Q&A With Steve Perlman

Back Door
October 2006 • Vol.6 Issue 10
Page(s) 108 in print issue

Q&A With Steve Perlman

Escaping The Uncanny Valley With Contour Motion Capture

If you noticed something eerie about Tom Hanks' character in "The Polar Express" or wondered why characters in most video games seem so wooden, you've encountered the Uncanny Valley Effect. Coined in 1970 by Masahiro Mori, the term basically describes how hard it is to convince anyone that something artificial looks real. Steve Perlman, who created WebTV and Moxi Digital, is trying to change that. The 45-year-old entrepreneur has created a new start-up: Mova (www.mova.com). Funded over the past four years by his own Rearden companies, researchers at Mova have created a 44-camera tool dubbed Contour that can capture human faces and convert them into a form that computer artists can easily manipulate. Perlman promises it will now be far easier to create realistic human characters, with proper facial and lip movements, in video games and animated movies with this new camera tool. Creating realistic characters is as easy as sponging fluorescent makeup on a person and then capturing the light reflected from the makeup into a 3D mapping system. Lastly, a computer process will convert the set-up points into a 3D wire-frame model that an artist manipulates with simple tools. The system goes on sale for game development and movie production houses by the fourth quarter of this year.

CPU: Where did the original idea for Contour come from?

SP: Contour started out with a very broad goal: develop a facial capture system that could cross the Uncanny Valley. By this time we were about to purchase our three-optical motion-capture system, so we had a pretty good idea of the trajectory of the technology; it was not evolving in the direction of achieving photorealistic faces in a production-efficient environment, so we set out and tried everything. Contour in its current form evolved out of a series of perhaps two dozen insights we gained from a succession of experiments, each getting us part of the way there. I'm afraid there was no single 'aha' moment. It was a long and difficult climb up a steep hill for four years of 'two steps forward, one step back;' one invention built upon another. And even after we got it working, it was so computationally intensive, it was not practical. Literally it would take almost a week to compute a single volumetric frame. After steadily optimizing the software and the hardware, we can now compute a single frame in less than a minute. In time, we will be able to compute Contour frames in real-time. At that point, directing will truly be a holodeck-like experience.

CPU: Can you give us the technical explanation in terms of how Contour works? How, for instance, do you get the depth information by triangulation?

SP: Let's consider two cameras each, say, 30 degrees apart from each other. The two cameras are looking at the same object in space, say, a person's face, but from different angles. Let's say we are trying to figure out the position in 'Z' of a spot on the person's right cheek. We use one camera and see the particular random phosphor pattern that happens to be on that spot on the cheek. Next, we take the second camera and compare every pattern that it sees with the first camera's random pattern, until we finally find the exact same pattern that the first camera saw. Then we triangulate between the two cameras; we measure the angle from each camera to that spot on the right cheek, and, using geometry, we determine the 'Z' of that spot.

In practice it is far more complex than that because the surfaces are not flat and it may very well be the case that the second camera can't see the same spot that the first one sees, there may be noise in the image, one camera may be rotated relative to the other, one image may be brighter than the other because of the angle, and so on. And, of course, there is a very precise calibration process that exactly determines where the cameras are relative to each other, their angle, the exact distortions of the lenses, etc. Then, on top of it, if we were to do what I described above without any optimizations to the code, it would take days to compute one frame.

Contour is a very advanced and sophisticated image processing system; highly optimized for practical use. And we have plans to make it even faster, even more accurate, and even easier to use. What you are seeing today is the tip of the iceberg.

CPU: What's your view on the future of digital entertainment? Will it take over?

SP: This is a pretty big question! The answer to that is, I think that you'll be seeing an increasing overlap between motion pictures and video games to the point where a motion picture as we know it today will be viewed as a purely linear entertainment experience and a video game as we know it today will be viewed as a highly interactive entertainment experience. And, I would expect that most experiences made would fall somewhere in between.

Already you can get DVDs with movies/TV shows where you can choose alternative endings. The first season of "24" has two choices for an ending. And, of course, you can get video games that have very sophisticated cinematics that approach movie-grade quality. These are very simple examples of convergence, but they do illustrate how the two worlds are reaching out to each other. Movies like "The Polar Express," "Monster House," and "Beowulf" are shot very much like video games. And video games like The Godfather are produced increasingly with cinematic values. But there is a difference today: For the movies, the emphasis is on the realism and the quality of the imagery, while the flexibility 3D affords the director is important but secondary. For the video games, the emphasis is on the interactivity, while the realism and image quality is important but secondary.

Once the production technology reaches the point where you don't have to compromise realism in order to have interactivity and vice versa, you can see where we will be at a point where there is really no longer a sharp distinction between video games and motion pictures. Video games will look as photorealistic as movies, and directors will have as much real-time 3D interactivity in creating a movie as a gamer has in controlling a character in a game. Then, the choice of the level of interactivity in a digital production will purely be a creative decision, not a decision driven by technology or cost limitations.

And, of course, that's where Contour comes in. Contour provides photorealistic image quality for movie and video games, but it does less expensively than a conventional production pipeline.

CPU: Besides good facial expressions, what else do you need to conquer the Uncanny Valley?

SP: Realistic motion. Conquering the Uncanny Valley takes both very accurate facial features and very realistic facial motion. An artist can hand paint a still image of a face that we think of as photorealistic and cross the Uncanny Valley, but it is extremely difficult and perhaps impossible to hand-animate a face that crosses the Valley.

CPU: Do you need a lip-sync technology? Is it already there in terms of quality?

SP: Contour has no trouble with lip-sync because we capture lip motion precisely. In fact, it is trivial. We just use a clacker at the start of the take, and we sync the clack sound with the frame where the clacker closes and we are done. Lip-sync is usually a nightmare. There are two issues with lip-sync: The first is just synching up lip motion with sound. That's not as hard as it used to be because there are tools that trigger animation based on phonetic transitions. The second is the biggie: creating realistic lip motion to follow what has been said. This is extremely difficult and expensive to do, and you rarely see it done convincingly and then only at a high cost.

One reason for this is lips are very complex organs that actually unfold as they move. (That's the reason we have lines in our lips. Those are folds where the lips fold up when they are closed. They unfold when they are expanded.) It is extraordinarily difficult to model that behavior in 3D.

CPU: How much processing power do you need to work on the faces?

SP: A lot. It depends on how fast you want results. All you need is a single laptop to run Contour. Right now we are running Contour on 40 custom-designed blade processors each with an Nvidia GPU for acceleration, and that is plenty of procession power for normal production needs.

CPU: Some of this technology has unnatural implications. If you can create more realistic faces now with Contour, could you get people to say things that they didn't say? The ability to make people say things they didn't say may unsettle the average user.

SP: You don't need Contour if your goal is just to make someone say something they didn't say. You can do that with readily available video-editing tools, as I'm sure you've see done on the Web for parodies. I don't know of a practical way you could do this with Contour. With Contour you are limited to capturing data from people that agree to have phosphor makeup put on their face and then being captured at Mova. When actors do performances (whether motion capture or live action) their contracts almost always have limits on how the recorded information may be used. For example, I'm sure Angelina

Jolie's "Beowulf" contract limits the use of her motion-capture data for that particular movie. So, we would not be allowed to use the data we've captured for any other purpose than the purpose listed in the contract. And, frankly, performer concerns are much more mundane than someone changing what they say. They are worried about outtakes or other embarrassing moments getting into the hands of paparazzi. And, that's a concern they have with just still photography, let alone video or motion capture data. With video, you can find video clips on the Web and then rework them if you want to fake a video. With Contour the only capture data we can use is data someone deliberately provides to us, and almost invariably it is with specific contractual limits.

CPU: Is there a reason why users should enthusiastically welcome this technology beyond the fact it will get them better games?

SP: Having photorealistic video games on the new game platforms is definitely one reason users should be excited. Also, you'll see more games that can approach photorealistic because the budgets for photorealistic will be much less.

But they also should be excited about the fact that movies will be coming that have much cooler characters and effects, and there will be more kinds of movies that now will be able to afford photorealistic effects. Right now you can only see photorealistic people in movies in the $200 million budget range like "Pirates of the Caribbean: Dead Man's Chest" or "King Kong." There simply aren't that many movies made each year that are so expensive. So, we'll be making it possible to get those kinds of effects in far more movies each year. And the movies will look better than what can be made today at any cost.

Also, educational videos and documentary videos, like a documentary about pharaohs that you might see on the Discovery or History Channel, now will be able to have far more realistic CG recreations of ancient cities and peoples and even major battles. And, someday, I'm hoping we can bring the cost down of the system to create a 'prosumer' version that someone could set up, say, in their garage, and that runs in real-time so they can make their own CG movies. Essentially it would be Machinima on steroids. We're not there yet, but there's nothing between us and a low-cost system but the steady advance of Moore's Law. ■