

Stanford site advances science of turning 2-D images into 3-D models



A three-dimensional 'fly around' image, above, was created from a two-dimensional image using an algorithm developed by Stanford computer scientists. Credit: Ashutosh Saxena

An artist might spend weeks fretting over questions of depth, scale and perspective in a landscape painting, but once it is done, what's left is a two-dimensional image with a fixed point of view. But the Make3d algorithm, developed by Stanford computer scientists, can take any two-dimensional image and create a three-dimensional "fly around" model of its content, giving viewers access to the scene's depth and a range of points of view.

"The algorithm uses a variety of visual cues that humans use for estimating the 3-D aspects of a scene," said Ashutosh Saxena, a doctoral student in computer science who developed the Make3d website with Andrew Ng, an assistant professor of computer science. "If we look at a grass field, we can see that the texture changes in a particular way as it becomes more distant."

The algorithm runs at <http://make3d.stanford.edu> .

The applications of extracting 3-D models from 2-D images, the researchers say, could range from enhanced pictures for online real estate sites to quickly creating environments for video games and improving the vision and dexterity of mobile robots as they navigate through the spatial world.

Extracting 3-D information from still images is an emerging class of technology. In the past, some researchers have synthesized 3-D models by analyzing multiple images of a scene. Others, including Ng and Saxena in 2005, have developed algorithms that infer depth from single images by combining assumptions about what must be ground or sky with simple cues such as vertical lines in the image that represent walls or trees. But Make3d creates accurate and smooth models about twice as often as competing approaches, Ng said, by abandoning limiting assumptions in favor of a new, deeper analysis of each image and the powerful artificial intelligence technique "machine learning."

Restoring the third dimension

To "teach" the algorithm about depth, orientation and position in 2-D images, the researchers fed it still images of campus scenes along with 3-D data of the same scenes gathered with laser scanners. The algorithm correlated the two sets together, eventually gaining a good idea of the trends and patterns associated with being near or far. For example, it learned that abrupt changes along edges correlate well with one object occluding another, and it saw that things that are far away can be just a little hazier and more bluish than things that are close.

To make these judgments, the algorithm breaks the image up into tiny planes called "superpixels," which

are within the image and have very uniform color, brightness and other attributes. By looking at a superpixel in concert with its neighbors, analyzing changes such as gradations of texture, the algorithm makes a judgment about how far it is from the viewer and what its orientation in space is. Unlike some previous algorithms, the Stanford one can account for planes at any angle, not just horizontal or vertical. This allows it to create models for scenes that have planes at many orientations, such as the curved branches of trees or the slopes of mountains.

A paper on the algorithm by Ng, Saxena and a fellow student, Min Sun, won the best paper award at the 3-D recognition and reconstruction workshop at the International Conference on Computer Vision in Rio de Janeiro in October 2007.

On the Make3d website, the algorithm puts images uploaded by users into a processing queue and will send an e-mail when the model has been rendered. Users can then vote on whether the model looks good, and can see an alternative rendering and even tinker with the model to fix what might not have been rendered right the first time.

Photos can be uploaded directly or pulled into the site from the popular photo-sharing site Flickr.

Although the technology works better than any other has so far, Ng said, it is not perfect. The software is at its best with landscapes and scenery rather than close-ups of individual objects. Also, he and Saxena hope to improve it by introducing object recognition. The idea is that if the software can recognize a human form in a photo it can make more accurate distance judgments based on the size of the person in the photo.

For many panoramic scenes, there is still no substitute for being there. But when flat photos become 3-D, viewers can feel a little closer—or farther.

Source: By David Orenstein, Stanford University

This document is subject to copyright. Apart from any fair dealing for the purpose of private study, research, no part may be reproduced without the written permission. The content is provided for information purposes only.