

New machine-learning approach brings digital photos back to life

4 May 2021, by Stephanie Jones



Ours focused on foreground

Ours focused on background

The single image view synthesis process can also be used to generate refocused images (shown above).
Credit: Nima Kalantari

Every day, billions of photos and videos are posted to various social media applications. The problem with standard images taken by a smartphone or digital camera is that they only capture a scene from a specific point of view. But looking at it in reality, we can move around and observe it from different viewpoints. Computer scientists are working to provide an immersive experience for the users that would allow them to observe a scene from different viewpoints, but it requires specialized camera equipment that is not readily accessible to the average person.

To make the process easier, Dr. Nima Kalantari, professor in the Department of Computer Science and Engineering at Texas A&M University, and graduate student Qinbo Li have developed a machine-learning-based approach that would allow users to take a single photo and use it to generate novel views of the [scene](#).

"The benefit of our approach is that now we are not limited to capturing a scene in a particular way," said Kalantari. "We can download and use any image on the internet, even ones that are 100

years old, and essentially bring it back to life and look at it from different angles."

Further details about their work were published in the journal *Association for Computing Machinery Transactions on Graphics*.

View synthesis is the process of generating novel views of an object or scene using images taken from given points of view. To create novel view images, information related to the distance between the objects in the scene is used to create a synthetic photo taken from a virtual camera placed at different points within the scene.

Over the past few decades, several approaches have been developed to synthesize these novel view images, but many of them require the user to manually capture multiple photos of the same scene from different viewpoints simultaneously with specific configurations and hardware, which is difficult and time-consuming. However, these approaches were not designed to generate novel view images from a single input image. To simplify the process, the researchers have proposed doing the same process but with just one image.

"When you have multiple images, you can estimate the location of objects in the scene through a process called triangulation," said Kalantari. "That means you can tell, for example, that there's a person in front of the camera with a house behind them, and then mountains in the background. This is extremely important for view synthesis. But when you have a single image, all of that information has to be inferred from that one image, which is challenging."

With the recent rise of deep learning, which is a subfield of machine learning where [artificial neural networks](#) learn from large amounts of data to solve complex problems, the problem of single image view synthesis has garnered considerable attention. Despite this approach being more

accessible for the user, it is a challenging application for the system to handle because there is not enough information to estimate the location of the objects in the scene.

To train a deep-learning network to generate a novel view based on a single input image, they showed it a large set of images and their corresponding novel view images. Although it is an arduous process, the network learns how to handle it over time. An essential aspect of this approach is to model the input scene to make the training process more straightforward for the network to run. But in their initial experiments, Kalantari and Li did not have a way to do this.

"We realized that scene representation is critically important to effectively train the network," said Kalantari.

To make the training process more manageable, the researchers converted the input image into a multiplane image, which is a type of layered 3D representation. First, they broke down the image into planes at different depths according to the objects in the scene. Then, to generate a photo of the scene from a new viewpoint, they moved the planes in front of each other in a specific way and combined them. Using this representation, the network learns to infer the location of the objects in the scene.

To effectively train the network, Kalantari and Li introduced it to a dataset of over 2,000 unique scenes that contained various objects. They demonstrated that their approach could produce high-quality novel view images of a variety of scenes that are better than previous state-of-the-art methods.

The researchers are currently working on extending their approach to synthesize videos. As videos are essentially a bunch of individual images played rapidly in sequence, they can apply their approach to generate novel views of each of those images independently at different times. But when the newly created video is played back, the picture flickers and is not consistent.

"We are working to improve this aspect of the

approach to make it suitable to generate videos from different viewpoints," said Kalantari.

The single image view synthesis method can also be used to generate refocused images. It could also potentially be used for virtual reality and augmented reality applications such as video games and various software types that allow you to explore a particular visual environment.

More information: Qinbo Li et al. Synthesizing light field from a single image with variable MPI and two network fusion, *ACM Transactions on Graphics* (2020). [DOI: 10.1145/3414685.3417785](https://doi.org/10.1145/3414685.3417785)

Provided by Texas A&M University College of Engineering

APA citation: New machine-learning approach brings digital photos back to life (2021, May 4) retrieved 28 May 2022 from <https://techxplore.com/news/2021-05-machine-learning-approach-digital-photos-life.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.