Algorithm can determine, with 80 percent accuracy, whether video is running forward or backward

June 19 2014, by Larry Hardesty

Credit: Jose-Luis Olivares/MIT

Einstein's theory of relativity envisions time as a spatial dimension, like height, width, and depth. But unlike those other dimensions, time seems to permit motion in only one direction: forward. This directional asymmetry—the "arrow of time"—is something of a conundrum for theoretical physics. But is it something we can see?

An international group of computer scientists believes that the answer is yes. At the IEEE Conference on Computer Vision and Pattern Recognition this month, they'll present a new algorithm that can, with roughly 80 percent accuracy, determine whether a given snippet of video is playing backward or forward.

"If you see that a clock in a movie is going backward, that requires a high-level understanding of how clocks normally move," says William Freeman, a professor of computer science and engineering at MIT and one of the paper's authors. "But we were interested in whether we could tell the direction of time from low-level cues, just watching the way the world behaves."

By identifying subtle but intrinsic characteristics of visual experience, the research could lead to more realistic graphics in gaming and film. But Freeman says that that wasn't the researchers' primary motivation.

"It's kind of like learning what the structure of the visual world is," he says. "To study shape perception, you might invert a photograph to make everything that's black white, and white black, and then check what you can still see and what you can't. Here we're doing a similar thing, by reversing time, then seeing what it takes to detect that change. We're trying to understand the nature of the temporal signal."

Word perfect

Freeman and his collaborators—his students Donglai Wei and YiChang Shih; Lyndsey Pickup and Andrew Zisserman from Oxford University; Changshui Zhang and Zheng Pan of Tsinghua University; and Bernhard Schölkopf of the Max Planck Institute for Intelligent Systems in Tübingen, Germany—designed candidate algorithms that approached the problem in three different ways. All three algorithms were trained on a set of short videos that had been identified in advance as running either

forward or backward.

The algorithm that performed best begins by dividing a frame of video into a grid of hundreds of thousands of squares; then it divides each of those squares into a smaller, four-by-four grid. For each square in the smaller grid, it determines the direction and distance that clusters of pixels move from one frame to the next.

The algorithm then generates a "dictionary" of roughly 4,000 four-by-four grids, where each square in a grid represents particular directions and degrees of motion. The 4,000-odd "words" in the dictionary are chosen to offer a good approximation of all the grids in the training data. Finally, the algorithm combs through the labeled examples to determine whether particular combinations of "words" tend to indicate forward or backward motion.

Following standard practice in the field, the researchers divided their training data into three sets, sequentially training the algorithm on two of the sets and testing its performance against the third. The algorithm's success rates were 74 percent, 77 percent, and 90 percent.

One vital aspect of the algorithm is that it can identify the specific regions of a frame that it is using to make its judgments. Examining the words that characterize those regions could reveal the types of visual cues that the algorithm is using—and perhaps the types of cues that the human visual system uses as well.

The next-best-performing algorithm was about 70 percent accurate. It was based on the assumption that, in forward-moving video, motion tends to propagate outward rather than contracting inward. In video of a break in pool, for instance, the cue ball is, initially, the only moving object. After it strikes the racked balls, motion begins to appear in a wider and wider radius from the point of contact.

Probable cause

The third algorithm was the least accurate, but it may be the most philosophically interesting. It attempts to offer a statistical definition of the direction of causation.

"There's a research area on causality," Freeman says. "And that's actually really quite important, medically even, because in epidemiology, you can't afford to run the experiment twice, to have people experience this problem and see if they get it and have people do that and see if they don't. But you see things that happen together and you want to figure out: 'Did one cause the other?' There's this whole area of study within statistics on, 'How can you figure out when something did cause something else?' And that relates in an indirect way to this study as well."

Suppose that, in a video, a ball is rolling down a ramp and strikes a bump that briefly launches it into the air. When the video is playing in the forward direction, the sudden change in the ball's trajectory coincides with a visual artifact: the bump. When it's playing in reverse, the ball suddenly leaps for no reason. The researchers were able to model that intuitive distinction as a statistical relationship between a mathematical model of an object's motion and the "noise," or error, in the visual signal.

Unfortunately, the approach works only if the object's motion can be described by a linear equation, and that's rarely the case with motions involving human agency. The algorithm can determine, however, whether the video it's being applied to meets that criterion. And in those cases, its performance is much better.

Provided by Massachusetts Institute of Technology

Citation: Algorithm can determine, with 80 percent accuracy, whether video is running forward or backward (2014, June 19) retrieved 17 April 2024 from https://phys.org/news/2014-06-algorithm-percent-accuracy-video.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.