



tectural features of the recognized block structure. For example, in a block structure interpreted as a building, the system uses simple pattern-matching to identify such components as the walls and roof. It will then enhance the visual appearance of the model by adding appropriate geometry and surface detail.

Sculpting Another Approach

The modeling-clay application relies on external sensors and computer vision, rather than embedded computing. The system uses a motorized rotary table, a digital camera, a laser striper, and the host computer to scan, recognize, interpret, and animate 3D clay models. Users first model the clay into the desired form, then place it on the rotary table. As the table rotates, the digital camera captures a sequence of images from which the computer creates a volumetric scan. Complex models may be difficult to capture with the digital camera alone, thus the laser striper can be used to refine the model.

To recognize and interpret the clay model, the system compares a set of parameterized object templates to the scanned


model. The templates are deformed to match the model, and the model is classified based on the degree of deformation. Next, an interpretation step parses the model into its constituent parts. For example, if the system recognizes the model as a biped, the deformation classification identifies the head, arms, legs, and torso based on their relation to the best-matched template. The biped can then be animated using traditional techniques.

The building-block and clay-based modeling systems are still in the prototype stage. Top on the ongoing research agenda is the development of alternative architectures for embedded computation. In its current guise, the building-block system requires a huge supply of power to recognize large block structures. The researchers are studying ways to capture a composed structure using the least possible active components in each block. They are also considering architectures that make use of

A clay model gets virtual. A digital camera and computer-vision techniques are used to scan, recognize, and graphically interpret a clay creature, which can then be animated with traditional tools. Shown here, from left: the clay model, its volumetric scan (a combination of serial digital images), its "best match" from a template library, component parts of the interpreted model, and a frame from the resulting animation.

a broadcast medium. And, to enable more user interaction, they are exploring modifications that would allow for incremental adjustment of the physical models.

Finally, they are thinking about "mixed-initiative" systems that take better advantage of the automated algorithms by allowing them to learn from users' histories of choices and make suggestions based on those.

To date, the most important outcome of this work may be the message that a modeling shift is on the horizon. "You shouldn't have to master a hugely complex modeling system to create virtual worlds and the characters that inhabit them," says Marks. "If users are given the ability to easily create and animate complex 3D models, I'm sure developers will have no trouble producing fun and educational applications that have 3D modeling at their core." 

Diana Phillips Mahoney is chief technology editor of Computer Graphics World.

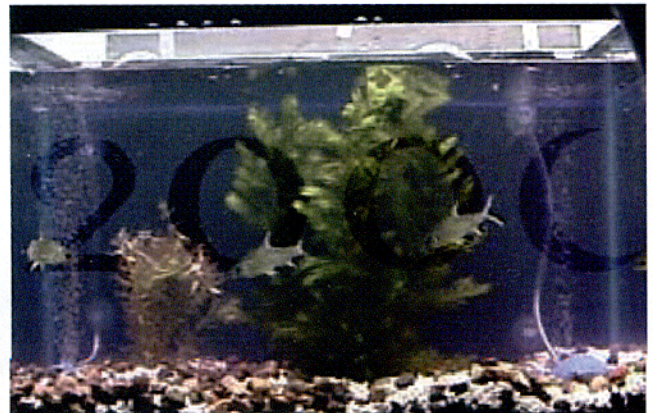
A New Take on Textures

Video-based rendering extends two-dimensional textures into the temporal domain

Somewhere between still images and digital video lies an innovative medium called a video texture that is poised to become invaluable for enhancing everything from Web displays to interactive games and movies.

While still images can portray important information, they fall short when it comes to capturing the essence of inherently dynamic phenomena, such as flickering flames or splashing water. The alternative is to use video, but in order to store video on a computer or other storage device, a clip has to record a specific period of time. So while it may capture the dynamic behavior of the event at hand, it is not able to depict the timeless quality of the phenomena in general.

In contrast, a video texture—so-called because of the similarity to the concept of image textures, which repeat visual patterns over defined periods—is a short, preprocessed video clip that can be



Multiple, independent video clips populate this fish tank. The bubbles, the swaying plants, and the fish are separate elements combined as video textures into one scene.

used as an animation in lieu of a regular, finite-duration video clip. "Video textures loop seamlessly, so the video appears to be playing without end, even though it was originally just a short piece of

video,” says Richard Szeliski of Microsoft Research, who developed the video-texture technology with colleagues Arno Schodl and Irfan Essa at Georgia Tech and David Salesin at the University of Washington. Examples include such visuals as trees blowing in the wind, a campfire burning, waves crashing, and waterfalls. “Such clips can be used to liven up a Web page, or they could be used as a synthetic element in a traditional computer-graphics application.”

The core of the video-texture technology is an algorithm that analyzes a standard video clip to discover where the natural loops occur. “Once the transition points have been discovered and catalogued, we can improve the quality of the resulting animation using tricks such as blending and morphing to disguise the transi-



To interactively guide the path of a fish that has been separated from its video background, the video-texture system selects and orders frames based on specific criteria, such as velocity.

tions,” says Szeliski. “We can also analyze the video by region in cases where several objects are moving independently.”

The first step in the video-texture process is locating potential transition points in the input video—places where the video can be looped back on itself in a visually effective, minimally noticeable way. The system the researchers have developed uses a three-component architecture to achieve this. The first component identifies good transition points by computing some measure of similarity between all the pairs of frames in the clip. When the differences between frames are slight, the likelihood of a good transition is high. The system stores the “good” transition points in a small data table that becomes part of the video-texture representation.

The second component synthesizes new video from the analyzed clip by deciding the order in which to play the original frames. This is achieved in one of two ways: using a random-play paradigm in which a table of frame-to-frame similarities computed by the analysis algorithm determines which frame should be played after a given frame, or by selecting a small number of transitions so the video is guaranteed to loop after a specified number of frames.

Once the frame set and sequence has been determined, the rendering component assembles the pieces. The process can be as simple as displaying or outputting the original video frames, or it may involve cross-fading or morphing across transitions or blending independently moving regions.

Extensive Application Possibilities

The basic video-texture concept can be extended to suit a broader range of applications. For example, to enable backward compatibility with existing video players and Web browsers, finite-duration video loops can be created to play continuously by using a novel optimization algorithm. Computer-vision techniques can also be employed to add complexity. For instance, the computer can be programmed to separate background objects and represent them as video sprites. Multiple sprites can be combined and rendered at arbitrary image locations.

Sound synthesis could also be integrated with the video textures, whereby an audio track is re-rendered along with the video texture by taking sound samples associated with each frame and playing them back with the frames that are selected for rendering. Additionally, video textures could be combined with traditional image-based rendering algorithms, such as view interpolation, to obtain 3D video textures that can be applied to simulate 3D motion.

Another possibility is the development of video-based animation, achieved by putting video textures under interactive control in order to drive them at a high level in real time. In such a scenario, the user could interactively specify a desired segment within the source video, then use a slider interface to speed up or slow down the action in that segment.

The video-texture system works best for simple objects defined by smooth, repetitive motion, such as swings and pendulums, and for complex phenomena with relatively smooth, unstructured motion, such as water pouring from a can. The system breaks down, however, with complex, highly structured phenomena, such as full-body motion. Also, says Szeliski, “we haven’t yet been able to deal with complex, sharp-edged scenes like fields of grass. We need to apply much more sophisticated motion- and scene-analysis techniques from computer vision to overcome these problems.”

In addition to focusing on these challenges, the researchers are considering a number of other enhancements. For example, they are looking at ways to improve the quality of the textures in order to make them more broadly applicable, and they hope to introduce automated techniques for programming variety in to long sequences of video from the same set of frames.

Another goal is the addition of creative control over video textures, either through parametric operations or a keyframe approach, to enable the creation of more complex scenes and behavior, such

A segmentation algorithm separated the original video of this scene into regions. The independent video textures were then rendered at arbitrary locations, adding visual complexity.



as crowds and flocking activity.

Along these lines, one of the researchers, Arno Schodl, is planning to extend the video-texture concept to truly controllable animations that could be used interchangeably with traditional computer animation.

Although plans for commercializing the video-texture technology are still far off, the researchers have already met one goal: to prove that a system that collects and synthesizes real-world footage can offer the same flexibility as traditional CG, while attaining a degree of photorealism and naturalness that the latter can’t match.

.....
Diana Phillips Mahoney is chief technology editor of Computer Graphics World.