Recovering 'lost dimensions' of images and video

October 16 2019, by Rob Matheson

A novel model developed at MIT recovers valuable data lost from images and video that have been "collapsed" into lower dimensions. It can, for instance, recreate video from motion-blurred images or from cameras that capture people's movement around corners as vague one-dimensional lines. Credit: Massachusetts Institute of Technology

MIT researchers have developed a model that recovers valuable data lost from images and video that have been "collapsed" into lower dimensions.

The model could be used to recreate video from motion-blurred images, or from new types of cameras that capture a person's movement around corners but only as vague one-dimensional lines. While more testing is needed, the researchers think this approach could someday could be used to convert 2-D medical images into more informative—but more expensive—3-D body scans, which could benefit medical imaging in poorer nations.

"In all these cases, the visual data has one dimension—in time or space—that's completely lost," says Guha Balakrishnan, a postdoc in the Computer Science and Artificial Intelligence Laboratory (CSAIL) and first author on a paper describing the model, which is being presented at next week's International Conference on Computer Vision. "If we recover that lost dimension, it can have a lot of important applications."

Captured visual data often collapses data of multiple dimensions of time and space into one or two dimensions, called "projections." X-rays, for example, collapse three-dimensional data about anatomical structures into a flat image. Or, consider a long-exposure shot of stars moving across the sky: The stars, whose position is changing over time, appear as blurred streaks in the still shot.

Likewise, "corner cameras," recently invented at MIT, detect moving people around corners. These could be useful for, say, firefighters finding people in burning buildings. But the cameras aren't exactly user-friendly. Currently they only produce projections that resemble blurry, squiggly lines, corresponding to a person's trajectory and speed.

The researchers invented a "visual deprojection" model that uses a

neural network to "learn" patterns that match low-dimensional projections to their original high-dimensional images and videos. Given new projections, the model uses what it's learned to recreate all the original data from a projection.

In experiments, the model synthesized accurate video frames showing people walking, by extracting information from single, one-dimensional lines similar to those produced by corner cameras. The model also recovered video frames from single, motion-blurred projections of digits moving around a screen, from the popular Moving MNIST dataset.

Joining Balakrishnan on the paper are: Amy Zhao, a graduate student in the Department of Electrical Engineering and Computer Science (EECS) and CSAIL; EECS professors John Guttag, Fredo Durand, and William T. Freeman; and Adrian Dalca, a faculty member in radiology at Harvard Medical School.

Clues in pixels

The work started as a "cool inversion problem" to recreate movement that causes motion blur in long-exposure photography, Balakrishnan says. In a projection's pixels there exist some clues about the high-dimensional source.

Digital cameras capturing long-exposure shots, for instance, will basically aggregate photons over a period of time on each pixel. In capturing an object's movement over time, the camera will take the average value of the movement-capturing pixels. Then, it applies those average values to corresponding heights and widths of a still image, which creates the signature blurry streaks of the object's trajectory. By calculating some variations in pixel intensity, the movement can theoretically be recreated.

As the researchers realized, that problem is relevant in many areas: X-rays, for instance, capture height, width, and depth information of anatomical structures, but they use a similar pixel-averaging technique to collapse depth into a 2-D image. Corner cameras—invented in 2017 by Freeman, Durand, and other researchers—capture reflected light signals around a hidden scene that carry two-dimensional information about a person's distance from walls and objects. The pixel-averaging technique then collapses that data into a one-dimensional video—basically, measurements of different lengths over time in a single line.

The researchers built a general model, based on a convolutional neural network (CNN)—a machine-learning model that's become a powerhouse for image-processing tasks—that captures clues about any lost dimension in averaged pixels.

Synthesizing signals

In training, the researchers fed the CNN thousands of pairs of projections and their high-dimensional sources, called "signals." The CNN learns pixel patterns in the projections that match those in the signals. Powering the CNN is a framework called a "variational autoencoder," which evaluates how well the CNN outputs match its inputs across some statistical probability. From that, the model learns a "space" of all possible signals that could have produced a given projection. This creates, in essence, a type of blueprint for how to go from a projection to all possible matching signals.

When shown previously unseen projections, the model notes the pixel patterns and follows the blueprints to all possible signals that could have produced that projection. Then, it synthesizes new images that combine all data from the projection and all data from the signal. This recreates the high-dimensional signal.

For one experiment, the researchers collected a dataset of 35 videos of 30 people walking in a specified area. They collapsed all frames into projections that they used to train and test the model. From a hold-out set of six unseen projections, the model accurately recreated 24 frames of the person's gait, down to the position of their legs and the person's size as they walked toward or away from the camera. The model seems to learn, for instance, that pixels that get darker and wider with time likely correspond to a person walking closer to the camera.

"It's almost like magic that we're able to recover this detail," Balakrishnan says.

The researchers didn't test their [model](#) on medical images. But they are now collaborating with Cornell University colleagues to recover 3-D anatomical information from 2-D medical images, such as X-rays, with no added costs—which can enable more detailed medical imaging in poorer nations. Doctors mostly prefer 3-D scans, such as those captured with CT scans, because they contain far more useful medical information. But CT scans are generally difficult and expensive to acquire.

"If we can convert X-rays to CT scans, that would be somewhat game-changing," Balakrishnan says. "You could just take an X-ray and push it through our algorithm and see all the lost information."

  More information: Visual Deprojection: Probabilistic Recovery of Collapsed Dimensions: arXiv:1909.00475 [cs.CV] arxiv.org/abs/1909.00475

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.