Animat Vision:
Active Vision in Artificial Animals

Demetri Terzopoulos and Tamer F. Rabie

Department of Computer Science, University of Toronto
10 King's College Road, Toronto, ON M5S 3G4, Canada

Abstract

We propose and demonstrate a new paradigm for active vision research that draws upon recent advances in the fields of artificial life and computer graphics. A software alternative to the prevailing hardware vision mindset, *animat vision* prescribes artificial animals, or animats, situated in physics-based virtual worlds as autonomous virtual robots with active perception systems. To be operative in its world, an animat must autonomously control its eyes and muscle-actuated body. Computer vision algorithms continuously analyze the retinal image streams acquired by the animat's eyes, enabling it to locomote purposefully through its world. We describe an initial animat vision implementation within lifelike artificial fishes inhabiting a physics-based, virtual marine world. Emulating the appearance, motion, and behavior of real fishes in their natural habitats, these animats are capable of spatially nonuniform retinal imaging, foveation, retinal image stabilization, color object recognition, and perceptually-guided navigation. These capabilities allow them to pursue moving targets such as other artificial fishes. Animat vision offers a fertile approach to the development, implementation, and evaluation of computational theories that profess sensorimotor competence for animal or robotic situated agents.

Keywords: Active Vision, Artificial Life, Artificial Animals, Autonomous Agents, Physics-Based Modeling, Biomimetic Vision Systems, Foveation, Visual Stabilization, Color Object Detection, Vision-Guided Navigation.

1 Introduction

Advances in the emerging field of artificial life (ALife) make possible a fresh approach to computational vision.[1] A major theme in ALife research is the synthesis of artificial animals, or "animats" [1]. Animats, a term coined by Wilson [2], are computational models of real animals situated in their natural habitats. A recent breakthrough in animat research has produced situated virtual agents that realistically emulate animals of nontrivial evolutionary complexity [3]. This advance prompts us to propose *animat vision*, a paradigm which prescribes the use of artificial animals as autonomous virtual robots for active vision research.[2]

Our zoomimetic approach to vision is made possible by the confluence of three recent trends:

1. advanced physics-based artificial life modeling of natural animals,
2. photorealistic computer graphics rendering and its efficient implementation in modern 3D graphics workstations, and
3. active computer vision algorithms.

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 individual and social functionality within these worlds. To this end, each animat is an autonomous agent possessing a muscle-actuated body capable of locomotion and a mind with perception, motor, and behavior centers. The animat is endowed with functional eyes that can image the dynamic 3D virtual world onto 2D virtual retinas. The perceptual center of the animat's brain exploits active vision algorithms to continually process the incoming stream of dynamic retinal images in order to make sense of what it sees and, hence, to purposefully navigate its world.

The uninitiated reader may doubt the possibility of implementing artificial animals rich enough to support serious active vision research. Fortunately, this hurdle has already been cleared. Recent animat theory encompasses the physics of the animal and its world, its ability to locomote using internal muscles, its adaptive, sensorimotor behavior, and its ability to learn. In particular, an

[1] For an engaging introduction to the ALife field, see, e.g., S. Levy, *Artificial Life* (Pantheon, 1992).

[2] The animat vision paradigm was introduced in an earlier version of this paper [4] and it is further developed in [5].

Figure 1: Artificial fishes in their physics-based virtual world as it appears to an underwater observer. The 3 reddish fish (center) are engaged in mating behavior, the greenish fish (upper right) is a predator hunting for small prey, the remaining 3 fishes are feeding on plankton (white dots). Seaweeds grow from the ocean bed and sway in the current.

animat with these essential capabilities has been implemented that emulates animals as complex as teleost fishes in their marine habitats [3, 6].

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, 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.

The challenge undertaken in this paper is to synthesize an active vision system for the fish animat which is based solely on retinal image analysis. The vision system should be extensible so that it will eventually support the broad repertoire of individual and group behaviors described above. It is important to realize that we need not restrict ourselves to modeling the perceptual mechanisms of natural 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 develop in this paper makes no attempt to model fish perception [7]. Instead, we have found it an interesting and challenging problem to endow the piscine animat with a biomimetic vision system that enables it to be a functional, active observer of its world.

The basic functionality of the animat vision system starts with binocular perspective projection of the 3D world onto the animat's 2D retinas. Retinal imaging is accomplished by photorealistic, color graphics rendering of the world from the animat's viewpoint. 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 the eyes in order to stabilize the visual fields during locomotion through compensatory eye movements (an optokinetic reflex), to attend to interesting colored targets, and to keep these dynamic targets fixated. The artificial fish is thus able to approach and track other artificial fishes under visual guidance. Eventually its arsenal of active vision algorithms will enable it to forage, evade predators, find mates, etc.

The remainder of the paper is organized as follows: Section 2 briefly motivates our approach vis-a-vis conventional active vision based on robot hardware and discusses the background of our work. Section 3 reviews the relevant details of the artificial fish model. Section 4 describes the active vision system that we have implemented in the animat, including the retinal imaging, foveation, color object detection, and retinal field stabilization algorithms. Section 5 presents results in vision-guided navigation and the pursuit of moving nonrigid targets—other artificial fish. Section 6 presents conclusions and discusses future research.

2 Motivation and Background

2.1 Problems with the Hardware Vision Mindset

Active vision research in most labs today is in reality the technologically driven pursuit of "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 mounts, 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 working order.

The animat vision methodology that we advocate in this paper can potentially liberate a significant segment of the computer vision research community from the tyranny of hardware. It addresses the needs of scientists who are motivated to understand and ultimately reverse engineer the powerful vision systems that nature has evolved in higher animals. These researchers realize that readily available hardware systems are woefully inadequate models of natural animals—clearly, animals do not have CCD chip eyes, electric motor muscles, and wheel legs. Moreover, their mobile robots typically lack the compute power necessary to achieve real-time response within a fully dynamic world while running active vision algorithms of much sophistication. Yet in their ambition to understand the complex sensorimotor functions of real animals, active vision researchers have been forced to struggle with whatever hardware is available to them, for lack of a better research strategy; that is, until now.

2.2 Further Advantages of the Animat Vision Approach

Animat vision offers an alternative research strategy for developing biologically inspired active vision systems. It 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 readily available 3D graphics workstations. Animat vision offers several additional advantages:

- One can arbitrarily 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 retarding the agent's responses relative to the temporal evolution of its virtual world. This in turn permits the development and evaluation of sophisticated new 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 no environmental ground truth data, but must glean visual information "the hard way"—from 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.[3]

2.3 Related Work

It should by now be clear to the reader that our goal in this paper is *not* to model the vision systems of real fishes, but instead to employ the artificial fish as a virtual robot for designing general-purpose active vision systems.[4]

J.J. Gibson [9], in a sense the grandfather of active vision, stresses in pre-computational terms the importance of modeling the active observer situated in the dynamic environment. Versions of this paradigm suitable for mainstream computer vision were introduced in the seminal papers of Bajcsy [10] and Ballard [11] under the names of active perception and animate vision, respectively.[5] The active vision approach was further developed by Aloimonos *et al.* [12] and many others (see, e.g., [13, 14, 15]) into the prevalent paradigm that it is today.

The artificial animals that we advocate in this paper are active "vehicles" in the sense of Braitenberg [16]. 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 [17] and have been an inspiration to numerous other robotics groups (see, e.g., the compilation [18]). 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

[3] It is often convenient to represent ground truth data iconically in the form of retinocentric intrinsic images, including intensity, range, illumination, reflectance, and object identity images, and these can be computed easily and quickly by the rendering pipelines of 3D graphics workstations.

[4] However, it is interesting that from the point of view of vision science, fishes, which account for some 22,000 species inhabiting almost every conceivable photic environment from the clear waters of tropical lagoons to the darkest deep seas, provide ideal models for the vertebrate visual system. Fish visual systems have been open to a wider variety of investigations, from biochemistry to ecology, than any other visual system, and many characteristics of vertebrate vision were initially demonstrated using fish (see [8]).

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

processors.

Artificial fishes are animats of unprecedented sophistication. They are autonomous virtual robots situated in a 3D continuous virtual world governed by physical dynamics. This makes them suitable for grounding active vision systems. By contrast, Wilson's original animat [2], which was proposed for exploring the acquisition of simple behavior rules, is a point marker on a non-physical 2D grid world that can move between squares containing food or obstacles. Other simple animats include the 2D cockroaches of Beer [19]. A more sophisticated animat is the kinematic dog described by Blumberg and Galyean [20]. Prior animats make use of "perceptual oracles"—schemes for directly interrogating the virtual world models to extract sensory information about the environment as needed by the animat. One can also find several instances of "oracle vision" in the behavioral animation literature [21, 22, 23]. Unlike oracle vision, for the animat vision approach to make sense, it is absolutely necessary that the animat and its world attain a high standard of visual fidelity.

3 Review of the Fish Animat

The artificial fish model is developed elsewhere [3, 6, 23]. This section reviews the animat to a level of detail that suffices to comprehend the remainder of the paper.

Each artificial fish is an autonomous agent with a deformable body comprising a graphical display model and a biomechanical model actuated by internal muscles. As Fig. 2 illustrates, the body also includes eyes (among other on-board sensors) and a brain with motor, perception, behavior, and learning centers. Through controlled muscle actions, artificial fishes are able to swim in simulated water in accordance with simplified 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 within a repertoire of piscine relevant behaviors, including collision avoidance, foraging, preying, schooling, and mating.

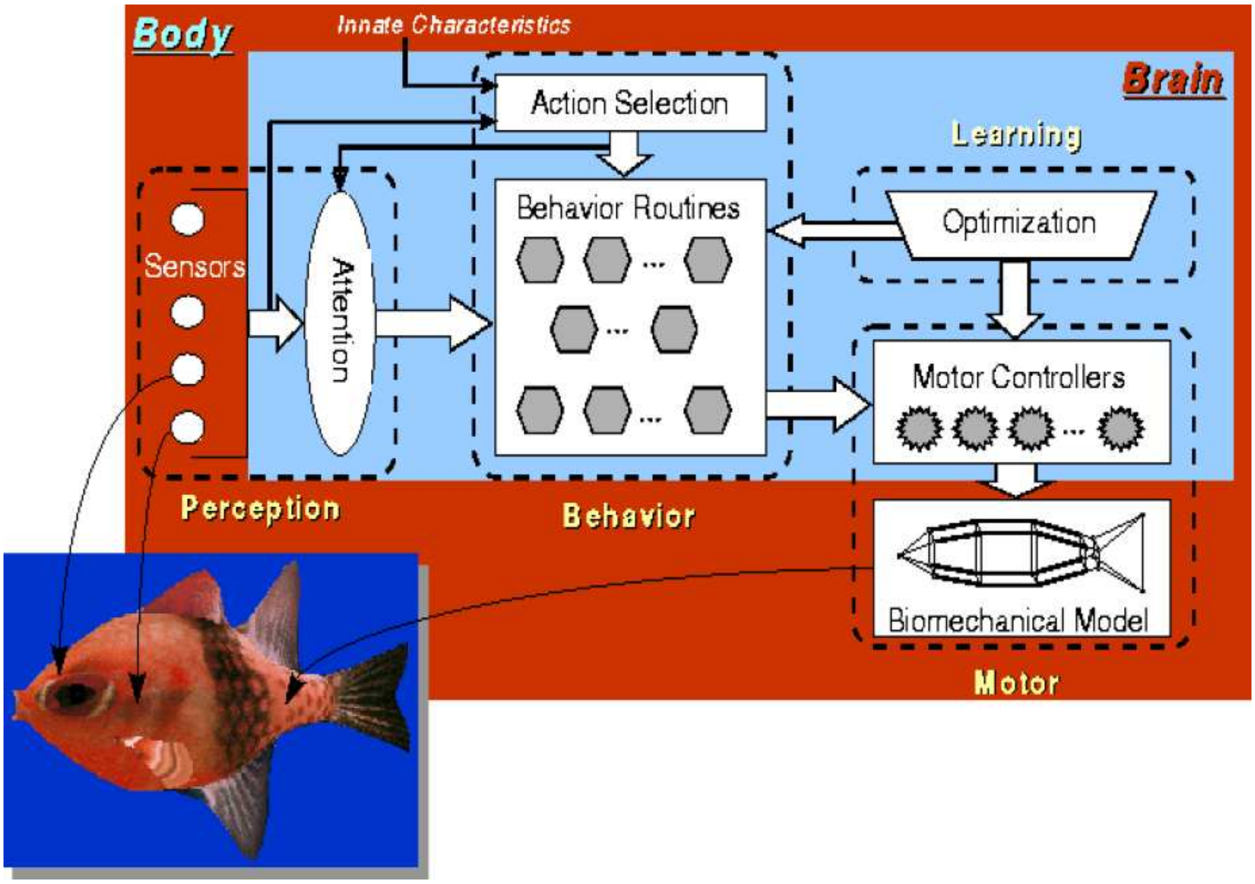
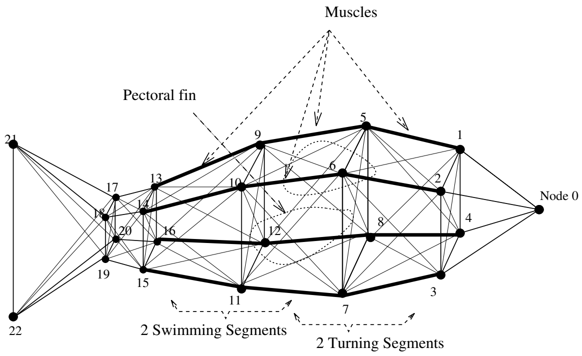


Figure 2: The body of an artificial fish comprises a muscle-actuated biomechanical model, perceptual sensors, and a brain with motor, perception, behavior, and learning centers. To the lower left is an artificial fish graphical display model.

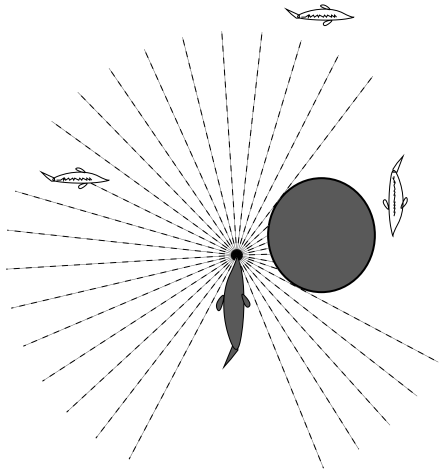
3.1 Motor System

The motor system (see Fig. 2) comprises the fish biomechanical model, including muscle actuators and a set of motor controllers (`MCs`). Fig. 3(a) illustrates the mechanical body model which produces realistic piscine locomotion using only 23 lumped masses and 91 uniaxial viscoelastic elements, 12 of which are actively contractile muscle elements. These mechanical components are interconnected so as to maintain the structural integrity of the body as it flexes due to the muscle actions.

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



(a)



(b)

Figure 3: Biomechanical fish model (a). Nodes denote lumped masses. Lines indicate springs (shown at their natural lengths). Bold lines indicate muscle springs. 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 animat at the center).

continually integrated through time by a numerical simulator.[6]

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 which enable it to navigate freely in its 3D world by pitching, rolling, and yawing its body. Additional motor controllers coordinate the fin actions.

3.2 Perception System

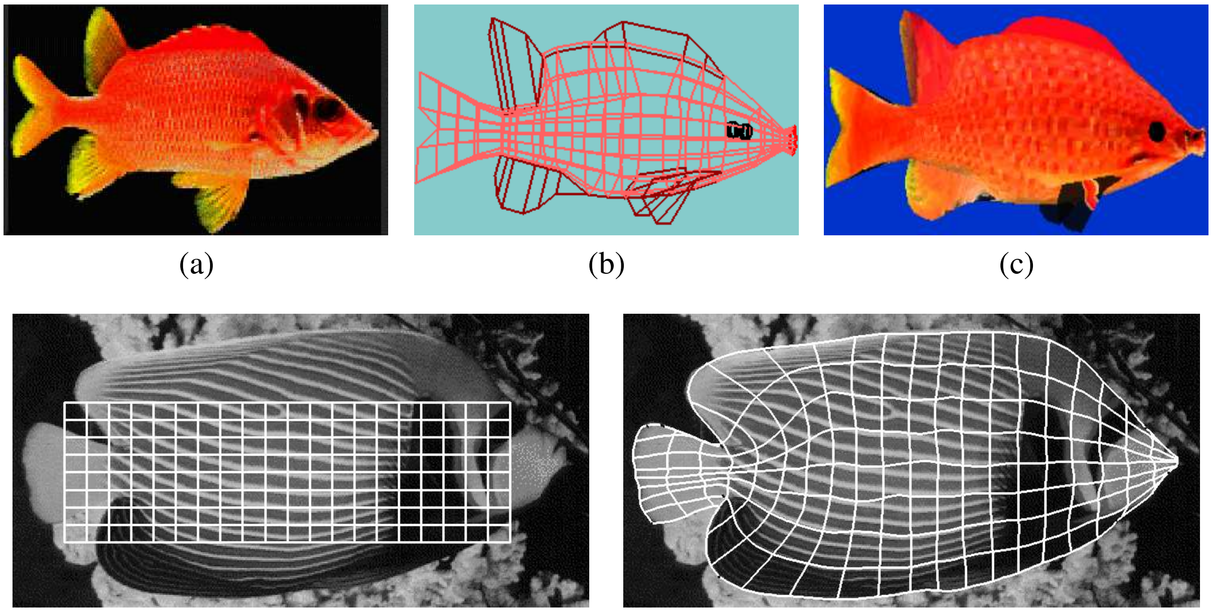
Artificial fishes gain awareness of their world through sensory perception. As Fig. 3(b) suggests, it is necessary to model not only the abilities but also the limitations of animal perception systems in order to achieve natural sensorimotor behaviors. Hence, the artificial fish has a limited field

[6]The artificial fish model achieves a good compromise between realism and computational efficiency. To give an example simulation rate, 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.

of view extending frontally and laterally to an effective radius consistent with visibility in the translucent water (Fig. 3(b)). An object may be detected only if some visible portion of it (i.e., not occluded behind some other opaque object) enters the fish's field of view (Fig. 3(b)). The perception center of the artificial fish's brain (see Fig. 2) includes a perceptual attention mechanism which allows the animat to attend to the world in a task-specific way, hence filtering out sensory information superfluous to its immediate behavioral needs. For example, the artificial fish attends to sensory information about nearby food sources when foraging. The animats in our previous ALife simulations (described in [3, 6, 23]) employ a perceptual oracle scheme according to which the artificial fish may satisfy its perceptual needs via direct interrogation of the 3D world model. In particular, subject to the appropriate perceptual limitations, the animat's on-board sensors query the geometric and photometric information that is available to the graphics rendering engine, as well as object identity and dynamic state information within the physics-based virtual world. We emphasize that our goal in the present paper is to replace the perceptual oracle with an artificial fish active vision system that elicits visual information from retinal images, as is described in 4.

3.3 Behavior System

The behavior center of the artificial fish's brain mediates between its perception system and its motor system (Fig. 2). A set of innate characteristics determines the (static) genetic legacy, which dictates whether the fish is male or female, predator or prey, etc. A (dynamic) mental state comprises variables representing hunger, fear, and libido, whose values depend on sensory inputs. The fish's cognitive faculty resides in the action selection component of its behavior center. At each simulation time step, action selection entails combining the innate characteristics, the mental state, and the incoming stream of sensory information to generate sensible, survival sustaining goals for the fish, such as to avoid an obstacle, to avoid predators, to hunt and feed on prey, or to court a potential mate. The action selector ensures that goals have some persistence by exploiting a single-item memory. The behavior memory reduces dithering, thereby improving the robustness of prolonged behaviors such as foraging, schooling, and mating. The action selector also controls the perceptual attention mechanism. At every simulation time step, the action selector activates behavior routines that attend to sensory information and compute the appropriate motor control parameters to carry the fish a step closer to fulfilling its immediate goals. The behavioral repertoire



(a) (b) (c)

(d) (e)

Figure 4: (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-mesh on an image of a different fish.

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 is dependent on the mental state.

3.4 Modeling Form and Appearance

Animat vision requires that artificial animals capture the form and appearance of natural animals with considerable visual fidelity. To this end, photographs of real fishes, such as the one shown in Fig. 4(a), are converted into 3D spline (NURBS) surface body models (Fig. 4(b)). The digitized photographs are analyzed semi-automatically using deformable models [24], in particular, a "snake-mesh" tool which is demonstrated in Fig. 4(d–e) on a different fish image. The snake mesh slides freely over the image and can be manipulated using the mouse. 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 with a smooth, nonuniform coordinate system (Fig. 4(e)). The coordinate system serves to map the appropriate image texture onto the spline surface to produce the final texture mapped fish body model (Fig. 4(c)).

4 The Animat Vision System

In this section we present the animat vision system that we have developed for the artificial fish, which makes exclusive use of color retinal images.

4.1 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 measured 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. The level $l = 0$ camera has the widest field of view (about 120°) and the horizontal and vertical fields of view for the level l camera are related by

$$f_x^l = 2\tan^{-1}\left(\frac{d_x/2}{2^l f_c^0}\right); \qquad f_y^l = 2\tan^{-1}\left(\frac{d_y/2}{2^l f_c^0}\right), \tag{1}$$

where d_x and d_y are the horizontal and vertical image dimensions and f_c^0 is the focal length of the wide field of view camera ($l = 0$).[7]

Fig. 5(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. Since the field of view decreases with increasing l, image l is a zoomed version of the central part of image $l - 1$. We refer to the image at level $l = 3$ as the fovea and the others as peripheral images. We magnify the level l image by a factor of 2^{3-l} and overlay in sequence the four images coincident on their centers starting with the $l = 0$ image at the bottom (to form an (incomplete) pyramid), thus compositing a 512×512 retinal image with a 64×64 fovea at the center of a periphery with radially decreasing resolution (and increasing smoothing) in 3 steps. Fig. 5(b) shows the 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.

[7] If f_c^0 is unknown, but the $l = 0$ field of view is known, then f_c^0 is first computed using (1) with $l = 0$ and this value is used to specify the field of view at the other levels.

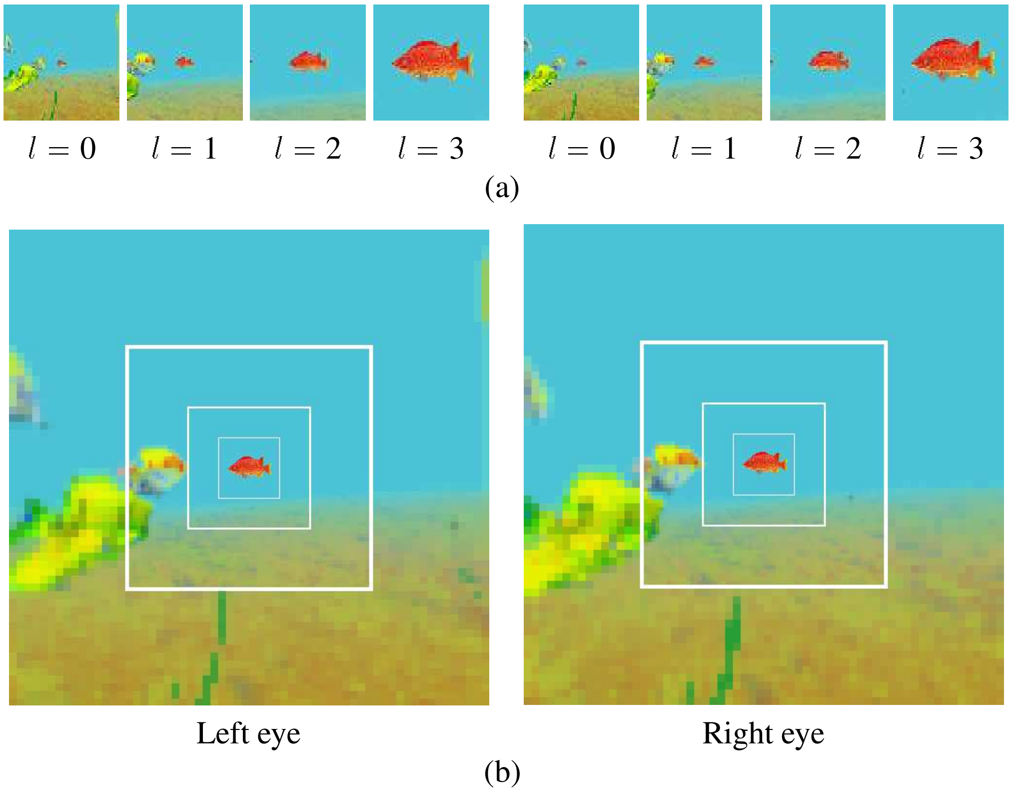


Figure 5: Binocular retinal imaging. (a) 4 component images; $l = 0, 1, 2,$ are peripheral images; $l = 3$ is foveal image. (b) Composited retinal images (borders of composited component images are shown in white).

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 would have to process a uniform 512×512 retinal image.

4.2 Active Vision System Overview

Fig. 6 illustrates a block diagram of one ocular channel of the binocular animat vision system. The system currently consists of two main modules—a *foveation* module and *stabilization* module. Together they implement a gaze control capability that enables the artificial fish to stabilize the visual world as it locomotes, as well as to detect a visual target in its field of view, foveate the target, and visually navigate towards the target. If the target is in motion, the artificial fish can track it visually and swim in pursuit.

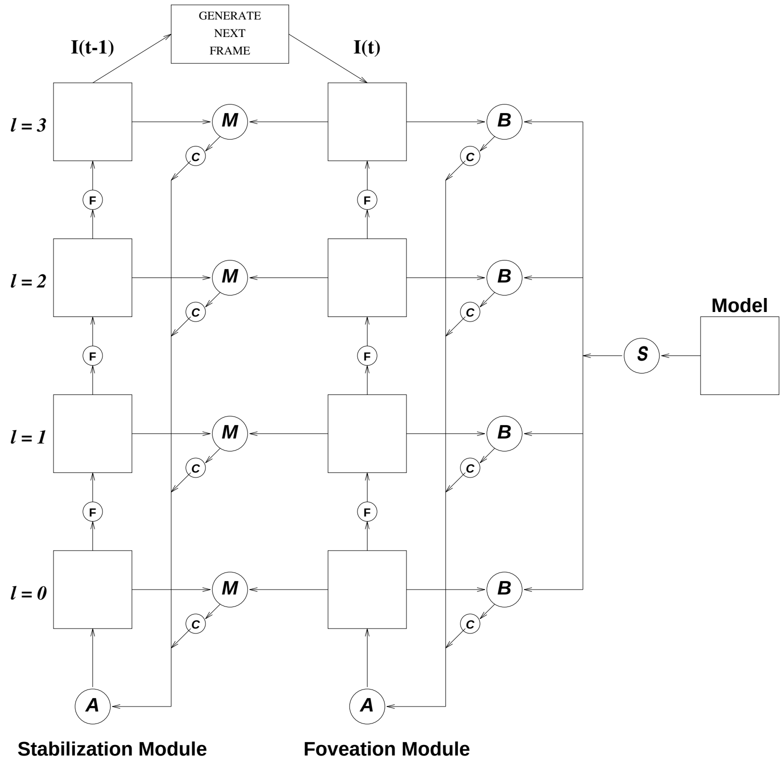


Figure 6: Gaze control for one eye in the animat vision system. The flow of the algorithm is from right to left. A: Update gaze angles (θ, ϕ) and saccade using these angles, B: Search current level for model target and if found localize it, else search lower level, C: Select level to be processed (see 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 model's color histogram for use by the current level.

4.3 Foveation using 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 a predator, it would possess models of prey fish. The models are stored as a list of 64×64 color images in the fish's memory.

We have adopted into the active vision system the color histogram methods of Swain [25]. The fish employs these methods to detect and localize any target that may be imaged in the low resolution periphery of its retinas. Since each model object has a unique color histogram (when the background is subtracted from the object) it can be detected in the periphery by histogram intersection and localized by histogram backprojection.

4.3.1 Modified Color Histogram Intersection Method

Swain [25] developed a technique called color indexing that efficiently identifies objects in a database in the presence of occlusion and over changes in viewpoint. He demonstrated that object color distributions without geometric information can provide a powerful cue for recognition.

The effectiveness of the algorithm degrades badly if the area of the object in the model image differs substantially from the area of the target object appearing in the image. Swain suggests scaling the initial model histogram by d_M^2/d^2 where d_M is the known range of the model initially and d is the computed range of the target object at the time of backprojection. We will show later that estimating range is straightforward with the eyes verged on a target. Unfortunately, this scaling technique did not work well for the artificial fish, apparently because of noisy depth measurements and the perspective nonlinearity associated with the wide field of view cameras.

We developed a more robust intersection measure that is invariant to scale changes. This new method iteratively scales down an initially large model color histogram to the approximate size of the target object appearing in the image.

Following Swain's notation in [25], his original histogram intersection measure is

$$H = \frac{\sum_{j=1}^{n} \min(I_j, M_j)}{\sum_{j=1}^{n} M_j}, \tag{2}$$

where I is the image color histogram, M is the model color histogram, and n is the number of histogram bins used. This measure is effective only if the model histogram is properly scaled. To overcome this limitation, note that the match value H gives the percentage of pixels from the model that have corresponding pixels of the same color in the image (hence, $H = 0.9$ means that there is 90% chance the model appeared in the image; however, the value of H can drop significantly; e.g. $H = 0.1$ if there is a scale difference between target and model). This suggests that we can use H itself to scale the model histogram M. Our experiments revealed that this is not always effective, but that it can be improved by scaling the histogram iteratively; i.e., recomputing H after every scaling of M until the value of H increases above a set threshold, say 0.9. This technique may be expressed as

$$M_k^{i+1} = M_k^i H^i = M_k^i \frac{\sum_{j=1}^{n} \min(I_j, M_j^i)}{\sum_{j=1}^{n} M_j^i}; \quad k = 1, \ldots, n, \tag{3}$$

where i is the iteration number. Equation 3 is iterated until the value of H^i either exceeds the threshold, indicating the presence of the model in the image, or remains constant below threshold (or decreases), indicating that the model matches nothing in the image. The equation converges after a few iterations (usually 2 to 4 if the target size is not too much smaller than the model).

The iterative technique may degenerate in cases when the model is not present in the image, while a similar color combination is. The problem is that the model histogram gets scaled to the size of the false target to yield a large intersection match value, hence a false alarm.

To overcome this problem, we employ a new intersection measure after scaling the model histogram using (3). Our measure makes use of a weighted histogram intersection method inspired by the local histogram method proposed by Ennesser and Medioni [26]. Our measure is

$$H_N = \frac{\sum_{j=1}^{n} W_j \min(I_j, M_j)}{\sum_{j=1}^{n} M_j}, \tag{4}$$

where the weighting histogram W is given by $W_j = M_j/2^l P_j$. Here P is the color histogram of the peripheral image (at level $l = 0$). As is noted by Ennesser and Medioni, the weighted intersection technique increases the selectivity of the method by placing more importance on colors that are specific to the model. In our experiments, H_N provided very good separation between intersection match values for false targets ($H_N < 0.2$) and true targets ($H_N > 0.8$).

An alternative method, which also gives good results is to incorporate the weighting histogram inside the iteration of equation (3) as follows:

$$M_k^{i+1} = M_k^i \frac{\sum_{j=1}^{n} (M_j^i/2^l P_j) \min(I_j, M_j^i)}{\sum_{j=1}^{n} M_j^i}. \tag{5}$$

The scaled M is then used iteratively to compute the intersection match value H_N as before.

4.3.2 Localization using Color Histogram Backprojection

Once the model histogram has been properly scaled as described above, Swain's backprojection algorithm works well in localizing the pixel position of the center of the detected target in the foveal image. A thorough description of this algorithm is available in [25].

4.3.3 Saccadic Eye Movements

When a target is detected in the visual periphery, the eyes will saccade to the angular offset of the target 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.

When the fish is executing a rapid turn, however, the target could partially exit the fovea ($l = 3$). Part of it will appear in the next coarser image ($l = 2$). Three saccades are then typically used to regain a foveal fix on the target. The first saccade detects the portion of the target still in the fovea and makes an initial attempt to foveate the target, on the second saccade the target is brought closer to the center of the fovea, and finally the third saccade accurately centers the target in both foveas to verge the eyes.

Module *A* in Fig. 6 performs the saccades by incrementing the gaze angles (θ, ϕ) with differential angles $(\Delta\theta, \Delta\phi)$ in order to rotate the eyes to the required gaze direction. When the pixel location of the target computed from the left or right images at level l is (x_c, y_c), the correction gaze angles for the eye are given by

$$\Delta\theta = \tan^{-1}\left(\frac{x_c}{2^l f_c}\right); \qquad \Delta\phi = \tan^{-1}\left(\frac{y_c}{2^l f_c}\right). \tag{6}$$

If the target object comes too near the eyes and fills the entire fovea, the algorithm foveates the target at the next coarser level (e.g., $l = 2$), where the field of view is broader and the target has a more reasonable size for detection and localization. Note that (6) computes the correction angles at level l, but the same corrected (θ, ϕ) are used to render all the other levels.

4.4 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. The optokinetic reflex in animals stabilizes vision by measuring image motion and producing compensatory eye movements. Once a target is verged in both foveas, the stabilization process assumes the task of keeping the target foveated as the fish locomotes.

Stabilization is achieved by computing the displacement (u, v) between the current and previous foveal images and updating the gaze angles to $(\theta + \Delta\theta, \phi + \Delta\phi)$. The displacement is computed as a translational offset in the retinotopic coordinate system by a least squares minimization of the optical flow constraint equation between image frames at times t and $t-1$ [27, 28]. The flow constraint equation is given by [29]

$$uI_x + vI_y + I_t = 0, \tag{7}$$

where I_x, I_y, and I_t are the derivatives of the image intensity

$$I(x,y,t) = \frac{1}{3}[R(x,y,t) + G(x,y,t) + B(x,y,t)], \tag{8}$$

where R, G, and B denote the color component channels. The error function

$$E(u,v) = \sum_{x,y \in \text{fovea}} (uI_x + vI_y + I_t)^2 \tag{9}$$

is minimized by simultaneously solving the two equations $\partial E/\partial u = 0$ and $\partial E/\partial v = 0$ for the image displacement (u,v).

The correction angles $(\Delta\theta, \Delta\phi)$ for a displacement of (u,v) between the images at level l are computed using (6) by replacing (x_c, y_c) with (u,v). If the displacement computed from the foveal images (at level $l = 3$) is too small (indicating the target is close enough to fill the fovea), the algorithm stabilizes at the next lower level where the target does not fill the entire image area.

The flow constraint displacement estimation method is accurate 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 fix is regained by invoking the foveation module 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 eyes must be correlated to keep them verged accurately on the target and not drifting in different directions. The correlation is performed by computing the displacement (u,v) between the left and right foveal images (at $l = 3$), and correcting the gaze angles of the right eye to $(\theta_R + \Delta\theta_R, \phi_R + \Delta\phi_R)$ using (6).

Once the eyes are verged on a target, it is straightforward for the active vision system to estimate

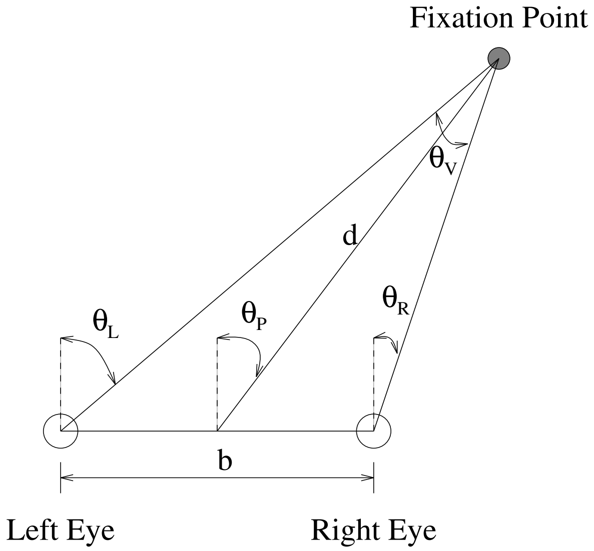


Figure 7: Gaze angles and range to target geometry.

the range to the target from the gaze angles. Referring to Fig. 7, the range is [29]

$$d = b\frac{\cos(\theta_R)\cos(\theta_L)}{\sin(\theta_L - \theta_R)\cos(\theta_P)}, \tag{10}$$

where b is the baseline between the two eyes, and $\theta_P = \frac{1}{2}(\theta_R + \theta_L)$ is the left/right turn angle. When the eyes are verged on a target the vergence angle is $\theta_V = (\theta_R - \theta_L)$ and its magnitude increases as the fish comes closer to the target [30].

5 Vision-Guided Navigation

The artificial fish can employ the direction of gaze of its eyes to effectively navigate 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. 7, and the ϕ angles are similarly used to compute an up/down turn angle $\phi_P = \frac{1}{2}(\phi_R + \phi_L)$. The fish's turn motor controllers are invoked to execute a left/right turn—`left-turn-MC` for negative θ_P and `right-turn-MC` for positive θ_P (see Section 3)—with $|\theta_P|$ as parameter when $|\theta_P| > 30°$. An up/down turn command is issued to the fish's pectoral fins if $|\phi_P| > 5°$, with a positive ϕ_P interpreted as up and negative as down.

5.1 Pursuit of Nonrigid Targets in Motion

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

We have carried out numerous experiments in which the moving target is a reddish fish whose color histogram model is stored in the memory of a predator fish equipped with the active vision system. Fig. 8 shows plots of the gaze angles and the turn angles obtained over the course of 100 frames in a typical experiment as the predator fixates on and actively pursues a prey target. Fig. 9 shows a sequence of image frames acquired by the observer fish during its navigation (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 saccaded to foveate the target (note that the size decrease of the target after foveation is a perspective effect). 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. Fixation is achieved by stabilizing the eyes with compensating saccade signals. The signals are indicated in Fig. 8 by the undulatory responses of the θ angles.

Fig. 8 shows that the vergence angle tends to increase in magnitude as the fish moves closer to the target (around 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; that is, (θ_R, ϕ_R) follows (θ_L, ϕ_L), with $\theta_R - \theta_L$ indicating the reciprocal of range to the target.

Notice that in frames 87–117 of Fig. 9, a yellow fish whose size is similar to the target fish passes behind the target. In this experiment the fish with active vision was instructed to treat all non-reddish objects as totally uninteresting and not worth foveating. 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.

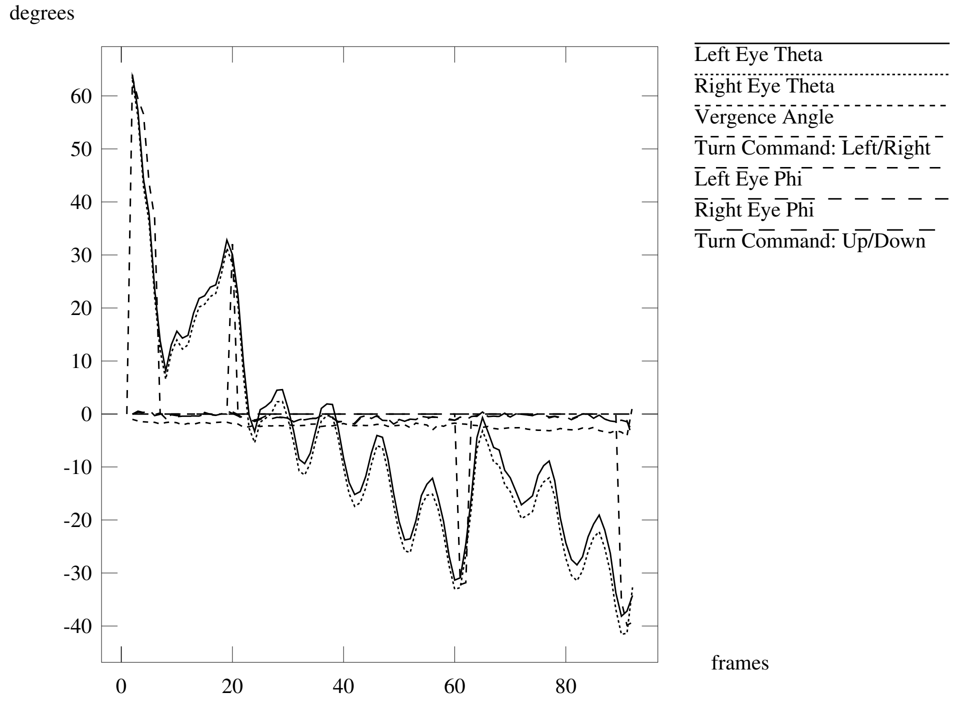


Figure 8: Gaze angles (saccade signals) vs time (frames) of the observer fish while pursuing the reddish target fish.

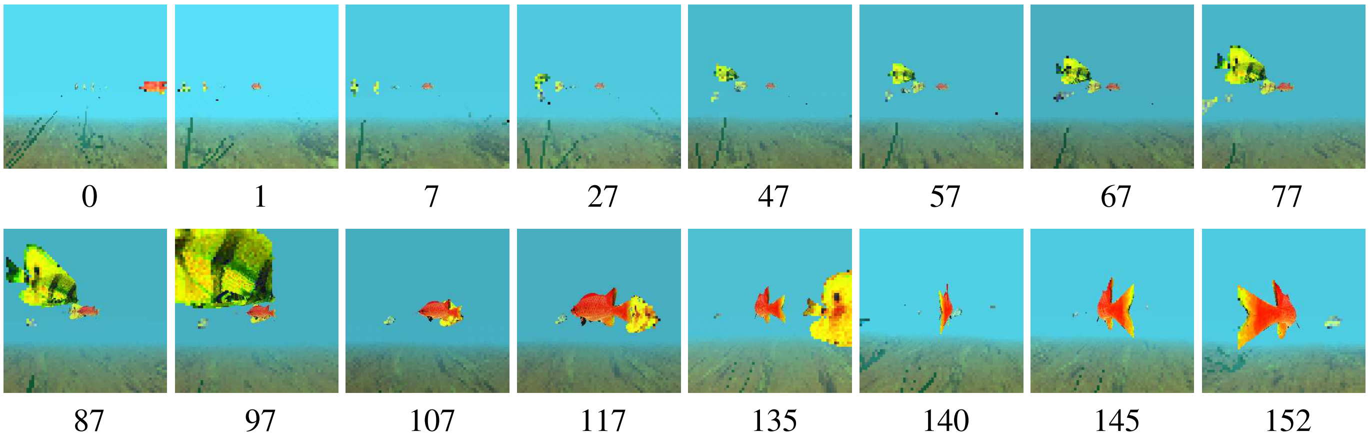


Figure 9: Retinal image sequence from the left eye of the active vision fish as it detects and foveates on a reddish fish target and swims in pursuit of the target. The target appears in the periphery (middle right) in frame 0 and is foveated in frame 1. The target remains fixated in the center of the 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).

6 Conclusion and Future Work

This paper has presented work that spans the fields of computer vision, artificial life, and computer graphics. Our research was motivated in part by the realization that many active vision researchers would rather not have their progress impeded by the limitations and complications of currently available hardware. Animat vision offers a viable, purely software alternative to the prevailing hardware vision mindset.

To demonstrate the animat vision approach, we have employed a physics-based, virtual marine world inhabited by lifelike artificial fishes that emulate the appearance, motion, and behavior of natural fishes in their physical habitats. Artificial fishes are virtual zoological robots ("zoobots") that offer active vision researchers greater mobility and maneuverability, lower cost, and higher reliability/repeatability than can be expected from present-day physical robots. Moreover, these physics-based virtual robots are governed in their virtual world by the same principles that physical robots are subject to in the physical world, hence they share the attraction of situated physical robots for the purposes of active vision research.

In a relatively short period of time we have successfully implemented within the framework of the artificial fish animat a set of active vision algorithms for foveation and vergence of interesting targets, for retinal image stabilization, and for pursuit of moving targets through visually-guided navigation. Note, however, that the automated analysis of the class of retinal images that confront our vision algorithms is by no means easy.

In future work we will endeavor to develop a more extensive arsenal of active vision algorithms to support the complete behavioral repertoire of artificial fishes. The animat vision approach allows us to do this in stages without ever compromising the complete functionality of the artificial fish. We anticipate that the active vision suite that we are developing will be relevant in whole or in part to physical robotics (e.g., autonomous underwater vehicles). In conclusion, it appears that artificial animals in their virtual worlds can serve as a proving ground for theories that profess sensorimotor competence in animal or robotic situated agents.

Acknowledgements

This research was made possible by the creativity, dedication, and cooperation of Xiaoyuan Tu, who developed the artificial fish animat. We thank her and Radek Grzeszczuk for their many important contributions to the artificial fishes project. We also thank the many persons who have discussed and debated the animat vision idea with us, especially John Tsotsos, Geoffrey Hinton, and Allan Jepson. The work was supported by a research grant 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. DT is a fellow of the Canadian Institute for Advanced Research.

References

[1] P. Husbands, editor. *From Animals to Animats: The 3rd International Conf. on Simulation of Adaptive Behavior*, Cambridge, MA, 1994. MIT Press.

[2] 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.

[3] 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.

[4] D. Terzopoulos and T.F. Rabie. Animat vision. In *Proc. Fifth International Conference on Computer Vision*, pages 801–808, Cambridge, MA, June 1995. IEEE Computer Society Press.

[5] T.F. Rabie and D. Terzopoulos. Motion and color analysis for animat perception. In *Proc. 13th National Conference on Artificial Intelligence*, pages 1090–1097, Portland, OR, August 1996. Morgan Kaufmann.

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

[8] E.R. Loew and W.N. McFarland, editors. *Fish Vision*, volume 31(3) of *Vision Research*, 1991.

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

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

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

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

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

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

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

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

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

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

[19] R. Beer. *Intelligence as Adaptive Behavior*. Academic press, NY, 1990.

[20] B. M. Blumberg and T. A. Galyean. Multi-level direction of autonomous creatures for real-time virtual environments. In *Computer Graphics* Proceedings, Annual Conference Series, Proc. SIGGRAPH '95 (Los Angeles, CA), pages 47–55. ACM SIGGRAPH, August 1995.

[21] C. W. Reynolds. Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics*, 21(4):25–34, 1987.

[22] O. Renault, N. Magnenat-Thalmann, and D. Thalmann. A vision-based approach to behavioural animation. *Visualization and Computer Animation*, 1:18–21, 1990.

[23] X. Tu and D. Terzopoulos. Perceptual modeling for the behavioral animation of fishes. In *Proc. 2nd Pacific Conf. on Computer Graphics*, Beijing, China, 1994.

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

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

[26] F. Ennesser and G. Medioni. Finding waldo, or focus of attention using local color information. In *Proc. Computer Vision and Pattern Recognition Conf. (CVPR'93)*, pages 711–712, 1993.

[27] P.J. Burt, J.R. Bergen, R. Hingorani, R. Kolczynski, W.A. Lee, A. Leung, J. Lubin, and H. Shvaytser. Object tracking with a moving camera: An application of dynamic motion analysis. *Proc. IEEE Workshop on Visual Motion*, pages 2–12, March 1989. Irvine, CA.

[28] M. Irani, B. Rousso, and S. Peleg. Recovery of ego-motion using image stabilization. *Proc. IEEE Workshop on Visual Motion*, pages 454–460, 1994.

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

[30] D. Coombs and C. Brown. Real-time binocular smooth pursuit. *Int. J. Computer Vision*, 11(2):147–164, 1993.