

Exploiting video game software yields broadcast-quality 3-D video of soccer games in real time

4 November 2015, by Larry Hardesty



Credit: Christine Daniloff/MIT

By exploiting the graphics-rendering software that powers sports video games, researchers at MIT and the Qatar Computing Research Institute (QCRI) have developed a system that automatically converts 2-D video of soccer games into 3-D.

The converted [video](#) can be played back over any 3-D device—a commercial 3-D TV, Google's new Cardboard system, which turns smartphones into 3-D displays, or special-purpose displays such as Oculus Rift.

The researchers presented the new system last week at the Association for Computing Machinery's Multimedia conference.

"Any TV these days is capable of 3-D," says Wojciech Matusik, an associate professor of [electrical engineering](#) and [computer science](#) at MIT and one of the system's co-developers. "There's just no content. So we see that the production of

high-quality content is the main thing that should happen. But sports is very hard. With movies, you have artists who paint the depth map. Here, there is no luxury of hiring 100 artists to do the conversion. This has to happen in real-time."

The system is one result of a collaboration between QCRI and MIT's Computer Science and Artificial Intelligence Laboratory. Joining Matusik on the conference paper are Kiana Calagari, a research associate at QCRI and first author; Alexandre Kaspar, an MIT graduate student in electrical engineering and computer science; Piotr Didyk, who was a postdoc in Matusik's group and is now a researcher at the Max Planck Institute for Informatics; Mohamed Hefeeda, a principal scientist at QCRI; and Mohamed Elgharib, a QCRI postdoc. QCRI also helped fund the project.

Zeroing in

In the past, researchers have tried to develop general-purpose systems for converting 2-D video to 3-D, but they haven't worked very well and have tended to produce odd visual artifacts that detract from the viewing experience.

"Our advantage is that we can develop it for a very specific problem domain," Matusik says. "We are developing a conversion pipeline for a specific sport. We would like to do it at broadcast quality, and we would like to do it in real-time. What we have noticed is that we can leverage video games."

Today's video games generally store very detailed 3-D maps of the virtual environment that the player is navigating. When the player initiates a move, the game adjusts the map accordingly and, on the fly, generates a 2-D projection of the 3-D scene that corresponds to a particular viewing angle.

The MIT and QCRI researchers essentially ran this process in reverse. They set the very realistic Microsoft soccer game "FIFA13" to play over and over again, and used Microsoft's [video-game](#) analysis tool PIX to continuously store screen shots of the action. For each screen shot, they also extracted the corresponding 3-D map.

Using a standard algorithm for gauging the difference between two images, they winnowed out most of the screen shots, keeping just those that best captured the range of possible viewing angles and player configurations that the game presented; the total number of screen shots still ran to the tens of thousands. Then they stored each screen shot and the associated 3-D map in a database.

Jigsaw puzzle

For every frame of 2-D video of an actual soccer game, the system looks for the 10 or so screen shots in the database that best correspond to it. Then it decomposes all those images, looking for the best matches between smaller regions of the video feed and smaller regions of the screen shots. Once it's found those matches, it superimposes the depth information from the screen shots on the corresponding sections of the video feed. Finally, it stitches the pieces back together.

The result is a very convincing 3-D effect, with no visual artifacts. The researchers conducted a user study in which the majority of subjects gave the 3-D effect a rating of 5 ("excellent") on a five-point ("bad" to "excellent") scale; the average score was between 4 ("good") and 5.

Currently, the researchers say, the system takes about a third of a second to process a frame of video. But successive frames could all be processed in parallel, so that the third-of-a-second delay needs to be incurred only once. A broadcast delay of a second or two would probably provide an adequate buffer to permit conversion on the fly. Even so, the researchers are working to bring the conversion time down still further.

"This is a clever use of game content, which leads to better results and easier acquisition of large and diverse reference data," says Hanspeter Pfister, a

professor of computer science at Harvard University. "One of the main insights of the paper is that domain-specific methods are able to yield bigger improvements than more general approaches. This is an important lesson that will have ramifications for other domains."

More information: Gradient-based 2D-to-3D Conversion for Soccer Videos. MM '15 Proceedings of the 23rd Annual ACM Conference on Multimedia Conference. Pages 331-340 [DOI: 10.1145/2733373.2806262](#)

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

Provided by Massachusetts Institute of Technology

APA citation: Exploiting video game software yields broadcast-quality 3-D video of soccer games in real time (2015, November 4) retrieved 25 October 2020 from <https://phys.org/news/2015-11-automatically-d-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.