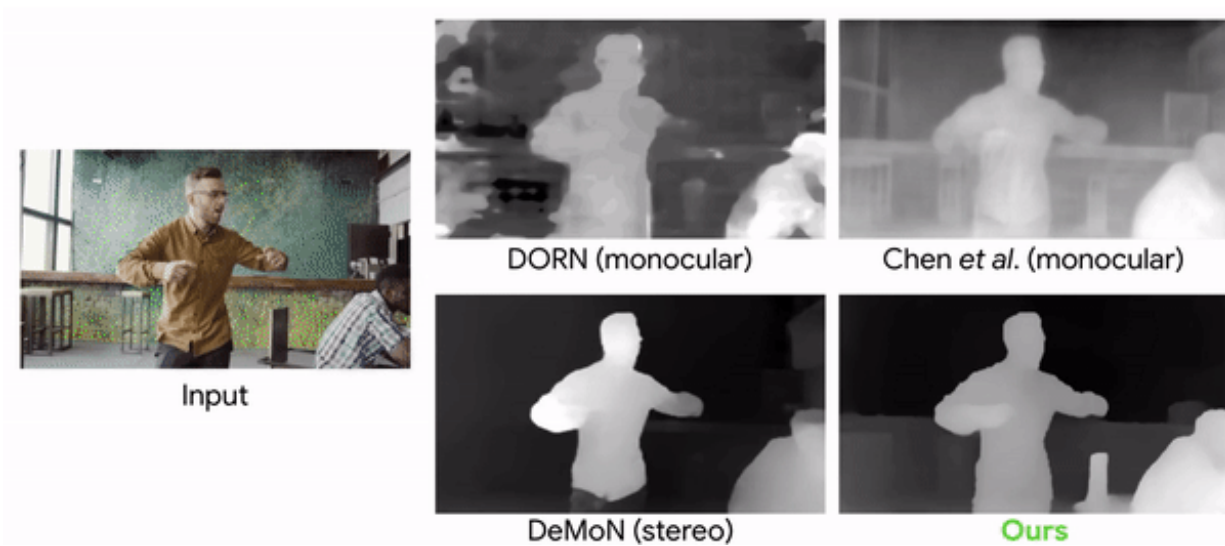


Researchers step back to mannequin viral wave to explore depth

June 29 2019, by Nancy Cohen



Comparison of depth prediction models to a video clip with moving cameras and people. Credit: Google

Who said the viral craze called Mannequin Challenge (MC) is done and dusted? Not so. Researchers have turned to the Challenge that won attention in 2016 to serve their goal. They used the MC for training a neural network that can reconstruct depth information from the videos.

"Learning the Depths of Moving People by Watching Frozen People" is the name of their paper, now up on arXiv, authored by Zhengqi Li, Tali Dekel, Forrester Cole, Richard Tucker, Noah Snavely, Ce Liu and

William Freeman. The paper was submitted in April this year.

The Mannequin Challenge? Who can forget? This was a YouTube trend gone viral. Anthony Alford in *InfoQ* brought readers back to 2016, when an internet meme had people teamed in groups impersonate mannequins. They were "frozen" but a videographer would make [moves](#) around the scene taking a [video](#) from different angles.

Alford wrote, because the camera is moving and the rest of the scene is static, parallax methods can easily reconstruct accurate [depth](#) maps of human figures in a variety of poses.

As the authors stated, the videos involved freezing in diverse, natural poses, while a hand-held camera toured the scene.

For training the [neural network](#), the team converted 2,000 of the videos into 2-D images with high-resolution depth data.

Alford said that out of the 2,000 YouTube MC videos, a dataset was produced of 4,690 sequences with a total of more than 170K valid image-depth pairs. The target of the learning system was the known depth map for the input image, computed from the MC videos. The DNN learned to take the input image, initial depth map, and human mask, and output a "refined" depth map where the depth values of humans were filled in.

Christine Fisher, *Engadget*: "To train the neural network, the researchers converted the clips into 2-D images, estimated the camera pose and created depth maps. The AI was then able to predict the depth of moving objects in [videos](#) with higher accuracy than previously possible."

Taking up the challenge was described by two of the paper's co-authors back in May in a Google blog.

"Because the entire scene is stationary (only the camera is moving), triangulation-based methods—like multi-view-stereo (MVS)—work, and we can get accurate depth maps for the entire scene including the people in it. We gathered approximately 2000 such videos, spanning a wide range of realistic scenes with people naturally posing in different group configurations." Tali Dekel, research scientist and Forrester Cole, [software engineer](#), machine perception, wrote more about the challenge they took on.

"The human [visual](#) system has a remarkable ability to make sense of our 3-D world from its 2-D projection. Even in complex environments with multiple moving objects, people are able to maintain a feasible interpretation of the objects' geometry and depth ordering. The field of computer vision has long studied how to achieve similar capabilities by computationally reconstructing a scene's geometry from 2-D image data, but robust reconstruction remains difficult in many cases."

Why this matters: "While there is a recent surge in using machine learning for depth prediction, this work is the first to tailor a learning-based approach to the case of simultaneous camera and human motion," they said in the May blog. "In this work, we focus specifically on humans because they are an interesting target for augmented reality and 3-D video effects."

Talking about results, Karen Hao, *MIT Technology Review*, said the researchers converted 2,000 of the videos into 2-D images with high-resolution depth data and used them to train a neural network. It was then able to predict the depth of moving objects in a video at much higher accuracy than was possible with previous state-of-the-art [methods](#).

More information: Learning the Depths of Moving People by Watching Frozen People, arXiv:1904.11111 [cs.CV]

arxiv.org/abs/1904.11111

Moving Camera, Moving People: A Deep Learning Approach to Depth Prediction: ai.googleblog.com/2019/05/moving-camera-moving-people-a-deep-learning-approach-to-depth-prediction.html

© 2019 Science X Network

Citation: Researchers step back to mannequin viral wave to explore depth (2019, June 29)
retrieved 19 September 2024 from
<https://techxplore.com/news/2019-06-mannequin-viral-explore-depth.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.