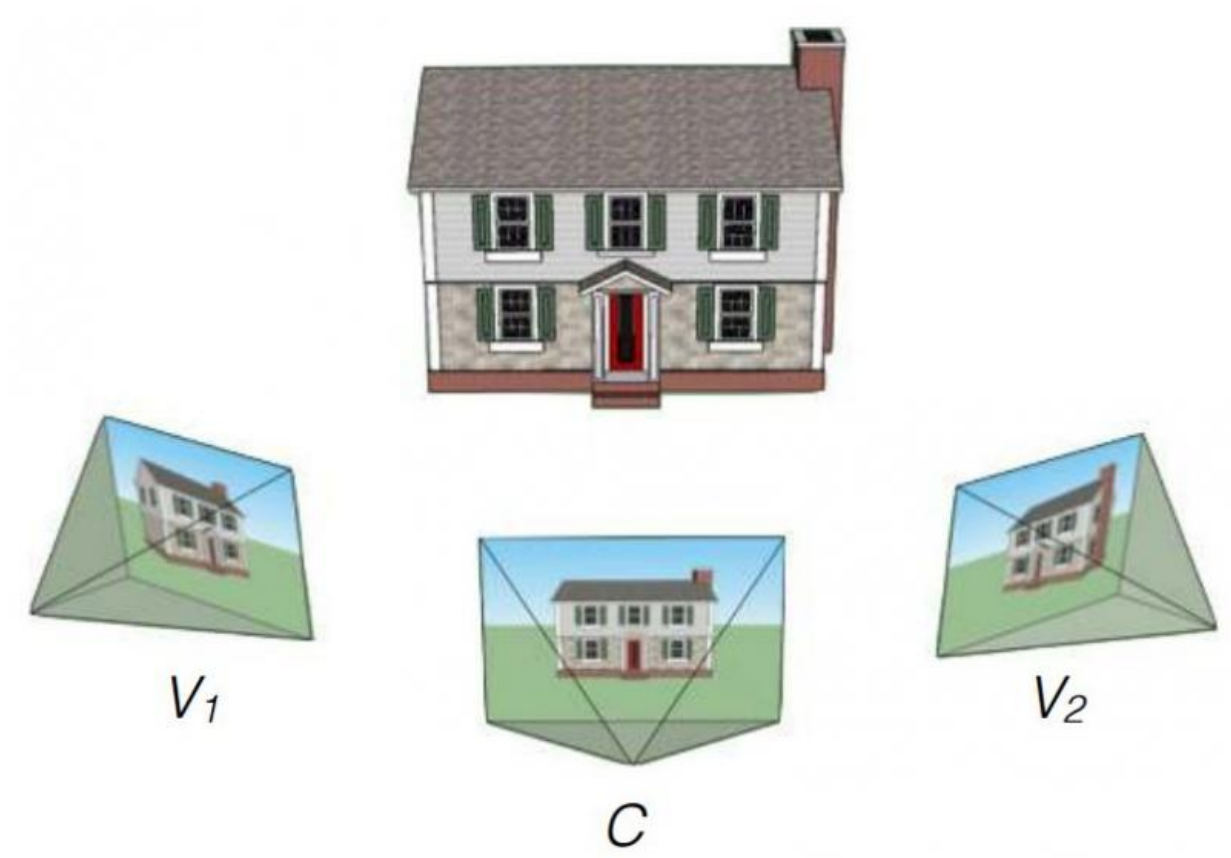
DeepStereo: Google quartet has method for new-view synthesis

July 9 2015, by Nancy Owano



Credit: John Flynn et al. arXiv:1506.06825 [cs.CV]

Four researchers from Google have been turning to deep networks—not for classification purposes in computer vision but this time for

application in graphics problems. Their work has shown interesting results, as evidenced in their paper, now on the arXiv server, titled "DeepStereo: Learning to Predict New Views from the World's Imagery."

The authors are John Flynn, Ivan Neulander, James Philbin and Noah Snavely. In brief, they have achieved as way to synthesize real-world images. As *MIT Technology Review*'s summary account of their work said, "Give Google's DeepStereo algorithm two images of a scene and it will synthesize a third image from a different point of view."

This was their research interest. Can one get a new-view synthesis directly from pixels? Their "deep architecture" involved end-to-end training from a large number of image sets.

(*MIT Technology Review*, explaining their research, said, "The task for the computer is to treat each image as a set of pixels and to determine the depth and color of each pixel given the depth and color of the corresponding pixels in the images that will appear before and after it in the movie.")

The team also has a video about their results; it shows movies made from Street View data. Source frames used in the model, shown in the upper right in the video, used 96 depth planes. They said their method "can convincingly reproduce known test views from nearby imagery."

"To our knowledge," they stated, "our work is the first to apply deep learning to the problem of new view synthesis from sets of real-world, natural imagery."

Think of it as image interpolation. Martin Anderson, editor, *The Stack*, noted, " By using deep networks to generate 'missing' frames from Google Street view, a team led by Google researcher and former visual

effects wizard John Flynn have discovered a technique, dubbed 'DeepStereo,' that can turn the staccato images of Google Maps' Street View into what appears to be genuine video footage."

Richard Chirgwin in *The Register* commented that "StreetView means Google owns one of the world's larger photo albums, so it's natural for Google to want to create a realistic 3D rendering of the world."

If, say, the whole idea were to give Google's viewers a motion-picture experience with StreetView images, they why not just play an image sequence from Street View images to create one movie? Not so practical when you want to see, and experience, something like an art show. *MIT Technology Review* explained: "Running these images at 25 frames per second or thereabouts makes the scenery run ridiculously quickly. That may be acceptable when the scenery does not change, perhaps along freeways and motorways or through unchanging landscapes. But it is entirely unacceptable for busy street views or inside an art gallery."

The alternative solution, to add additional frames between the ones recorded by the Street View cameras, posed another challenge, in what the frames should look like.

Enter the researchers from Google, who worked out what these missing frames should look like by studying the frames on either side—"a computational movie machine," said *MIT Technology Review*, designed for interpolating missing frames.

Discussing their training method, the paper's authors said they used images of street scenes captured by a moving vehicle. "The images were posed using a combination of odometry and traditional structure-from-motion techniques. The vehicle captures a set of images, known as a rosette, from different directions for each exposure. The capturing camera uses a rolling shutter sensor, which is taken into account by our

camera model. We used approximately 100K of such image sets during training."

The authors also discussed where they want to take their research from here. They said their method currently needs "reprojecting each input image to a set of depth planes; we currently use 96 depth planes, which limits the resolution of the output images that we can produce."

Increasing the resolution would call for a larger number of depth planes, "which would mean that the network takes longer to train, uses more RAM and takes longer to run. This is a drawback shared with other volumetric stereo methods; however, our method requires reprojected images per rendered frame, rather than just once when creating the scene. We plan to explore pre-computing parts of the network and warping to new views before running the final layers."

More information: DeepStereo: Learning to Predict New Views from the World's Imagery, arXiv:1506.06825 [cs.CV] arxiv.org/abs/1506.06825

Abstract
Deep networks have recently enjoyed enormous success when applied to recognition and classification problems in computer vision, but their use in graphics problems has been limited. In this work, we present a novel deep architecture that performs new view synthesis directly from pixels, trained from a large number of posed image sets. In contrast to traditional approaches which consist of multiple complex stages of processing, each of which require careful tuning and can fail in unexpected ways, our system is trained end-to-end. The pixels from neighboring views of a scene are presented to the network which then directly produces the pixels of the unseen view. The benefits of our approach include generality (we only require posed image sets and can easily apply our method to different domains), and high quality results

on traditionally difficult scenes. We believe this is due to the end-to-end nature of our system which is able to plausibly generate pixels according to color, depth, and texture priors learnt automatically from the training data. To verify our method we show that it can convincingly reproduce known test views from nearby imagery. Additionally we show images rendered from novel viewpoints. To our knowledge, our work is the first to apply deep learning to the problem of new view synthesis from sets of real-world, natural imagery.

© 2015 Tech Xplore

Citation: DeepStereo: Google quartet has method for new-view synthesis (2015, July 9) retrieved 2 May 2024 from
https://techxplore.com/news/2015-07-deepstereo-google-quartet-method-new-view.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.