

New method reduces amount of training data needed for facial performance capture system

October 25 2016

Disney Research has found a way to tailor a facial capture system to the characteristics of a specific actor's expressions while dramatically reducing the time and effort it would normally require.

Instead of exhaustively recording the actor displaying a variety of facial expressions under numerous combinations of lighting conditions and camera viewpoints, the researchers found that they could use a small sample of recordings and then synthetically generate the data necessary to train the system.

This method of generating data enabled them to determine a set of training data that was much smaller than normal - tens to hundreds of times smaller - without affecting facial capture accuracy.

The researchers will present their method Oct. 25 at the International Conference on 3D Vision in Palo Alto, Calif.

"Real-time marker-less facial performance capture has been growing in popularity for film and video game production, thanks to advances in machine learning," said Markus Gross, vice president for Disney Research. "By reducing the amount of facial imagery necessary to train these systems, our team has taken a big step toward increasing the flexibility and efficiency of this approach."

Machine learning techniques have made it possible to rapidly infer facial geometries from video, but this requires exhaustively training a program using a lot of annotated face images.

"It takes a lot of labor to capture not only a full spectrum of [facial expressions](#), but to do so under a variety of lighting conditions and from different camera angles," said Kenny Mitchell, senior research scientist. "Our idea was that if we could strategically capture an actor's expressions under certain conditions, we could synthesize all of the training data for a target scenario and save a lot of time."

The researchers used a multi-camera capture setup to initially record about 70 expressions on the actor's face under uniform [lighting conditions](#). This data is used to create a face rig, a movable, posable model of the actor's face. The face rig is then used to generate the synthetic training data tailored for environmental conditions and camera properties similar to those the filmmakers expect on the actual set, said Martin Kludiny, a post-doctoral associate at Disney Research.

The researchers determined that they could achieve the best accuracy by focusing more of the training data on expressions and changes in illumination, with variations in camera perspective having relatively less impact on the final result.

"Our experimental results showed that the best design strategy can reduce training image counts by one-to-two orders of magnitude and result in proportional computational savings with no visible loss of accuracy," added Steven McDonagh, another key post-doctoral researcher on the team.

Combining creativity and innovation, this research continues Disney's rich legacy of inventing new ways to tell great stories and leveraging technology required to build the future of entertainment.

Provided by Disney Research

Citation: New method reduces amount of training data needed for facial performance capture system (2016, October 25) retrieved 18 April 2024 from <https://phys.org/news/2016-10-method-amount-facial-capture.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.