Modeling Living Systems for Computer Vision

Demetri Terzopoulos

Department of Computer Science, University of Toronto
10 King's College Road, Toronto, Ontario, M5S 1A4, Canada
e-mail: dt@cs.toronto.edu

Abstract

This paper presents research spanning the fields of computer vision, computer graphics, and artificial life, with implications for AI. Although the modeling of all aspects of living systems is a worthwhile endeavor in its own right, the emphasis here will be on the modeling of animals, including humans, for computer vision. First, I present a new breed of artificial animals in a physics-based virtual marine world, whose muscle-actuated bodies harbor brains with motor, perception, behavior, and learning centers. In these mobile autonomous agents, sensorimotor control for the purposes of perceptually-guided navigation employs on-board, active computer vision systems that continually analyze the visual world. Second, turning my attention to human animals, I describe new algorithms that can construct artificial human heads with expressive faces. Through the use of range scanners, generic biomechanical facial models may be automatically personalized to individuals. Currently, artificial faces support a model-based approach to facial image analysis. In the future, it should be possible to incorporate brains and some degree of intelligence into them as well.

1 Introduction

Modeling has been a central theme of artificial intelligence. This is very evidently so in computer vision, where the modeling of objects has preoccupied researchers since the dawn of the field some three decades ago [1]. A good strategy for progress on the vision problem is through physics-based modeling (see, e.g., [2]). The idea is to incorporate principles of physical dynamics into conventional geometric models in order to be able to represent not only the shapes of objects, but their physical behaviors as well. Indeed, many of the products of the physics-based modeling movement in computer graphics (see, e.g., [3]) are also useful in vision. This fuels our long-standing philosophy in the Visual Modeling Group at the University of Toronto that vision and graphics are mutually converse problems, and that the two fields should advance synergistically.

In this paper, I will demonstrate that we have now taken an important next step and introduced some *artifical life* into the vision-graphics marriage. This enables us to set our sights on the most complex objects known—objects that are alive. I will demonstrate how recent advances in the emerging field of A-Life are inspiring fresh approaches to computer vision.[1] These advances center around the idea of artificial animals, or "animats" a term coined by Wilson [4]. In particular, I will review two of our ongoing research projects, which relate to the modeling of living systems for computer vision. The first involves the modeling of complete animals of nontrivial complexity on the evolutionary ladder, such as teleost fishes in their natural habitats (Fig. 1). The second project, involves the modeling of faces, a vitally important communicative medium of the most highly evolved living systems known—human beings (Fig. 2).

The presentation is in two parts. Sections 2 and 3 present our work on *artificial fishes* and *animat vision*. The basic idea in a nutshell is to implement, entirely in software, realistic artificial animals and to give them the ability to locomote, perceive, and in some sense understand the realistic virtual worlds in which they are situated so that they may achieve both individual and social functionality within these worlds. To this end, each animat is an autonomous agent possessing a muscle-actuated body that can locomote and a mind with motor, perception, behavior, and learning centers. The animat is endowed with functional eyes that can image the dynamic 3D virtual world onto its 2D virtual retinas. The perceptual center of the animat's brain exploits "active vision" algorithms to continually process the incoming retinal image stream in order to make sense of what it sees and, hence, to purposefully navigate its world. I hope to begin to convince the uninitiated reader that it is now within our means to implement realistic virtual worlds inhabited by artificial animals rich enough to support some serious computer vision (and AI) research.

In the second part of the paper, Sections 4 and 5, I briefly present our work on physical and anatomical modeling of human faces. Our goal has been to develop *artificial faces* that are capable of synthesizing realistic facial expressions. At different levels of abstraction, these hierarchical models capture knowledge about facial expression from psychology, facial anatomy and facial tissue histology, and continuum biomechanics. I will show that a generic facial model of this sort can be "personalized", or made to conform closely to

[1] For an engaging survey of the A-Life field, see, e.g., S. Levy, *Artificial Life* (Pantheon, 1992). Journals such as *Artificial Life* and *Adaptive Behavior* document the state of the art.

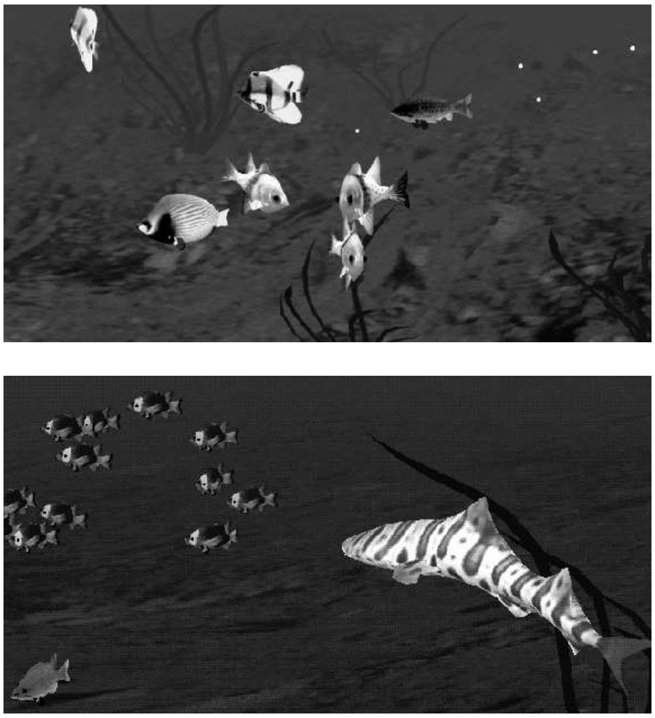


Figure 1: Artificial fishes in their physics-based virtual world as it appears to an underwater observer (monochrome version of original color images). Top: The 3 reddish fish (center) are engaged in a mating ritual, the greenish fish (upper right) is a predator hunting for small prey, the remaining 3 fishes are feeding on plankton (white dots). Dynamic seaweeds grow from the ocean bed and sway in the current. Bottom: A predator shark stalking a school of prey.

Figure 2: An artificial face (monochrome version of original color images). The functional face model was constructed automatically from an RGB/range laser scan of an individual, "George". Artificial George is shown here engaged in synthesizing various facial expressions and pretending to read a technical paper about how he was constructed.

individuals once the geometry and photometry of their faces has been captured by a range sensor. Finally, I will describe how this sophisticated model can be used in a model-based analysis-by-synthesis strategy to analyze facial images and image sequences, an important computer vision problem that relates to visual communication.

Section 6 concludes the paper with a brief discussion of where I hope that our approach will lead us and a preview of future work.

2 Artificial Fishes

Imagine a virtual marine world inhabited by a variety of realistic fishes (Fig. 1). In the presence of underwater currents, the fishes employ their muscles and fins to swim gracefully around immobile obstacles and among moving aquatic plants and other fishes. They autonomously explore their dynamic world in search of food. Large, hungry predator fishes stalk smaller prey fishes in the deceptively peaceful habitat. The sight of predators compels prey fishes to take evasive action. When a dangerous predator appears in the distance, similar species of prey form schools to improve their chances of survival. As the predator nears a school, the fishes scatter in terror. A chase ensues in which the predator selects victims and consumes them until satiated. Some species of fishes seem untroubled by predators. They find comfortable niches and feed on floating plankton when they get hungry. Driven by healthy libidos, they perform elaborate courtship rituals to attract mates.

Each artificial fish is an autonomous agent with a deformable body actuated by internal muscles. The body also harbors eyes and a brain with motor, perception, behavior, and learning centers, as Fig. 3 illustrates. Through controlled muscle actions, artificial fishes are able to swim through simulated water in accordance with hydrodynamics. Their functional fins enable them to locomote, maintain balance, and maneuver in the water. Thus the artificial fish model captures not just 3D form and appearance, but also the basic physics of the animal and its environment. Though rudimentary compared to real animals, the minds of artificial fishes are nonetheless able to learn some basic motor functions and carry out perceptually guided motor tasks. In accordance with their perceptual awareness of the virtual world, their minds arbitrate a repertoire of piscatorial behaviors, including collision avoidance, foraging, preying, schooling, and mating.

The details of the artificial fish model are presented in [5; 6]. I will summarize its main features in the remainder of this section.

2.1 Motor System

The motor system comprises the dynamic model of the fish including its muscle actuators and a set of motor controllers (MCs). Fig. 4(a) illustrates the biomechanical body model which produces realistic piscatorial locomotion using only 23 lumped masses and 91 elastic elements. These mechanical components are interconnected so as to maintain the structural integrity of the body as it flexes due to the action of its 12 contractile muscles.

Artificial fishes locomote like real fishes, by autonomously contracting their muscles. As the body flexes it displaces virtual fluid which induces local reaction forces normal to the

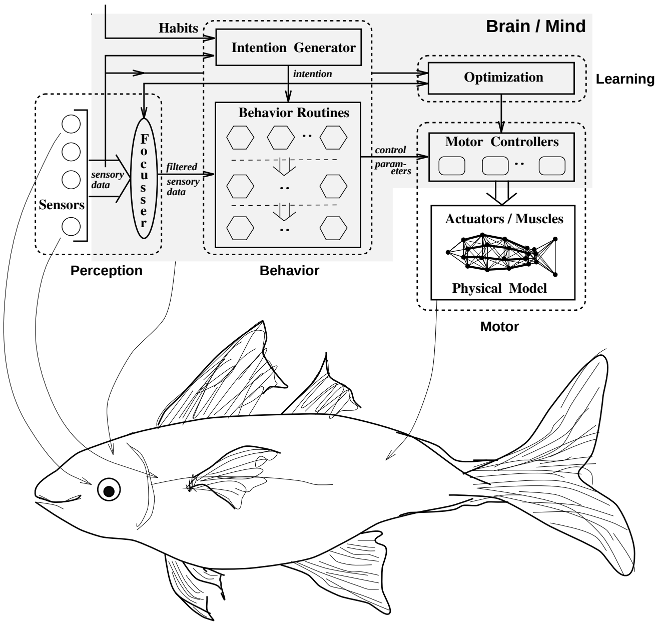
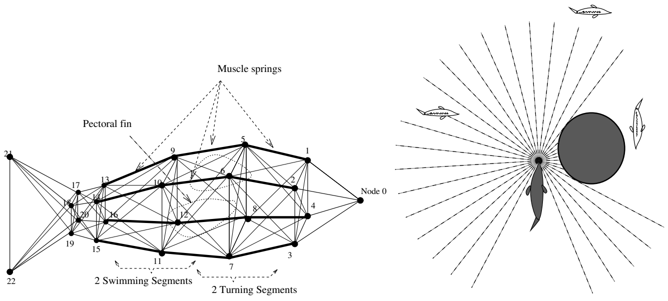


Figure 3: Control and information flow in artificial fish.



(a) (b)

Figure 4: Biomechanical fish model (a). Nodes denote lumped masses. Lines indicate uniaxial elastic elements (shown at natural length). Bold lines indicate muscle elements. Artificial fishes perceive objects (b) within a limited field view if objects are close enough and not occluded by other opaque objects (only the fish towards the left is visible to the one at the center).

body. These hydrodynamic forces generate thrust that propels the fish forward. The model mechanics are governed by Lagrange equations of motion driven by the hydrodynamic forces. The system of coupled second-order ordinary differential equations are continually integrated through time by a numerical simulator.[2]

The model is sufficiently rich to enable the design of motor controllers by gleaning information from the fish biomechanics literature. The motor controllers coordinate muscle actions to carry out specific motor functions, such as swimming forward (swim-MC), turning left (left-turn-MC), and turning right (right-turn-MC). They translate natural control parameters such as the forward speed or angle of the turn into detailed muscle actions that execute the function. The artificial fish is neutrally buoyant in the virtual water and has a pair of pectoral fins that enable it to navigate freely in its 3D world by pitching, rolling, and yawing its body. Additional motor controllers coordinate the fin actions.

2.2 Perception System

Artificial fishes are aware of their world through sensory perception. The perception system relies on a set of on-board virtual sensors to gather sensory information about the dynamic environment. As Fig. 4(b) suggests, it is necessary to model not only the abilities but also the limitations of animal perception systems in order to achieve natural sensorimotor behaviors. The perception center of the brain includes a perceptual attention mechanism (see [7] for a review of attention mechanisms) which allows the artificial fish to sense the world in a task-specific way, hence filtering out sensory information superfluous to its current behavioral needs. For example, the artificial fish attends to sensory information about nearby food sources when foraging. Our early artificial fishes—those that are *not* equipped with the active vision system described in Section 3—employ *simulated* perception, a "perceptual oracle" which satisfies the fish's sensory needs by directly interrogating the 3D world model—the fish can directly access the geometric and photometric information that is available to the graphics rendering engine, as well as object identity and dynamic state information about the physics-based world model.

2.3 Behavior

The behavior center of the artificial fish's brain mediates between its perception system and its motor system. An intention generator, the fish's cognitive faculty, harnesses the dynamics of the perception-action cycle. The innate character of the fish is established by a set of habits that determine if it is male or female, predator or prey, etc. At each simulation time step, the intention generator takes into account the habits of the fish and the incoming stream of sensory information to generate dynamic goals for the fish, such as to avoid an obstacle, to hunt and feed on prey, or to court a potential mate. It ensures that goals have some persistence by exploiting a single-item memory. Persistence makes sustained behaviors such as foraging, schooling, and mating more robust. The intention generator also controls the perceptual attention mechanism. At every simulation time step, the intention generator activates behavior routines that attend to sensory information and compute the appropriate motor control parameters to carry the fish one step closer to fulfilling its current intention. The behavioral repertoire of the artificial fish includes primitive, reflexive behavior routines, such as obstacle avoidance, as well as more sophisticated motivational behavior routines such as schooling and mating whose activation depends on the dynamic mental state of the fish as represented by hunger, fear, and libido mental variables (see [5] for the details).

[2] The artificial fish model achieves a good compromise between realism and computational efficiency. For example, the implementation can simulate a scenario with 10 fishes, 15 food particles, and 5 static obstacles at about 4 frames/sec (with wireframe rendering) on a Silicon Graphics R4400 Indigo^2 Extreme workstation. More complex scenarios with large schools of fish, dynamic plants, and full color texture mapped GL rendering at video resolution can take 5 seconds or more per frame.

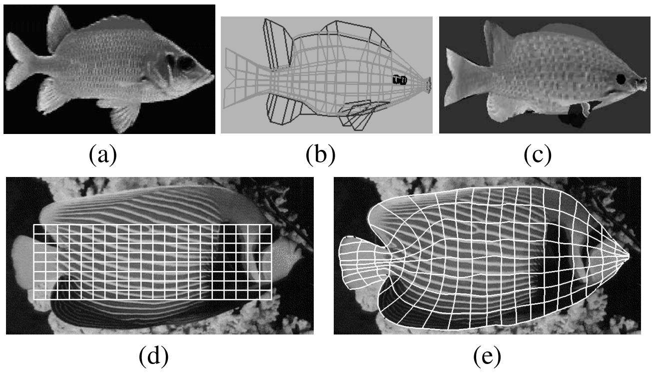


Figure 5: (a) Digitized color image of a fish photo. (b) 3D NURBS surface fish body. (c) Color texture mapped 3D fish model. Initial (d) and final (e) snake-grid on an image of a different fish.

2.4 Learning

The learning center of its mind enables the artificial fish to learn how to locomote through practice and sensory reinforcement. Through optimization, the motor learning algorithms discover muscle controllers that produce efficient locomotion. Muscle contractions that produce forward movements are "remembered". These half-success then form the basis for the fish's subsequent improvement in its swimming technique. Their brain's learning center also enable these artificial animals to train themselves to accomplish higher level sensorimotor tasks, such as maneuvering to reach a visible target (see [5; 8] for the details).

2.5 Modeling Form and Appearance

Of course, we want our animats to capture the form and appearance of real fishes with considerable visual fidelity. Visual fidelity is especially important in the application of our animats to computer vision, which I will describe shortly. To this end, photographs of real fishes, such as the one shown in Fig. 5(a), are converted into 3D spline (NURBS) surface body models (Fig. 5(b)). The digitized photographs are analyzed semi-automatically using a "snake-grid" tool which is demonstrated in Fig. 5(d–e) on a different fish image. The grid of snakes [9] floats freely over an 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 (Fig. 5(e)) for mapping the texture onto the spline surface to produce the final texture mapped fish body model (Fig. 5(c)).

3 Animat Vision

The psychologist J.J. Gibson stressed in pre-computational terms the importance of modeling the active observer situated in the dynamic environment [10]. Versions of this paradigm suitable for mainstream computer vision were introduced in the seminal papers of Bajcsy [11] and Ballard [12] under the names of active perception and animate vision, respectively.[3]

[3]"Animat" vision should not be confused with Ballard's "animate" vision; the latter does not involve artificial animals.

The active vision approach has developed into a prevailing paradigm [13; 14; 15; 16].

3.1 Problems with the "Hardware Vision" Mindset

As active vision is practiced in most labs today, however, it is in reality little more than technology-driven "hardware vision". To be sure, applications-minded researchers have legitimate reasons for building robot vision systems, but the necessary hardware paraphernalia—CCD cameras, pan-tilt units, ocular heads, frame-rate image processors, mobile platforms, manipulators, controllers, interfaces, etc.—can be expensive to fabricate or acquire commercially and a burden to maintain in good working order.

The animat vision methodology that we propose in [17] can potentially liberate a significant segment of the computer vision research community from the tyranny of robot hardware. It addresses the needs of scientists who are motivated to understand and ultimately reverse engineer the powerful vision systems found in higher animals. These researchers are well aware that animals do not have CCD chip eyes, electric motor muscles, and wheel legs. That is to say, they realize that readily available hardware systems are terrible models of biological animals. For lack of a better alternative, however, they have been struggling with inappropriate hardware in their ambition to understand the complex sensorimotor functions of real animals. Moreover, their mobile robots typically lack the compute power necessary to achieve real-time response within a fully dynamic world while permitting active vision research of much significance.

Artificial animals such as the fish are active "vehicles" in the sense of Braitenberg [18]. We believe that they are as appropriate for grounding active vision systems as are the hardware "mobots" that have come out of the situated robotics work of Brooks and his group [19; 20] and have been an inspiration to numerous other robotics groups (see, e.g., the compilation [21]). Undeniably, however, efforts to equip real-time mobile robots with general-purpose active vision systems have been hampered by the hardware and the relatively modest abilities of on-board processors.

3.2 Get A–Life!

I will now describe a zoomimetic approach to vision [17] which is made possible by the confluence of

1. advanced physics-based artificial life modeling of animals, as I presented it in the previous section
2. photorealistic computer graphics rendering, especially as implemented in modern 3D graphics workstations, and
3. active computer vision algorithms.

Our animat vision approach offers an alternative strategy for developing biologically inspired active vision systems that circumvents the aforementioned problems of hardware vision. The animat vision concept is realized with realistic artificial animals and active vision algorithms implemented entirely in software on 3D graphics workstations. Animat vision offers several additional advantages:

- One can slow down the "cosmic clock" of the virtual world relative to the cycle time of the CPU on which it is being simulated. This increases the amount of computation that each agent can consume between clock ticks without unduly retarding the agent's responses relative to the temporal evolution of its virtual world. This in turn permits the development and evaluation of new and/or computationally complex active vision algorithms that are not presently implementable in real-time hardware.

- The quantitative photometric, geometric, and dynamic information that is needed to render the virtual world is available explicitly. Generally, the animats are privy to none of this environmental ground truth data, but must glean visual information the hard way—from their retinal image streams. However, the readily available ground truth can be extremely useful in assaying the effective accuracy of the vision algorithms or modules under development.

Our challenge has been to synthesize a prototype active vision system for the fish animat which is based solely on retinal image analysis [17]. The vision system should be extensible so that it will eventually support the broad repertoire of individual and group behaviors of artificial fishes. It is important to realize that we need not restrict ourselves to modeling the perceptual mechanisms of real fishes. In fact, the animat vision paradigm applies to animats that model any animal—even a human being—to the level of fidelity that the artificial fish models a real fish. Indeed, the animat vision system that we have developed does not attempt to model fish perception [22]. Given a piscine animat that is an active observer of its world, we have found it interesting and challenging to endow the animat with human-like retinal imaging capabilities!

3.3 Active Vision System

The basic functionality of the animat vision system starts with binocular perspective projection of the color 3D world onto the animat's 2D retinas. Retinal imaging is accomplished by photorealistic graphics rendering of the world from the animat's point of view. This projection respects occlusion relationships among objects. It forms spatially variant visual fields with high resolution foveas and low resolution peripheries. Based on an analysis of the incoming color retinal image stream, the visual center of the animat's brain supplies saccade control signals to its eyes and stabilize the visual fields during locomotion, to attend to interesting targets based on color, and to keep a moving/deforming target fixated. The artificial fish is thus able to approach and track other artificial fishes using visual feedback. Eventually its arsenal of active vision algorithms will enable it to forage, evade predators, find mates, etc.

Fig. 6 is a block diagram of the active vision system showing two main modules that control foveation of the eyes and retinal image stabilization.

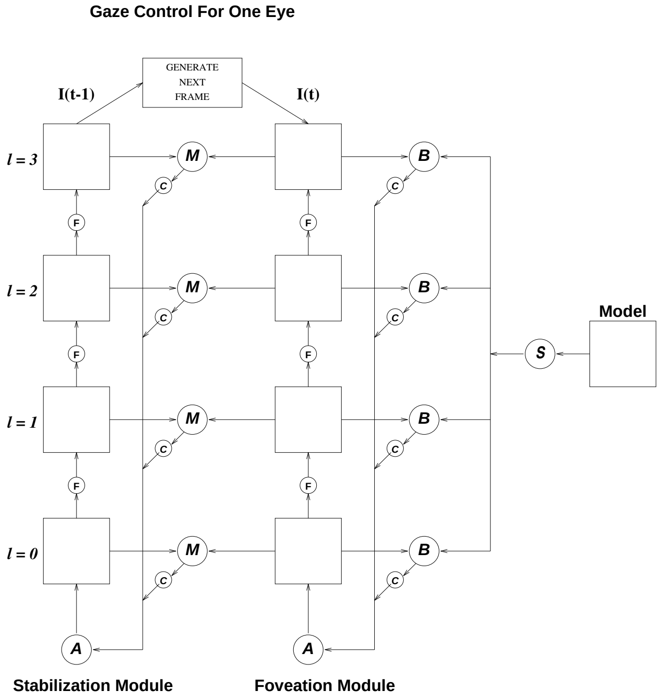


Figure 6: The animat vision system. The flow of the algorithm is from right to left. A: Update gaze angles (θ, ϕ) and saccade using these angles, B: Search current level for model target and if found localize it, else search lower level, C: Select level to be processed (see text), F: Reduce field of view for next level and render, M: Compute a general translational displacement vector (u, v) between images $I(t-1)$ and $I(t)$, S: Scale the color histogram of the model for use by the current level.

Eyes and Retinal Imaging

The artificial fish has binocular vision. The movements of each eye are controlled through two gaze angles (θ, ϕ) which specify the horizontal and vertical rotation of the eyeball, respectively. The angles are given with respect to the head coordinate frame, such that the eye is looking straight ahead when $\theta = \phi = 0°$.

Each eye is implemented as four coaxial virtual cameras to approximate the spatially nonuniform, foveal/peripheral imaging capabilities typical of biological eyes. Fig. 7(a) shows an example of the 64×64 images that are rendered by the four coaxial cameras (using the GL library and SGI graphics pipeline) of the left and right eye. The level $l = 0$ camera has the widest field of view (about 120°) and the lowest resolution. The resolution increases and the field of view decreases with increasing l. The highest resolution image at level $l = 3$ is the fovea and the other images form the visual periphery. Fig. 7(b) shows the 512×512 binocular retinal images composited from the coaxial images at the top of the figure. To reveal the retinal image structure in the figure, we have placed a white border around each magnified component image.

The advantages of the multiresolution retina are significant. Vision algorithms which process the four 64×64 component images are 16 times more efficient than those that process a uniform 512×512 retinal image.

Foveation by Color Object Detection

The mind of the fish stores a set of color models of objects that are of interest to it. For instance, if the fish is by habit a predator, it would possess models of prey fish. The models are stored as a list of 64×64 RGB color images in the fish's memory.

To detect and localize any target that may be imaged in the low resolution periphery of its retinas, the animat vision system of the fish employs an improved version of a color indexing algorithm proposed by Swain [23]. Since each model object has a unique color histogram signature, it can be detected in the retinal image by histogram intersection and

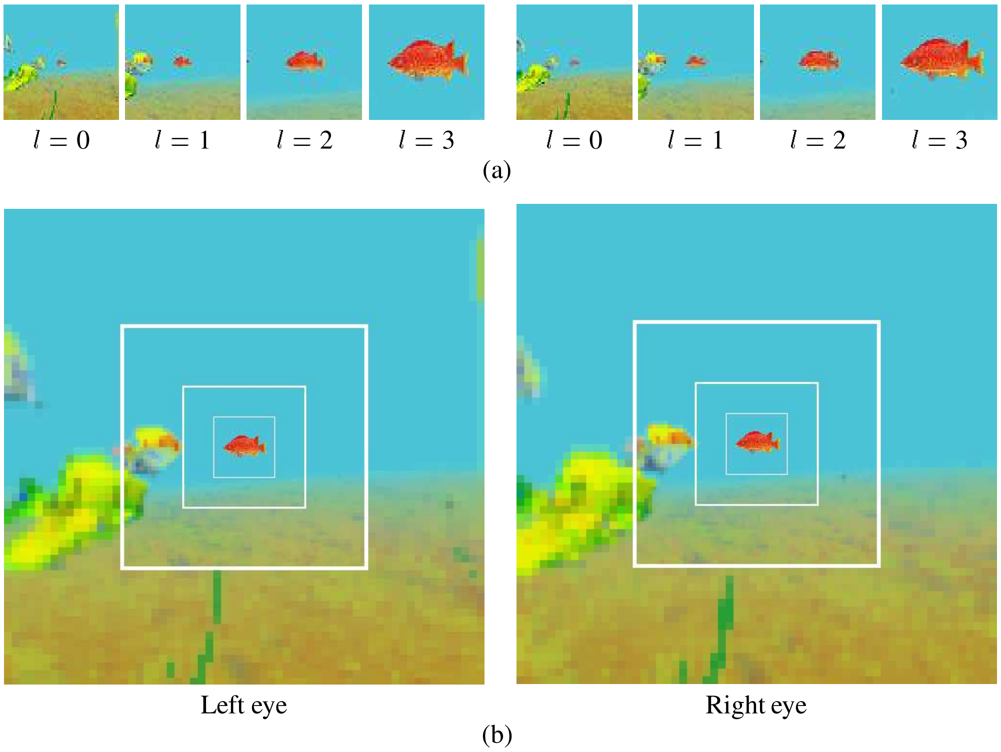


Figure 7: Binocular retinal imaging (monochrome versions of original color images). (a) 4 component images; $l = 0, 1, 2$, are peripheral images; $l = 3$ is foveal image. (b) Composited retinal images (borders of composited component images are shown in white).

localized by histogram backprojection.

Saccadic Eye Movements

When a target is detected in the visual periphery, the eyes will saccade to the angular offset of the object to bring it within the fovea. With the object in the high resolution fovea, a more accurate foveation is obtained by a second pass of histogram backprojection. A second saccade typically centers the object accurately in both left and right foveas, thus achieving vergence.

Module A in Fig. 6 performs the saccades by incrementing the gaze angles (θ, ϕ) in order to rotate the eyes to the required gaze direction.

Visual Field Stabilization using Optical Flow

It is necessary to stabilize the visual field of the artificial fish because its body undulates as it swims. Once a target is verged in both foveas, the stabilization process (Fig. 6) assumes the task of keeping the target foveated as the fish locomotes.

Stabilization is achieved by computing the overall translational displacement (u, v) of light patterns between the current foveal image and that from the previous time instant, and updating the gaze angles to compensate. The displacement is computed as a translational offset in the retinotopic coordinate system by a least squares minimization of the optical flow between image frames at times t and $t - 1$ [24].

The optical flow stabilization method is robust only for small displacements between frames. Consequently, when the displacement of the target between frames is large enough that the method is likely to produce bad estimates, the foveation module is invoked to re-detect and re-foveate the target as described earlier.

Each eye is controlled independently during foveation and stabilization of a target. Hence, the two retinal images must be correlated to keep them verged accurately on the target. Referring to Fig. 8, the vergence angle is $\theta_V = (\theta_R - \theta_L)$ and its magnitude increases as the fish comes closer to the target. Therefore, once the eyes are verged on a target, it is straightforward for the fish vision system to estimate the range to the target from the gaze angles.

3.4 Vision-Guided Navigation

The fish can use the gaze direction for the purposes of navigation in its world. In particular, it is natural to use the gaze angles as the eyes are fixated on a target to navigate towards the target. The θ angles are used to compute the left/right turn angle θ_P shown in Fig. 8, and the ϕ angles are similarly used to compute an up/down turn angle ϕ_P. The fish's turn motor controllers are invoked to execute a left/right turn—left-turn-MC for an above-threshold positive θ_P and right-turn-MC for negative θ_P (see Section 2)—with $|\theta_P|$ as parameter. Up/down turn motor commands are issued to the fish's pectoral fins, with an above-threshold positive ϕ_P interpreted as "up" and negative as "down".

The problem of pursuing a moving target that has been fixated in the foveas of the fish's eyes is simplified by the gaze control mechanism described above. The fish can robustly track a target in its fovea and follow it around by using the turn angles (θ_P, ϕ_P) computed from the gaze angles that are continuously updated by the foveation/stabilization algo-

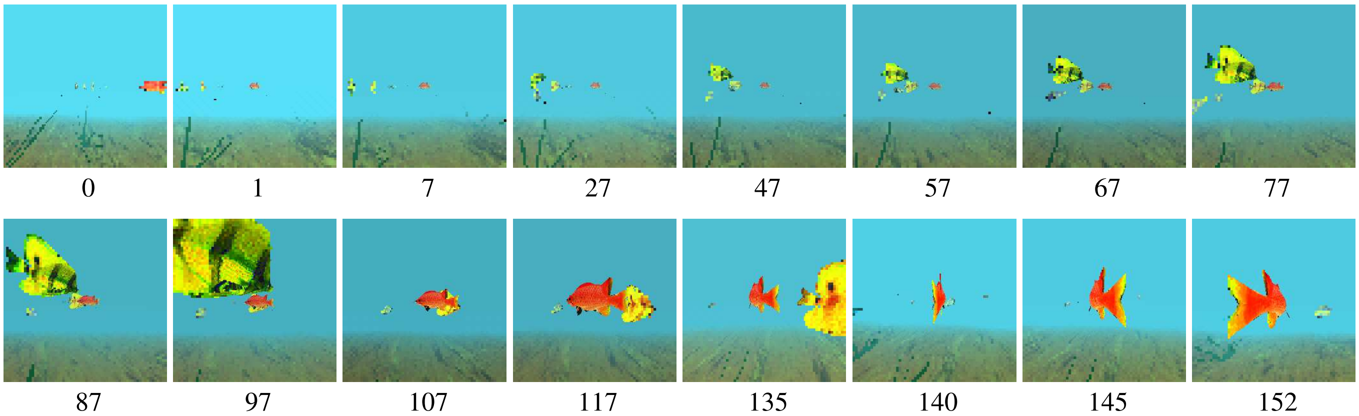


Figure 10: Retinal image sequence from the left eye of the active vision fish as it detects and foveates on a reddish fish target and swims in pursuit of the target (monochrome versions of original color images). The target appears in the periphery (middle right) in frame 0 and is foveated in frame 1. The target remains fixated in the center of the fovea as the fish uses the gaze direction to swim towards it (frames 7–117). The target fish turns and swims away with the observer fish in visually guided pursuit (frames 135–152).

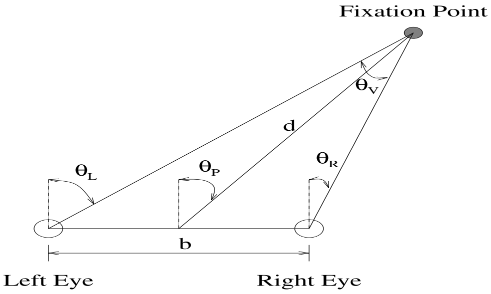


Figure 8: Gaze angles and range to target geometry.

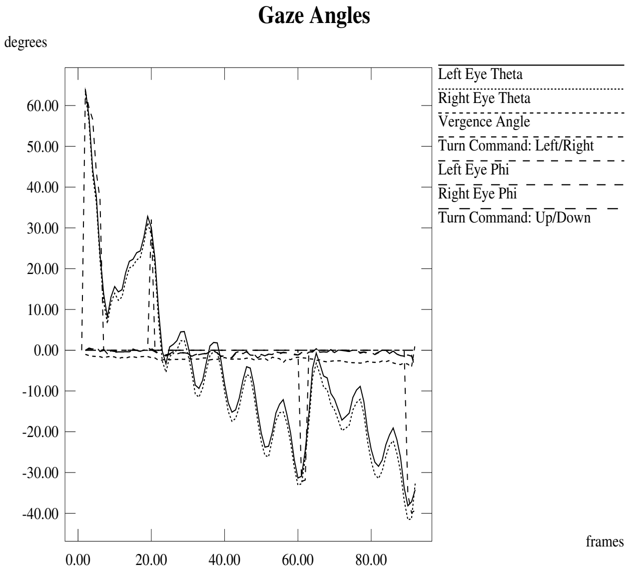


Figure 9: Gaze angles resulting from the pursuit of a target by the AV fish.

rithms.

We have carried out numerous experiments in which the moving target is a reddish prey fish whose color histogram model is stored in the memory of a predator fish equipped with the active vision system. Fig. 9 shows plots of the gaze angles and the turn angles obtained over the course of 100 frames in a typical experiment as the predator is fixated on and actively pursuing a prey target. Fig. 10 shows a sequence of image frames acquired by the fish during its navigation (monochrome versions of only the left retinal images are shown). Frame 0 shows the target visible in the low resolution periphery of the fish's eyes (middle right). Frame 1 shows the view after the target has been detected and the eyes have performed a saccade to foveate the target (the scale difference of the target after foveation is due to perspective distortion). The subsequent frames show the target remaining fixated in the fovea despite the side-to-side motion of the fish's body as it swims towards the target.

The saccade signals that keep the predator's eyes fixated on its prey as both are swimming are reflected by the undulatory responses of the gaze angles in Fig. 9. The figure also shows that the vergence angle increases as the predator approaches its target (near frame 100). In comparison to the θ angles, the ϕ angles show little variation, because the fish does not undulate vertically very much as it swims forward. It is apparent from the graphs that the gaze directions of the two eyes are nicely correlated.

Note that in frames 87–117 of Fig. 10, a yellow fish whose size is similar to the target fish passes behind the target. In this experiment the predator was instructed to be totally uninterested in and not bother to foveate non-reddish objects. Because of the color difference, the yellow object does not distract the fish's gaze from its reddish target. This demonstrates the robustness of the color-based fixation algorithm.

4 Artificial Faces

I will now shift gears and discuss the modeling of human animals, focusing on the important and challenging problem of modeling human faces.

The human face has attracted much attention in several disciplines, including psychology, computer vision, and computer graphics. Psychophysical investigations clearly indicate that faces are very special visual stimulii. Psychologists have studied various aspects of human face perception and recognition [25; 26]. They have also examined facial expression—the result of a confluence of voluntary muscle articulations which deform the neutral face into an expressive face. The facial pose space is immense. The face is capable of generating on the order of 55,000 distinguishable facial expressions with about 30 semantic distinctions. Studies have identified six primary expressions that communicate anger, disgust, fear, happiness, sadness, and surprise in all cultures.

Ekman and Friesen's "Facial Action Coding System" (FACS) provides a quantification of facial expressions [27]. The FACS quantifies facial expressions in terms of 44 "action units" (AU) involving one or more muscles and associated activation levels.

4.1 A Functional Facial Model

We have developed a hierarchical model of the face which provides natural control parameters and is efficient enough to run at interactive rates [28]. Conceptually, the model decomposes into six levels of abstraction. These representational levels encode specialized knowledge about the psychology of human facial expressions, the anatomy of facial muscle structures, the histology and biomechanics of facial tissues, and facial skeleton geometry and kinematics:

1. *Expression.* At the highest level of abstraction, the face model executes expression (or phoneme) commands. For instance, it can synthesize any of the six primary expressions within a specific time interval and with a specified degree of emphasis.
2. *Control.* A muscle control process, a subset of Ekman and Friesen's FACS, translates expression (or phoneme) instructions into a coordinated activation of actuator groups in the facial model.
3. *Muscles.* As in real faces, muscles comprise the basic actuation mechanism of the model. Each muscle model consists of a bundle of muscle fibers. When fibers contract, they displace their points of attachment in the facial tissue or the jaw.
4. *Physics.* The face model incorporates a physical approximation to human facial tissue. The tissue model is a lattice of point masses connected by nonlinear elastic springs. Large-scale synthetic tissue deformations, subject to volume constraints, are simulated numerically by continuously propagating through the tissue lattice the stresses induced by activated muscle fibers.
5. *Geometry.* The geometric representation of the facial model is a non-uniform mesh of polyhedral elements whose sizes depend on the curvature of the neutral face. Muscle-induced synthetic tissue deformations distort the neutral geometry into an expressive geometry.
6. *Images.* After each simulation time step, standard visualization algorithms implemented in dedicated graphics hardware render the deformed facial geometry in accordance with viewpoint, light source, and skin reflectance information to produce the lowest level representation in the modeling hierarchy, a continuous stream of facial images.

The hierarchical structure of the model encapsulates most of the complexities of the underlying representations, relegating the details of their computation to automatic procedures. At the higher levels of abstraction, our face model offers a semantically rich set of control parameters which reflect the natural constraints of real faces.

Our synthetic facial tissue model is motivated by histology and tissue biomechanics. Human skin has a nonhomogeneous and nonisotropic layered structure consisting of the epidermis, dermis, subcutaneous fatty tissue, fascia, and muscle layers. The synthetic tissue is a deformable lattice, an assembly of discrete finite elements (see [28] for the details).

4.2 Personalizing the Functional Model

We have developed a highly automated approach to constructing realistic, functional models of human heads [29]. These physics-based models are anatomically accurate and may be made to conform closely to specific individuals. Currently, we begin by scanning a subject with a laser sensor which circles the head to acquire detailed range and reflectance information. Next, an automatic conformation algorithm adapts a triangulated face mesh of predetermined topological structure to these data. 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.

The conformed polygonal mesh forms the epidermal layer of a physics-based 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 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 meaningful expressions. To increase realism, we include constraints to emulate tissue incompressibility and to enable the tissue to slide over the skull as real skin does.

Fig. 11 illustrates the aforementioned steps. The figure shows a 360° head-to-shoulder scan of a woman, "Heidi," acquired by a Cyberware Color 3D Digitizer. The data set consists of a radial range map (Fig. 11(a)) and a registered RGB photometric map (Fig. 11(b)). The range and RGB maps are high-resolution 512×256 arrays in cylindrical coordinates, where the x axis is the latitudinal angle around the head and the y axis is vertical distance. Fig. 11(c) shows the generic mesh projected into the 2D cylindrical domain and overlayed on the RGB map. The triangle edges in the mesh are elastic springs, and the mesh has been conformed automatically to the woman's face using both the range and RGB maps [29]. The nodes of the conformed mesh serve as sample points in the range map. Their cylindrical coordinates and the sampled range values are employed to compute 3D Euclidean space coordinates for the polygon vertices. In addition, the nodal coordinates serve as polygon vertex texture map coordinates

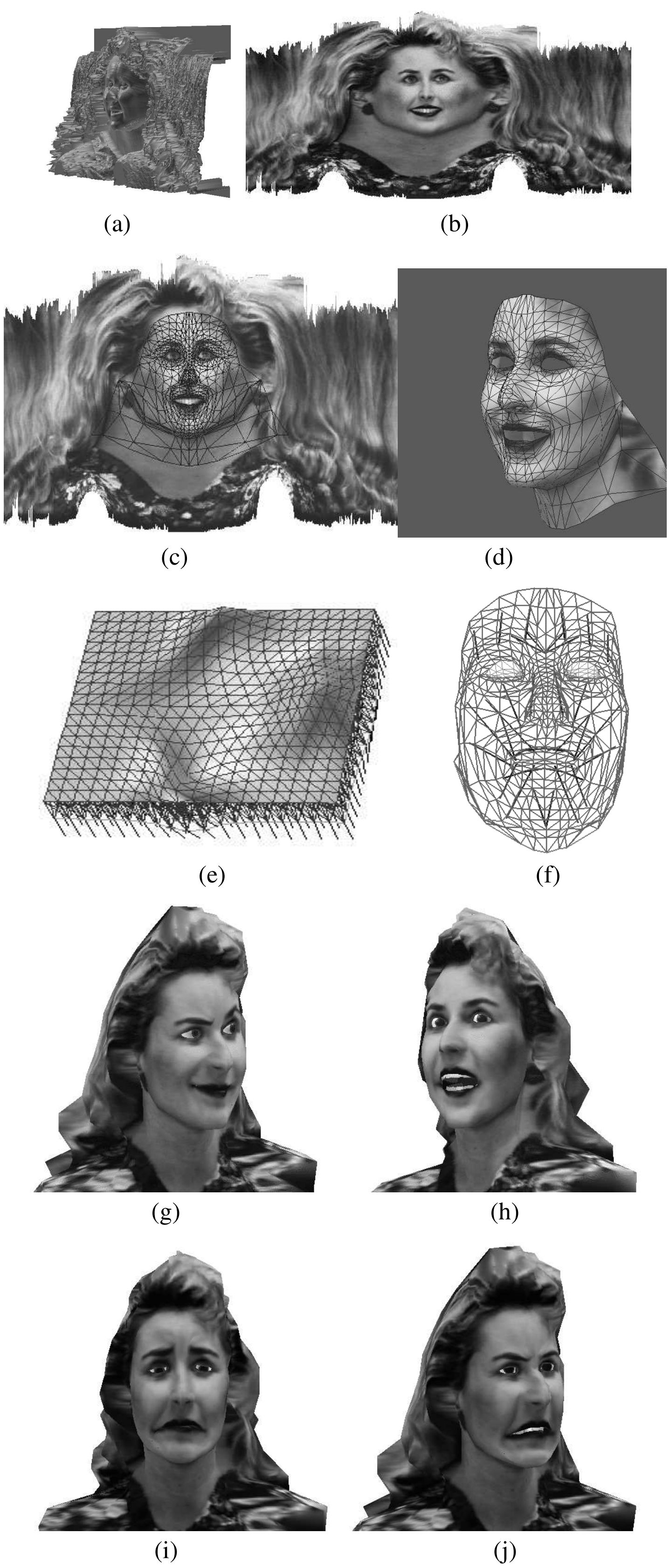


Figure 11: Facial modeling using scanned data. (a) Radial range map. (b) RGB photometric map. (c) RGB map with conformed epidermal mesh overlayed. (d) 3D mesh and texture mapped triangles. (e) Physics-based skin model. (f) Muscles under facial mesh. (g–j) Animate face model.

into the RGB map. Fig. 11(d) shows the 3D facial mesh with the texture mapped photometric data.

Once we have reduced the scanned data to the 3D epidermal mesh of Fig. 11(d), we can assemble a physics-based face model of Heidi, including the synthetic skin (Fig. 11(e) shows a skin patch undergoing deformation) and muscles (Fig. 11(e) shows the contractile muscles (vectors) underneath the epidermal mesh). Fig. 11(g–j) demonstrates the resulting facial model producing animated expressions by contracting facial muscles. The same technique was applied to animate the facial model of George shown in Fig. 2.

5 Model-Based Facial Image Analysis

Facial image analysis and synthesis is necessary for numerous applications. Among them is low bandwidth teleconferencing which may involve the real-time extraction of facial control parameters from live video at the transmission site and the reconstruction of a dynamic facsimile of the subject's face at a remote receiver. Teleconferencing and other applications require facial models that are computationally efficient and also realistic enough to accurately synthesize the various nuances of facial structure and motion. We have argued that the anatomy and physics of the human face, especially the arrangement and actions of the primary facial muscles, provide a good basis for facial image analysis and synthesis [30].

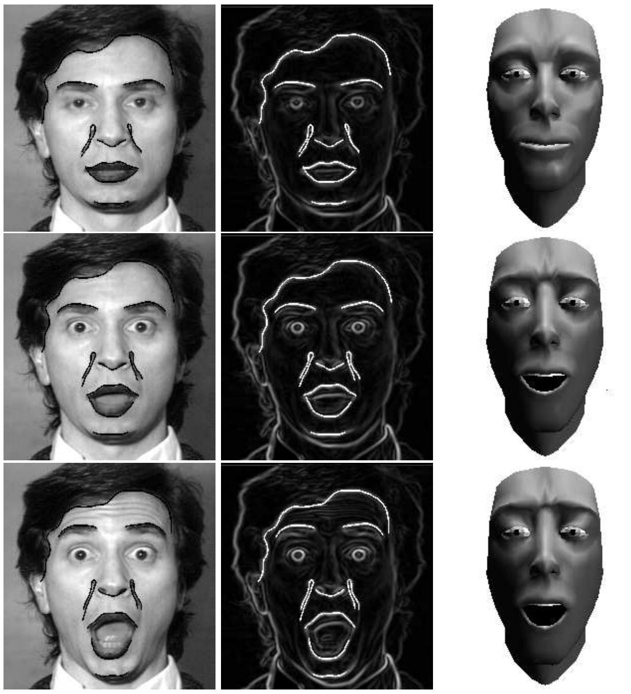
The physics-based anatomically motivated facial model has allowed us to develop a new approach to the analysis of dynamic facial images for the purposes of estimating and resynthesizing dynamic facial expressions [30]. Part of the difficulty of facial image analysis is that the face is highly deformable, particularly around the forehead, eyes, and mouth, and these deformations convey a great deal of meaningful information. Techniques for tracking the deformation of facial features include "snakes" [9]. Motivated by the anatomically consistent musculature in our model, we have considered the estimation of dynamic facial muscle contractions from video sequences of expressive faces. We have developed an analysis technique that uses snakes to track the nonrigid motions of facial features in video. Features of interest include the eyebrows, nasal furrows, mouth, and jaw in the image plane. We are able to estimate dynamic facial muscle contractions directly from the snake state variables. These estimates make appropriate control parameters for resynthesizing facial expressions through a generic face model at real-time rates.

Fig. 13 shows a plot of the estimated muscle contractions versus the frame number. They are input to the physics-based model as a time sequence. The model resynthesizes the facial expression. Three rendered images are shown in Fig. 12(c).

6 Where do we go from here?

I have presented two of our ongoing research projects that span the fields of computer vision, artificial life, and computer graphics. The projects are related in that they involve the development of nontrivial models of living systems. The models are founded upon computational physics. I have demonstrated applications of each of the models to computer vision.

To summarize, on one front, we have made significant progress over the past two years in developing a model which captures the essential features of most living systems—biomechanics, locomotion, perception, behavior, and learning. We are now using this piscatorial model as a situated



(a) (b) (c)

Figure 12: Dynamic facial image analysis and expression resynthesis. Sample video frames with superimposed deformable contours tracking facial features; (a) intensity images with black snakes, (b) image potentials with white snakes. (c) Facial model resynthesizes surprise expression from estimated muscle contractions.

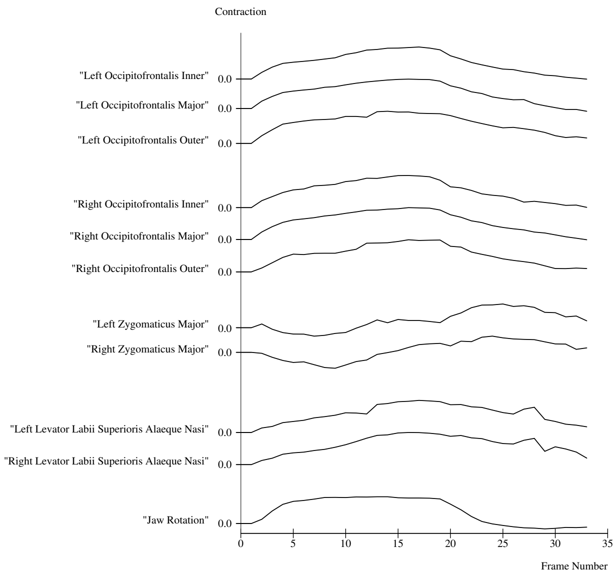


Figure 13: Estimated facial muscle contractions plotted as time series.

virtual robot for active vision research. It is my hope that the active vision systems that we are synthesizing in this way will be relevant in whole or in part to physical robotics. It seems to me that virtual animats in their dynamic world can serve as an useful proving ground for theories that profess sensorimotor competence in animal or robotic situated agents.[4]

On another front, we have been able to generate functional models of people's heads and use them for model-based facial image analysis. At the biomechanical and anatomical levels, the face models are as faithful to human faces as the fish models are to real fishes. Unlike artificial fishes, however, the disembodied artificial heads do not yet have a much of a brain—just a simple motor center that blinks eyelids, moves the eyes, flexes the neck, and coordinates the facial muscles to produce meaningful expressions.

We would like to construct a brain model for the artificial heads that is as at least as comprehensive as the brain of the artificial fish, in the sense that it should be capable of dynamic perception and cognitively motivated behavior depending on environmental influences. We can get to this goal in an interesting way.

An Artificial Mermaid: A logical next step, given what we have implemented already, would be to couple the artificial head model to the posterior of the artificial fish model with an anthropomorphic torso to create an artificial mermaid. The mermaid will be able to locomote through its virtual underwater world as the fishes now do. Using an animat vision system, it will be able to perceive and interact with fish and other mermaids. Unlike fish, however, the mermaid will have some of the expressive and behavioral capabilities of a human. Far from being frivolous, this virtual creature could serve to smoothly bridge the gap between the humble cognitive abilities of an artificial fish and—maybe some day—human level intelligence.

The Artificial Life of a Virtual Human: Naturally, an exciting long-term goal that should elicit little controversy within the AI community is to develop an intelligent artificial human that is at least as convincing on virtual terra firma as our artificial fishes are in their virtual seas. It seems to me that we are well on the way to this end. To achieve this goal, however, more progress will obviously be necessary on several challenging problems, not excluding the AI problem.

To be continued...

[4]**Doom vision:** As a further test of the animat vision paradigm, we are developing an active vision system, similar to the one in artificial fishes, within an autonomous agent situated in a "doom" world ("Doom" is an amazingly popular video game). The challenge is for this agent to assume the role of the human doom player, given the same dynamic graphical image(s) that a human player would see displayed on the screen. The agent's "brain" will interpret the incoming retinal image stream and generate motor commands, analogous to the keyboard commands a human player would issue, to locomote through the amusingly hostile doom world. A successful doom agent would be able to do what a skilled human player does—explore, accumulate points, and avoid being killed.

Acknowledgements

I would like to thank my students for their important contributions to the research described herein. Xiaoyuan Tu developed the artificial fish animat. Tamer Rabie developed the animat vision system. Radek Grzeszczuk developed the animat learning algorithms. Yuenchen Lee developed the physics-based face model. I thank Keith Waters for our years of collaboration on facial modeling. I also thank the many persons who have discussed and debated with me some of the ideas contained herein, especially John Tsotsos (who also provided valuable comments on a draft), Geoffrey Hinton, and Allan Jepson. Range/RGB facial data were provided courtesy of Cyberware, Inc., Monterey, CA. The research described herein was supported by grants from the Natural Sciences and Engineering Research Council of Canada and by the ARK (Autonomous Robot for a Known environment) Project, which receives its funding from PRECARN Associates Inc., Industry Canada, the National Research Council of Canada, Technology Ontario, Ontario Hydro Technologies, and Atomic Energy of Canada Limited. The author is a fellow of the Canadian Institute for Advanced Research.

References

[1] L. G. Roberts. Machine perception of three-dimensional solids. In Tippett et al., editors, *Optical and Electro-Optical Information Processing*, chapter 9, pages 159–197. MIT Press, Cambridge, Massachusetts, 1965.

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

[3] D. Terzopoulos, J. Platt, A. Barr, D. Zeltzer, A. Witkin, and J. Blinn. Physically-based modeling: Past, present, and future. *Computer Graphics*, 23(5):191–209, 1989.

[4] S. W. Wilson. The animat path to AI. In J.-A. Meyer and S. Wilson, editors, *From Animals to Animats*, pages 15–21. MIT Press, Cambridge, MA, 1991.

[5] 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.

[6] X. Tu and D. Terzopoulos. Artificial fishes: Physics, locomotion, perception, behavior. In *Computer Graphics* Proceedings, Annual Conference Series, Proc. SIGGRAPH '94 (Orlando, FL), pages 43–50. ACM SIGGRAPH, July 1994.

[7] J. K. Tsotsos, S. M. Culhane, W. Wai, Y. Lai, N. Davis, and F. Nuflo. Modeling visual attention via selective tuning. *Artificial Intelligence*, 1995. in press.

[8] R. Grzeszczuk and D. Terzopoulos. Automated learning of muscle actuated locomotion through control abstraction. In *Computer Graphics* Proceedings, Annual Conference Series, Proc. SIGGRAPH '95 (Los Angeles, CA), pages 43–50. ACM SIGGRAPH, August 1995.

[9] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[10] J. J. Gibson. *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston, MA, 1979.

[11] R. Bajcsy. Active perception. *Proceedings of the IEEE*, 76(8):996–1005, 1988.

[12] D. Ballard. Animate vision. *Artificial Intelligence*, 48:57–86, 1991.

[13] Y. Aloimonos, A. Bandyopadhyay, and I. Weiss. Active vision. *Int. J. Computer Vision*, pages 333 – 356, 1987.

[14] D.H. Ballard and C.M. Brown. Principles of animate vision. *CVGIP: Image Understanding*, 56(1):3 – 21, July 1992.

[15] A. Blake and A. Yuille, editors. *Active Vision*. MIT Press, Cambridge, MA, 1992.

[16] M.J. Swain and M.A. Stricker (Eds.). Promising directions in active vision. *Int. J. Computer Vision*, 11(2):109 – 126, 1993.

[17] D. Terzopoulos and T.F. Rabie. Animat vision. In *Proc. Fifth International Conference on Computer Vision*, Cambridge, MA, June 1995.

[18] V. Braitenberg. *Vehicles, Experiments in Synthetic Psychology*. MIT Press, Cambridge, MA, 1984.

[19] R. A. Brooks. Elephants don't play chess. In P. Maes, editor, *Designing Autonomous Agents*. MIT Press, Cambridge, MA, 1990.

[20] R.A. Brooks. Intelligence without representation. *Artificial Intelligence*, 47:139–160, 1991.

[21] P. Maes, editor. *Designing Autonomous Agents*. MIT Press, Cambridge, MA, 1991.

[22] R. D. Fernald. Vision. In D. H. Evans, editor, *The Physiology of Fishes*, chapter 6, pages 161–189. CRC Press, Boca Raton, FL, 1993.

[23] M. Swain and D. Ballard. Color indexing. *Int. J. Computer Vision*, 7:11 – 32, 1991.

[24] B. K. P. Horn. *Robot Vision*. MIT Press, Cambridge, MA, 1986.

[25] G. M. Davies, H. D. Ellis, and G. M. Shepherd. *Perceiving and Remembering Faces*. Academic Press, New York, 1981.

[26] V. Bruce. *Recognizing Faces*. Lawrence Erlbaum, Hillsdale, 1988.

[27] P. Ekman and W. V. Friesen. *Manual for the Facial Action Coding System*. Consulting Psychologist Press, Palo Alto, CA, 1977.

[28] D. Terzopoulos and K Waters. Physically-based facial modelling, analysis, and animation. *Journal of Visulization and Computer Animation*, 1:73–80, 1990.

[29] Y. Lee, D. Terzopoulos, and K. Waters. Realistic facial modeling for animation. In *Computer Graphics* Proceedings, Annual Conference Series, Proc. SIGGRAPH '95 (Los Angeles, CA). ACM SIGGRAPH, August 1995.

[30] D. Terzopoulos and K. Waters. Analysis and synthesis of facial image sequences using physical and anatomical models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6):569–579, 1993.