Published in *Computer Graphics,* **33**(4):42-45, Nov. 1999 (Special Issue on Applications of Computer Vision to Computer Graphics).

Visual Modeling for Computer Animation: Graphics with a Vision

Demetri Terzopoulos
University of Toronto

1 Introduction

Over the past decade, the Visual Modeling Research Program at the University of Toronto has consistently championed the *concerted* exploration of computer graphics and computer vision. Our premise has been this: graphics, the forward/synthesis/models-to-images problem, and vision, the inverse/analysis/images-to-models problem, pose mutually converse challenges, which may best be tackled synergistically through the development of advanced modeling techniques catering to the needs of both fields. With illustrated case studies of three projects spanning a twelve-year period, this brief article presents a personal retrospective on image-based modeling for computer animation. As we shall see, one of the projects has also created opportunities for computer animation to contribute to the study of visual perception in living systems.

I shall begin in the next section by reviewing an early computer animation project that pioneered the use of image-based modeling to combine natural and synthetic imagery. The animation *Cooking with Kurt*, produced in 1987 at the Schlumberger Palo Alto Research center, introduced a paradigm in which computer vision was applied to acquire 3D models of objects from their images. The acquired models were then dynamically animated in a simulated physical scene reconstructed from the image of a real scene. The approach demonstrated a promising alternative to the established convention of keyframe animating manually constructed geometric models.

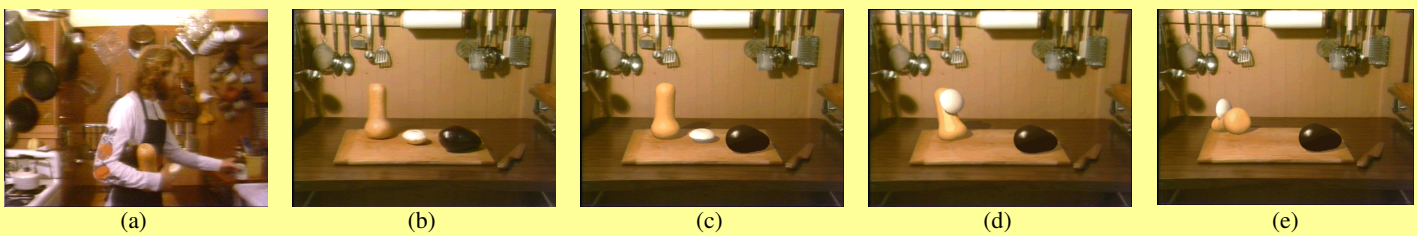
The human face is a natural objective for the image-based modeling approach. I will next describe a facial animation project that uses specialized imaging devices to capture models of human heads with functional, biomechanically simulated faces conforming closely to human subjects. Facial animation is by no means the sole benefactor of an exciting area of advanced graphics modeling that lies at the intersection of virtual reality and artificial life [1]. Accordingly, I will also describe a virtual "seaquarium" populated by artificial marine animals whose 3D shapes and appearances were captured from images of real fishes. As I discuss in the final section, lifelike, self-animating creatures such as these now serve as biomimetic autonomous agents in the study of animal perception.

2 Vision Meets Graphics

Figure 1 summarizes the animation *Cooking with Kurt* [2]. The action begins with video of an actor walking into a kitchen and placing several vegetables on a cutting board. The vegetables "come to life" behind the actor's back in what appears to be the physical kitchen counter environment. They bounce, slide, roll, tumble, and collide with one another and with the table-top, cutting board, and back wall.

Among the novel features of this innovative animation project was the creative use of newly developed computer vision techniques known as deformable models [3]. Deformable models enabled us to reconstruct the 3D shapes of the vegetables from their 2D images, as described in Figure 2. The reconstructed vegetables were physics-based, elastic models. They were animated by numerically simulating their equations of nonrigid motion, producing physically realistic actions [4]. The actions were induced by internal "thruster" driving forces and "servo control" forces, enabling the synthetic vegetables to bounce in an upright manner along choreographed paths. The motions were also affected by external interaction forces due to friction and collision among the models and planar surfaces comprising the simulated kitchen table environment. A physically realistic collision sequence in which the small white squash leaps into the large yellow squash is illustrated in Figure 1(d-e).

Another novel feature of this animation was the compelling illusion that the vegetables are situated in the three-dimensional, physical kitchen scene when, in fact, these animated graphical elements are simply matted into the two-dimensional, background image (Figure 1(b) minus the real vegetables). To achieve the effect, we first employed photogrammetric techniques from computer vision to reconstruct a simplified, 3D geometric scene model from the 2D background image. We also estimated a viewpoint into that virtual scene model consistent with the background image. In particular, we developed interactive optimization methods that positioned three invisible virtual planes and adjusted the virtual camera such that the planes project correctly into the table top, cutting board, and back wall visible in the background image. Similar optimization



(a) (b) (c) (d) (e)

Figure 1. Images from the 1987 computer-animated short "Cooking with Kurt." (a) Kurt Fleischer carrying vegetables in kitchen. (b) Video frame of real vegetables. (c) Reconstructed deformable vegetable models in scene from frame (b). (d–e) Animation frames from a physics-based vegetable collision sequence.

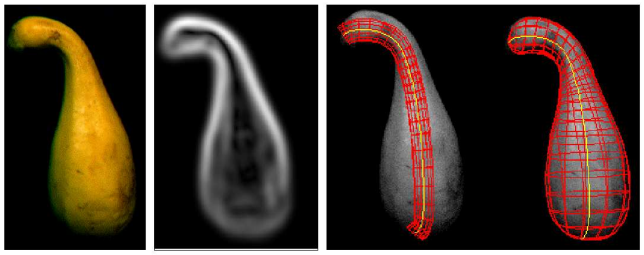


Figure 2. Deformable model 3D reconstruction of a squash from its image. The image (left) is processed into multiscale magnitude-of-gradient potential functions (one scale is shown). These induce forces that attract and constrain the deformable cylinder as it inflates from a crude initial approximation to reconstruct a 3D model, which accurately captures the squash shape (right).

techniques also served in adjusting surface colors and albedos, as well as in positioning a synthetic light source to light the synthetic vegetables and cast shadows on the invisible planes that are consistent with scene shadows evident in the background image. Thus, we could matte in our animate vegetables, as shown in Figure 1(c-e).

Combining real and synthetic imagery has now become a popular special effects technique in the movie industry. For example, state-of-the-art "matchmove" methods enable the rendering of graphical objects into moving background video, by estimating the parameters (pose, motion, focal length, etc.) of the camera that shot the video (see Doug Roble's article in this issue). This requires the tracking in the video sequence of a small set of fiducial points, such as object corners or markings. Computer vision technology promises to mitigate the labor intensiveness of this traditionally manual post-production process. Several start-up companies, among them SynaPix, Science.D.Visions, and REALVIZ, have developed software products that exploit vision techniques for this purpose [5]. The latter company also markets image-based modeling software that does a job analogous to that illustrated in Figure 2.

3 Image-Based Facial "Cloning"

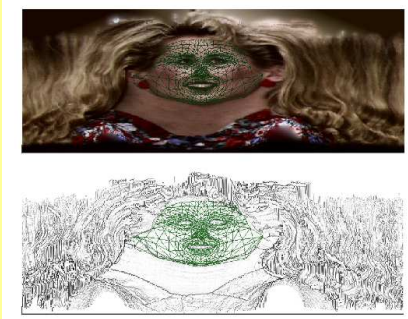
Next, I shall describe a highly automated image-based approach to constructing anatomically accurate, functional models of human heads that can be made to conform closely to specific individuals [6]. Figure 3(a) shows example input images and the resulting functional model, which is suitable for animation.

The image acquisition phase begins by scanning a person with a laser sensor, which circles around the subject's head to acquire detailed range and reflectance images. The figure shows a head-to-shoulder, 360° cylindrical scan of a woman, "Heidi", acquired using a Cyberware Color 3D Digitizer., producing a range image and a registered RGB photometric image, both 512x256 pixel arrays in cylindrical coordinates.

In the image analysis phase, an automatic conformation algorithm adapts an elastic triangulated face mesh of predetermined topological structure to the acquired images. The generic mesh, which is reusable with different individuals, reduces the range data to an efficient, polygonal approximation of the facial geometry and supports a high-resolution texture mapping of the skin reflectivity. Figure 3(b) shows the elastic mesh after it has conformed to the woman's facial area in both the range and RGB images using a feature-based matching algorithm that encodes structural knowledge about the face, specifically the relative arrangement of nose, eyes, ears, mouth, and chin [6]. The 2D positions of the nodes of the conformed mesh serve as texture map coordinates in the RGB image, as well as range map sampling locations from which 3D Euclidean space coordinates are computed for the polygon vertices. The visual quality of the face model is comparable to a 3D display of the original high resolution data, despite the significantly coarser mesh geometry.

After reducing the scanned data to the 3D epidermal mesh, the final phase assembles the physics-based, functional face model. The conformed polygonal mesh forms the epidermal layer of a biomechanical model of facial tissue. An automatic algorithm constructs the multilayer synthetic skin and estimates an underlying skull substructure with a jointed jaw. Finally, the algorithm inserts two dozen synthetic muscles into the deepest layer of the facial tissue. These contractile actuators, which emulate the primary muscles of facial expression, generate forces that deform the synthetic tissue into

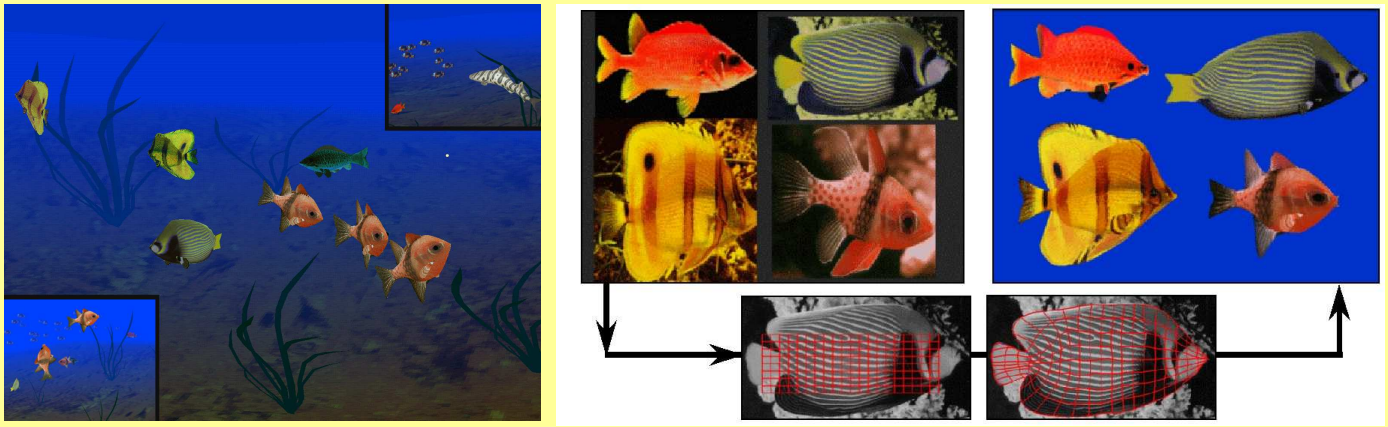
(a)



(b)

(c)

Figure 3. Image-based facial modeling. (a) Cylindrical range and texture images of the head of a real person captured using a Cyberware 3D Color Digitizer. The back of the head is depicted on either side of the facial area at center. From the pair of images at the top, our algorithms "clone" a functional model of the subject, incorporating a textured, biomechanically-simulated deformable facial skin with embedded muscles of facial expression. The synthetic face at the bottom is rendered in neutral and expressive poses dynamically generated through coordinated muscle contractions. (b) Fitting the generic mesh to both RGB texture and edge-enhanced range images. (c) Scenes from the computer-animated short "Bureaucrat Too".



(a) (b)

Figure 4. (a) Artificial fishes in their virtual marine world. (b) From images of real fishes, to textured 3D spline surface fish models. The procedure for converting the images at the upper left to the models at the upper right is illustrated underneath. A deformable mesh model is interactively adjusted from its initial rectangular configuration at the lower left to extract the shape of the fish's body from the image and to produce the nonuniform coordinate system at the lower right for mapping the extracted texture.

meaningful expressions. To increase realism, we include constraints emulating tissue incompressibility and constraints enabling it to slide over the skull substructure.

The lower portion of Figure 3(a) demonstrates that we can animate the resulting face model through the coordinated activation of its internal muscles. Figure 3(c) shows scenes from the animation *Bureaucrat Too*, in which a functional model, which was cloned from a male subject "George", was similarly animated [7].

A alternative to reconstructing geometric facial models from images that are acquired by specialized scanners is to capture multiple, high-resolution photographs or video recordings of the face. The viability of this approach for facial animation was demonstrated in recent papers [8, 9]. Moreover, using high-definition studio photographs, the company Virtual Celebrity Productions has created stunning, stylistically photorealistic animated digital clones of Marlene Dietrich and other legendary stars [10].

4 Image-Based Modeling of Animals

I now turn to our work on creating physics-based virtual worlds inhabited by realistic artificial animals. These sophisticated graphical models are of interest because they are self-animating creatures that dramatically advance the state of the art of character animation and interactive games. For example, we have developed artificial fishes that possess muscle-actuated biomechanical bodies, sensors, and brains with motor, perception, behavior, and learning centers [11]. We have employed artificial fishes to produce two computer-animated shorts for SIGGRAPH's Electronic Theater venue [12].

Artificial fish models, such as the ones illustrated in Figure 4(a), must capture the form and appearance of real fishes with reasonable visual fidelity. To this end, we have once again used image-based modeling techniques. We convert photographs of real fishes into 3D NURBS surface body models using an interactive image-based modeling strategy. The digitized photographs are analyzed semi-automatically using a mesh of "snakes" (deformable contours [3]) that floats freely over the image. The border snakes adhere to intensity edges demarcating the fish from the background, and the remaining snakes relax elastically to cover the imaged fish body. This yields a smooth, nonuniform coordinate system for mapping the texture onto the spline surface to produce the final texture mapped fish body model.

5 Graphics Meets Vision

As biomimetic autonomous agents situated in realistic virtual worlds, artificial animals also foster a deeper understanding of biological information processing, including perception, learning, and cognition. For example, they have enabled an advantageously novel, purely software approach to the design of active vision systems, an activity that has heretofore had to rely on mobile robot hardware [13].

We have been developing an active vision system using artificial fishes and, more recently, artificial humans (see Figure 5), demonstrating that virtual autonomous agents can support serious experimentation with image analysis algorithms and sensorimotor control strategies. Perception begins with a pair of virtual eyes that afford the agent high-acuity foveal vision plus wider field-of-view albeit lower-acuity peripheral vision. With mobile, foveated eyes, controlling gaze through eye movements becomes an important issue. The active vision system includes a stabilization module and a foveation module. By continually minimizing optical flow over the retina, the stabilization module implements an optokinetic reflex, producing egomotion-compensating eye movements that stabilize the visual field during locomotion. The foveation module directs the gaze to objects of interest based on visual models stored in the agent's brain. For example, a primary visual cue for recognition is color. Given a mental model of a tan colored uniform, the observer soldier in the figure recognizes another virtual soldier wearing a tan uniform, tracks him visually, and autonomously locomotes in pursuit of the moving target.

Acknowledgements

I am indebted to my present and former students, Radek Grzeszczuk, Yuencheng Lee, Tamer Rabie, Xioyuan Tu, and colleages, Kurt Fleischer, Michael Kass, Keith Waters, Andrew Witkin, for their invaluable contributions to the research reviewed herein.

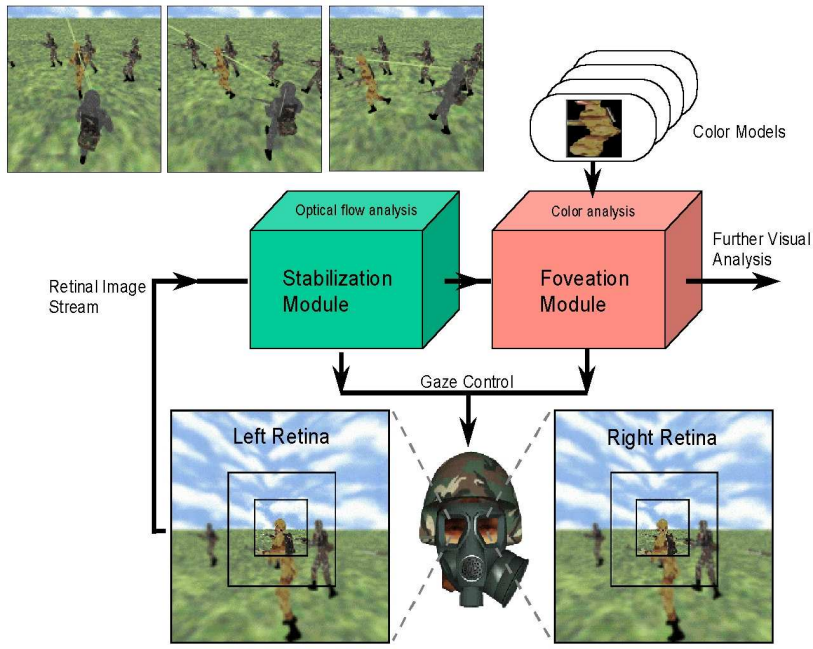


Figure 5. Virtual humans with active vision. The images at the upper left show a virtual soldier, the observer, visually tracking and following on foot another virtual soldier, the target, wearing a tan colored uniform (lines from the observer's eyes indicate the gaze direction). The remainder of the figure illustrates the active vision system that we have incorporated into the observer soldier (the soldiers are animated using the "*DI-Guy*" API from Boston Dynamics, Inc.). Each eye, which is capable of eye movements, is implemented as a set of coaxial virtual cameras that render small (64x64) images of the virtual world with progressively wider fields of view. These images are appropriately expanded and composited (as indicated by the black borders) to model biomimetically foveated retinas, here illustrated as the observer gazes at the target. The observer's vision system comprises a stabilization module and a foveation module, which are responsible for actively controlling the eyes. The stabilization module stabilizes the field of view of the moving observer by inducing (egomotion-compensating) eye movements that minimize optical flow over the retinas. The foveation module matches color (mental) models to regions in the retinal images in order to recognize targets of interest according to their distinctive colors. It produces eye movements necessary to center a recognized target within the high-acuity, foveal region in the retinas for further visual analysis. The observer soldier follows the target soldier using a sensorimotor control loop that steers the body in accordance with the gaze direction.

Funding was provided in part by the Natural Sciences and Engineering Research Council of Canada and the Canada Council for the Arts.

References

[1] D. Terzopoulos. Artificial life for computer graphics. *Communications of the ACM*, **42**(8):32-42, 1999.

[2] K. Fleischer, A. Witkin, M. Kass, and D. Terzopoulos. "*Cooking with Kurt*". Animation in ACM SIGGRAPH Video Review Issue 36: SIGGRAPH 87 Film & Video Show, 1998.

[3] D. Terzopoulos, A. Witkin, and M. Kass. Constraints on deformable models: Recovering 3D shape and nonrigid motion. *Artificial Intelligence*, **36**(1):91-123, 1988.

[4] D. Terzopoulos and A. Witkin. Physically-based models with rigid and deformable components. *IEEE Computer Graphics and Applications*, **8**(6):41-51, 1988.

[5] SynaPix, Inc., Lowell, MA; see www.synapix.com. Science.D.Visions, Dortmund, Germany; see www.sci-d-vis.com. REALVIZ, S.A., Sophia Antipolis, France; see www.realviz.com.

[6] Y. Lee, D. Terzopoulos, and K. Waters. Realistic facial modeling for animation. In *Computer Graphics* Proc. ACM SIGGRAPH 95 (Los Angeles, CA) August 1995, 55-62.

[7] K. Waters, Y.Lee, and D. Terzopoulos. "*Bureaucrat Too*". Animation in ACM SIGGRAPH Video Review Issue 109: Selections from the Facial Animation Workshop, 1995.

[8] B. Guenter, C. Grimm, D. Wood, H. Malvar, F. Pighin., Making faces. In *Computer Graphics* Proc. ACM SIGGRAPH 98 (Orlando, FL), July 1998, 55-66.

[9] F. Pighin, J. Hecker, D. Lishinski, R. Szeliski, D. Salesin. Synthesizing realistic facial expressions from photographs. In *Computer Graphics* Proc. ACM SIGGRAPH 98 (Orlando, FL), July 1998, 75-84.

[10] Virtual Celebrity Productions, LLC, Los Angeles, CA; see www.virtualcelebrity.com.

[11] D. Terzopoulos, X. Tu, and R. Grzeszczuk. Artificial fishes: Autonomous locomotion, perception, behavior, and learning in a simulated physical world. *Artificial Life*, **1**(4):327-351, 1994.

[12] X. Tu, R.Grzeszczuk, D. Terzopoulos, "*The Undersea World of Jack Cousto.*" Animation screened at the ACM SIGGRAPH 95 Electronic Theater, Los Angeles, August 1995. See also "*Go Fish!*" Animation excerpted in ACM SIGGRAPH Video Review Issue 91: SIGGRAPH 93 Electronic Theater.

[13] D. Terzopoulos and T. Rabie. Animat vision: Active vision in artificial animals. *Videre: Journal of Computer Vision Research*, **1**(1):2-19, 1997. See www.cs.toronto.edu/~dt/animat-vision.