D experiences

26 August 2020, by Melanie Lefkowitz



Credit: CC0 Public Domain

Using publicly available tourist photos of world landmarks such as the Trevi Fountain in Rome or Top of the Rock in New York City, Cornell researchers have developed a method to create maneuverable 3-D images that show changes in appearance over time.

The method, which employs deep learning to ingest and synthesize tens of thousands of mostly untagged and undated photos, solves a problem that has eluded experts in computer vision for six decades.

"It's a new way of modeling scenes that not only allows you to move your head and see, say, the fountain from different viewpoints, but also gives you controls for changing the time," said Noah Snavely, associate professor of computer science at Cornell Tech and senior author of "Crowdsampling the Plenoptic Function," presented at the European Conference on Computer Vision, held virtually Aug. 23-28.

"If you really went to the Trevi Fountain on your vacation, the way it would look would depend on what time you went—at night, it would be lit up by

floodlights from the bottom. In the afternoon, it would be sunlit, unless you went on a cloudy day," Snavely said. "We learned the whole range of appearances, based on time of day and weather, from these unorganized photo collections, such that you can explore the whole range and simultaneously move around the scene."

Representing a place in a photorealistic way is challenging for traditional computer vision, partly because of the sheer number of textures to be reproduced. "The real world is so diverse in its appearance and has different kinds of materials—shiny things, water, thin structures," Snavely said.

Another problem is the inconsistency of the available data. Describing how something looks from every possible viewpoint in space and time—known as the plenoptic function—would be a manageable task with hundreds of webcams affixed around a scene, recording data day and night. But since this isn't practical, the researchers had to develop a way to compensate.

"There may not be a photo taken at 4 p.m. from this exact viewpoint in the data set. So we have to learn from a photo taken at 9 p.m. at one location, and a photo taken at 4:03 from another location," Snavely said. "And we don't know the granularity of when these photos were taken. But using [deep learning](#) allows us to infer what the scene would have looked like at any given time and place."

The researchers introduced a new scene representation called Deep Multiplane Images to interpolate appearance in four dimensions—3-D, plus changes over time. Their method is inspired in part on a classic animation technique developed by the Walt Disney Company in the 1930s, which uses layers of transparencies to create a 3-D effect without redrawing every aspect of a scene.

"We use the same idea invented for creating 3-D

effects in 2-D animation to create 3-D effects in [real-world](#) scenes, to create this deep multilayer image by fitting it to all these disparate measurements from the tourists' photos," Snavely said. "It's interesting that it kind of stems from this very old, classic technique used in animation."

In the study, they showed that this model could be trained to create a [scene](#) using around 50,000 publicly available images found on sites such as Flickr and Instagram. The method has implications for [computer vision](#) research, as well as virtual tourism—particularly useful at a time when few can travel in person.

"You can get the sense of really being there," Snavely said. "It works surprisingly well for a range of scenes."

**More information:** Crowdsampling the Plenoptic Function. [research.cs.cornell.edu/crowdplenoptic/](https://research.cs.cornell.edu/crowdplenoptic/)

Provided by Cornell University

APA citation: Tool transforms world landmark photos into 4-D experiences (2020, August 26) retrieved 28 May 2022 from <https://techxplore.com/news/2020-08-tool-world-landmark-photos-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.*