Motion and Color Analysis for Animat Perception

Tamer F. Rabie and **Demetri Terzopoulos**

Department of Computer Science, University of Toronto
10 King's College Road, Toronto, Ontario, M5S 3G4, Canada
e-mail: {tamer|dt}@cs.toronto.edu

Abstract

We propose novel gaze control algorithms for active perception in mobile autonomous agents with directable, foveated vision sensors. Our agents are realistic artificial animals, or animats, situated in physics-based virtual worlds. Their active perception systems continuously analyze photorealistic retinal image streams to glean information useful for controlling the animat's eyes and body. The vision system computes optical flow and segments moving targets in the low-resolution visual periphery. It then matches segmented targets against mental models of colored objects of interest. The eyes saccade to increase acuity by foveating objects. The resulting sensorimotor control loop supports complex behaviors, such as predation.

Introduction

Animals are active observers of their environment (Gibson 1979). This fact has inspired a trend in the computer vision field popularly known as "active vision" (Bajcsy 1988; Ballard 1991; Swain & Stricker 1993). Unfortunately, efforts to create active vision systems for physical robots have been hampered by hardware and processor limitations. The recently proposed *animat vision* paradigm (Terzopoulos & Rabie 1995) offers an approach to developing biomimetic active vision systems that does not rely on robot hardware. Instead of physical robots, animat vision prescribes the use of virtual robots that take the form of artificial animals, or animats, situated in physics-based virtual worlds. Animats are autonomous virtual agents possessing mobile, muscle-actuated bodies and brains with motor, perception, behavior and learning centers. In the perception center of the animat's brain, computer vision algorithms continually analyze the incoming perceptual information. Based on this analysis, the behavior center dispatches motor commands to the animat's body, thus forming a complete sensorimotor control system. Animat vision, implemented entirely in software, has several important advantages over conventional "hardware vision", at least for research purposes (refer to (Terzopoulos & Rabie 1995; Terzopoulos 1995) for a discussion).

In many biological eyes, the high-acuity fovea covers only a small fraction of a visual field whose resolution decreases monotonically towards the periphery. Spatially nonuniform retinal imaging provides opportunities for increased computational efficiency through economization of photoreceptors and focus of attention, but it forces the visual system to solve problems that do not generally arise with a uniform field of view. A key problem is determining where to redirect the fovea when a target of interest appears in the periphery. In this paper we present a solution to this problem through the exploitation of motion and color information.

Figure 1: Artificial fishes swimming among aquatic plants in a physics-based virtual marine environment.

Motion and color play an important role in animal perception. Birds and insects exploit optical flow for obstacle avoidance and to control their ego-motion (Gibson 1979). Some species of fish are able to recognize the color signatures of other fish and use this information in certain piscene behaviors (Adler 1975). The human visual system is highly sensitive to motion and color. We tend to focus our attention on moving colorful objects. Motionless objects whose colors blend in to the background are not as easily detectable, and several camouflage strategies in the animal kingdom rely on this fact (Cedras & Shah 1995).

Following the animat vision paradigm, the motion and color based gaze control algorithms that we propose in this paper are implemented and evaluated within artificial fishes in a virtual marine world (Fig. 1). The fish animats are the result of research in the domain of artificial life (see (Terzopoulos, Tu, & Grzeszczuk 1994) for the details). In the present work, the fish animat serves as an autonomous mobile robot situated in a photorealistic, dynamic environment.

Our new gaze control algorithms significantly enhance the prototype animat vision system that we implemented in prior work (Terzopoulos & Rabie 1995) and they support more robust vision-guided navigation abilities in the artificial fish. We review the animat vision system in the next section before presenting our new work on integrating motion and color analysis for animat perception in subsequent sections.

A Prototype Animat Vision System

The basic functionality of the animat vision system, which is described in detail in (Terzopoulos & Rabie 1995), 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 progressively lower 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 to stabilize the visual fields during locomotion, to attend to interesting targets based on color, and to keep moving targets fixated. The artificial fish is thus able to approach and track other artificial fishes visually. Fig. 2 provides a block diagram of the active vision system showing two main modules that control retinal image stabilization and foveation of the eyes.

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. 3(a) shows an example of the 64×64 images that are rendered by the coaxial cameras in each eye (rendering employs the GL library and graphics pipeline on Silicon Graphics workstations). 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. 3(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. 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 brain of the artificial 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.

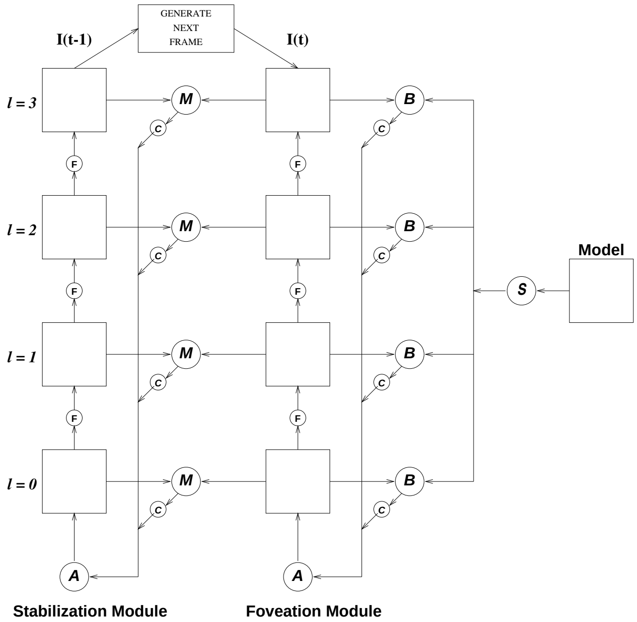


Figure 2: The animat vision system. The flow of the gaze control algorithm is from right to left. A: Update gaze angles (θ, ϕ) and saccade using these angles, B: Search current level for model target and if found localize it, else search lower level, C: Select level to be processed (see text), F: Reduce 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.

The mental models are stored as a list of 64×64 RGB color images.

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 (Swain & Ballard 1991).[1] Since each model object has a unique color histogram signature, it can be detected in the retinal image by histogram intersection and 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. 2 performs the saccades by incrementing

[1]Our improvements, which include iterative model histogram scaling and weighted histograms, make the technique much more robust against the large variations in scale that occur in our application. The details of the improved algorithm are presented in (Terzopoulos & Rabie 1995).

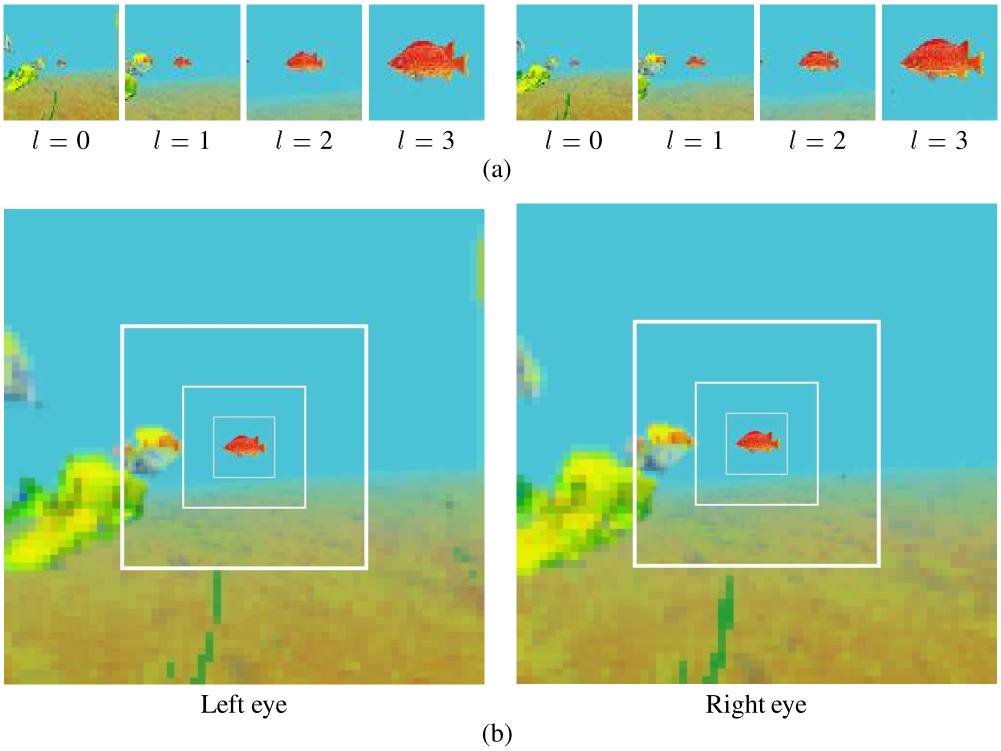


Figure 3: Binocular retinal imaging (monochrome versions of original color images). (a) 4 component images; $l = 0, 1, 2$, are peripheral images; $l = 3$ is foveal image. (b) Composited retinal images (borders of composited component images are shown in white).

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. 2) assumes the task of keeping the target foveated during locomotion.

Stabilization is achieved by computing the overall translational displacement (u, v) of intensities 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 $t-1$ (Horn 1986).

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. 4, 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

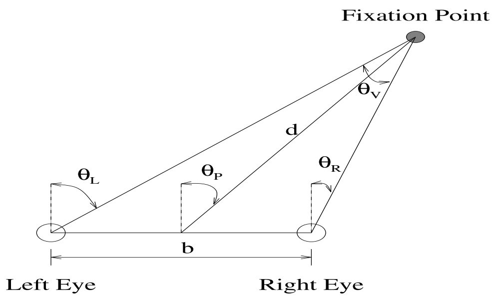


Figure 4: Gaze angles and range to target geometry.

straightforward for the vision system to estimate the range to the target from the gaze angles.

Vision-Guided Navigation

The artificial fish can also employ the gaze direction (i.e., the gaze angles) while 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. 4, 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—right-turn-MC for an above-threshold positive θ_P and left-turn-MC for negative θ_P—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 motor controllers are explained in (Terzopoulos, Tu, & Grzeszczuk 1994).

The remainder of the paper presents our new work on integrating color and motion analysis in active vision.

Integrating Motion and Color for Attention

Selective attention is an important mechanism for dealing with the combinatorial aspects of search in vision (Tsotsos *et al.* 1995). Deciding where to redirect the fovea can involve a complex search process (Tsotsos *et al.* 1995; Rimey & Brown 1992; Maver & Bajcsy 1990). In this section we offer an efficient solution which integrates motion and color to increase the robustness of our animat's perceptual functions.

Motion and color have been considered extensively in the literature in a variety of passive vision systems, but rarely have they been integrated for use in dynamic perception systems. The conjunction of color and motion cues has recently been exploited to produce more exact segmentations and for the extraction of object contours from natural scenes (Dubuisson & Jain 1993). Color and motion features of video images have been used for color video image classification and understanding (Gong & Sakauchi 1992).

Integrating motion and color for object recognition can improve the robustness of moving colored object recognition. Motion may be considered a bottom-up *alerting* cue, while color can be used as a top-down cue for model-based recognition (Swain, Kahn, & Ballard 1992). Therefore, integrating motion and color can increase the robustness of the recognition problem by bridging the gap between bottom-up and top-down processes, thus, improving the selective attention of dynamic perceptual systems such as the animat vision system that we are developing.

Where to Look Next

Redirecting gaze when a target of interest appears in the periphery can be a complex problem. One solution would be to section the peripheral image into smaller patches or focal probes (Burt *et al.* 1989) and search of all the probes. The strategy will work well for sufficiently small images, but for dynamic vision systems that must process natural or photorealistic images the approach is not effective.

We choose a simple method based on motion cues to help narrow down the search for a suitable gaze direction (Campani, Giachetti, & Torre 1995). We create a saliency image by initially computing a reduced optical flow field between two stabilized peripheral image frames (an advantage of the multiresolution retina is the small 64×64 peripheral image). Then an affine motion model is fitted to the optical flow using a robust regression method that will be described momentarily. The affine motion parameters are fitted to the dominant background motion. A saliency map is determined by computing an error measure between the affine motion parameters and the estimated optical flow as follows:

$$S(x,y) = \sqrt{[v_x(x,y) - u(x,y)]^2 + [v_y(x,y) - v(x,y)]^2}, \tag{1}$$

where (u, v) is the computed optical flow and

$$\begin{aligned} v_x(x,y) &= a + bx + cy, \\ v_y(x,y) &= d + ex + fy \end{aligned} \tag{2}$$

is the affine flow at retinal image position (x, y). The saliency image S is then convolved with a circular disk of area equal to the expected area of the model object of interest as it appears in the peripheral image.[2]

The blurring of the saliency image emphasizes the model object in the image. The maximum in S is taken as the location of the image probe. The image patches that serve as probes in consecutive peripheral frames form the image sequence that is processed by the motion segmentation module described later. Fig. 5 shows four consecutive peripheral images with the image probes outlined by white boxes. The blurred saliency image is shown at the end of the sequence in Fig. 5. Clearly the maximum (brightness) corresponds to the fast moving blue fish in the lower right portion of the peripheral image.

Robust Optical Flow

A key component of the selective attention algorithm is the use of optical flow. Given a sequence of time-varying images, points on the retina appear to move because of the relative motion between the eye and objects in the scene (Gibson 1979). The vector field of this apparent motion is usually called optical flow (Horn 1986). Optical flow can give important information about the spatial arrangement of objects viewed and the rate of change of this arrangement.

For our specific application, however, we require efficiency, robustness to outliers, and an optical flow estimate at all times. Recent work by Black and Anandan (Black & Anandan 1990; 1993) satisfies our requirements. They propose incremental minimization approaches using robust statistics for the estimation of optical flow which are geared towards dynamic environments. As is noted by Black, the goal is incrementally to integrate motion information from new images with previous optical flow estimates to obtain more accurate information about the motion in the scene over time. A detailed description of this method can be found in (Black 1992). Here we describe our adaptation of the algorithm to the animat vision system.

Ideally optical flow is computed continuously[3] as the animat navigates in its world, but to reduce computational cost and to allow for new scene features to appear when no interesting objects have attracted the attention of the animat, we choose to update the current estimate of the optical flow every four frames. The algorithm is however still "continuous" because it computes the current estimate of the optical flow at time t using image frames at t-3, t-2, t-1, and t in a short-time batch process. Fig. 6 shows this more

[2] Reasonably small areas suffice, since objects in the 64×64 peripheral image are typically small at peripheral resolution. Methods for estimating appropriate areas for the object, such as Jagersand's information theoretic approach (Jagersand 1995), may be applicable.

[3] By continuously, we mean that there is an estimate of the optical flow at every time instant.

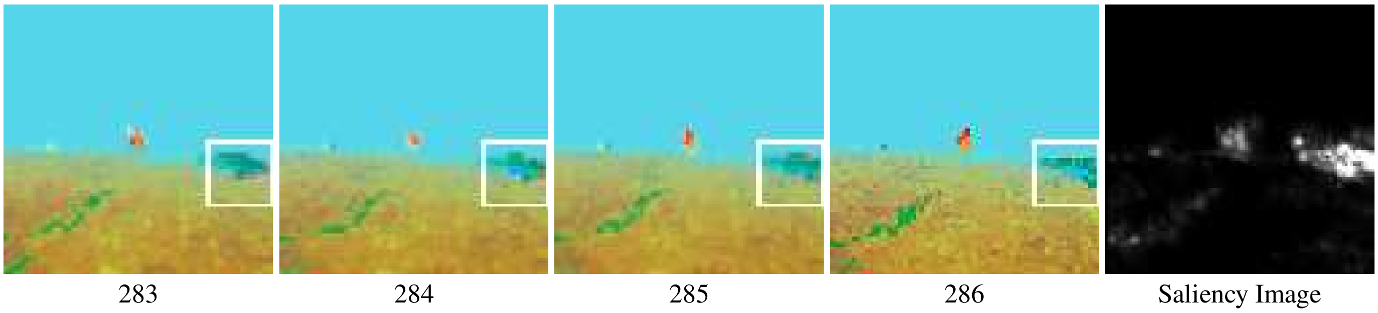


Figure 5: Four consecutive peripheral images with image probes outlined by white squares. Saliency image (right), with bright areas indicating large motions.

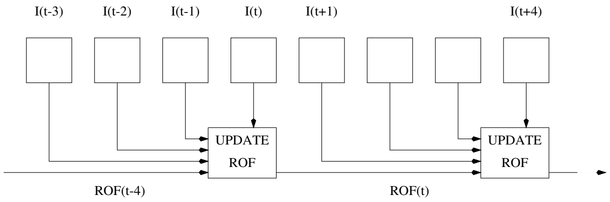


Figure 6: Incremental estimation of robust optical flow over time.

clearly. This arrangement requires storage of the previous three frames for use by the estimation module.

The flow at $t + 1$ is initialized with a predicted flow computed by forward warp of the flow estimate at t by itself[4] and then the optical flow at $t + 4$ is estimated by spatiotemporal regression over the four frames.

We compute our optical flow estimate by incrementally minimizing the cost function

$$E(u, v) = \lambda_D E_D(u, v) + \lambda_S E_S(u, v) + \lambda_T E_T(u, v), \quad (3)$$

where E_D is the data conservation constraint and is given in terms of the intensity constraint equation as

$$E_D = \rho_{\sigma_D}(uI_x + vI_y + I_t), \quad (4)$$

and E_S is the spatial coherence constraint and is given as

$$E_S = \sum_{m,n \in N} [\rho_{\sigma_S}(u - u(m,n)) + \rho_{\sigma_S}(v - v(m,n))], \quad (5)$$

where N is the 4-connected neighbors of the current pixel position. We formulate our temporal continuity constraint E_T by imposing some coherence between the current flow estimate and its previous and next estimate:

$$E_T = \rho_{\sigma_T}(\mathbf{u} - \mathbf{u}_{BW}) + \rho_{\sigma_T}(\mathbf{u} - \mathbf{u}_{FW}), \quad (6)$$

where $\mathbf{u} = (u, v)$ is the current optical flow estimate at time t, $\mathbf{u}_{BW}$ is the previous estimate at $t - 1$ obtained by setting it to the most recent estimate, and $\mathbf{u}_{FW}$ is a prediction of what the optical flow will be at $t + 1$ and is computed by forward warp of the current estimate by itself.[5] The λ parameters in (3) control the relative importance of the terms, and the ρ_σ functions in the above equations are taken to be the Lorentzian robust estimator:

$$\rho_\sigma(x) = log\left(1 + \frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right), \quad (7)$$

and its influence function, $\psi_\sigma(x)$, is the first derivative with respect to x. This function characterizes the bias that a particular measurement has on the solution (Hampel 1974; Black & Anandan 1993).

This robust formulation of our cost function E causes it to be non-convex. A local minimum can, however, be obtained using a gradient-based optimization technique. We choose the successive over relaxation minimization technique. The iterative equations for minimizing E are

$$u^{i+1} = u^i - \frac{\mu}{T_u}\frac{\partial E}{\partial u}, \quad (8)$$

where $1 < \mu < 2$ is an overrelaxation parameter that controls convergence. A similar iterative equation for v is obtained by replacing u with v in (8). The terms T_u, T_v are upper bounds on the second partial derivatives of E, and can be given as

$$T_u = \frac{\lambda_D I_x^2}{\sigma_D^2} + \frac{4\lambda_S}{\sigma_S^2} + \frac{2\lambda_T}{\sigma_T^2}, \quad (9)$$

and similarly for T_v by replacing u with v and x with y. The partial derivative in (8) is

$$\begin{aligned}\frac{\partial E}{\partial u} = {} & \lambda_D I_x \psi_{\sigma_D}(uI_x + vI_y + I_t) + \\ & \lambda_S \sum_{m,n \in N} \psi_{\sigma_S}(u - u(m,n)) + \\ & \lambda_T[\psi_{\sigma_T}(u - u_{BW}) + \psi_{\sigma_T}(u - u_{FW})], \quad (10)\end{aligned}$$

and similarly for $\partial E / \partial v$.

The above minimization will generally converge to a local minimum. A global minimum may be found by constructing an initially convex approximation to the cost function

[4] The flow estimate is being used to warp itself, thus predicting what the motion will be in the future.

[5] Note that $\mathbf{u}_{BW}$ can also be estimated by backward warping of $\mathbf{u}$ by itself.

Figure 7: The robust optical flow vectors estimated for the four image probe sequence (Fig. 5). Large vectors indicate large motion of the fish object.

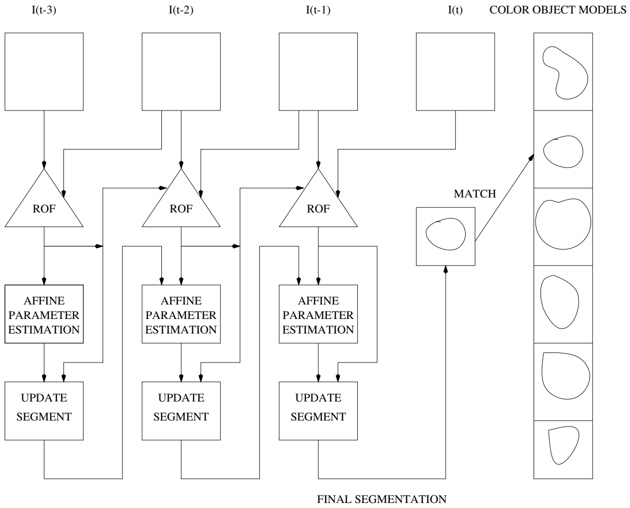


Figure 8: Incremental motion segmentation and object recognition using multi-resolution robust optical flow (ROF) estimation, affine parametric motion segmentation and color object recognition.

by choosing initial values of the σ parameters to be sufficiently large (equal to the maximum expected outlier in the argument of $\rho_i(\cdot)$), so that the Hessian matrix of E is positive definite at all points in the image. The minimum is then tracked using the graduated non-convexity (GNC) continuation method (Blake & Zisserman 1987) by decreasing the values of the σ parameters from one iteration to the next, which serves to gradually return the cost function to its non-convex shape, thereby introducing discontinuities in the data, spatial, and temporal terms. These discontinuities are, however, dealt with by the robust formulation and are rejected as outliers, thus producing more accurate optical flow estimates. The values of the λ parameters are determined empirically (typically $\lambda_D = 10, \lambda_S = \lambda_T = 1$).

To deal with large motions in the image sequence, we perform the minimization using a coarse-to-fine flow-through strategy. A Gaussian pyramid (Burt & Adelson 1983) is constructed for each image in the sequence, and minimization starts at the coarsest level and flows through to the finest resolution level. Our flow-through technique is based on the assumption that displacements which are less than 1 pixel are estimated accurately at each individual level and thus need not be updated from a coarser level's estimate, while estimates that are greater than 1 pixel are most probably more accurately computed at the coarser level, and are updated by projecting the estimate from the coarser level.

This incremental minimization approach foregoes a large number of relaxation iterations over a 2 frame sequence in favor of a small number of relaxation iterations over a longer sequence. Fig. 7 shows the optical flow estimated for the sequence of four image probes of Fig. 5. The figure clearly shows the complex motion of the target fish. It is a non-trivial task to segment such motions.

Motion Segmentation and Color Recognition

For the animat to recognize objects moving in its periphery it must first detect their presence by means of a saliency map as described earlier. Once it detects something that might be worth looking at, it must then segment its region of support out from the whole peripheral image and then match this segmentation with mental models of important objects. Fig. 8 shows the steps involved in an incremental segmentation of the detected object over the duration of the four probe images as explained above.

Segmentation of the optical flow at each time instant is performed by fitting an affine parametric motion model to the robust optical flow (ROF) estimated so far at the current time instant. This is done by incrementally minimizing the cost function given as

$$E(a,b,c,d,e,f) = E_x(a,b,c) + E_y(d,e,f), \tag{11}$$

where (a,b,c,d,e,f) are the affine motion parameters. E_x and E_y are formulated using robust estimation to account for outliers

$$\begin{aligned} E_x &= \sum_{x,y \in R} \rho_\sigma(v_x(x,y) - u(x,y)), \\ E_y &= \sum_{x,y \in R} \rho_\sigma(v_y(x,y) - v(x,y)), \end{aligned} \tag{12}$$

where R is the current region of support of the segmented object (initially equal to the full frame image size). v_x and v_y are horizontal and vertical affine motion flow vectors according to (2). (u,v) is the ROF estimated at the current instant, and $\rho_\sigma(x)$ is taken to be the Lorentzian robust estimator. We use successive over relaxation and GNC to minimize this cost function by using a small number of iterations over a sequence of four image probes and updating the segmentation at every time instant.

The estimated affine motion parameters at the current time instant are then used to update the segmentation by calculating an error norm between the affine flow estimate (v_x, v_y) and the ROF estimate as in (1). This norm is then thresholded by an appropriate threshold taken to be the minimum outlier in the affine fit. The updated segmentation serves as the region of support R for the next frame's affine minimization step.

If more than one moving object is present in the probe sequence, the current segmentation is subtracted from the image, and another affine motion model is fitted to the remaining pixels thus segmenting other moving objects. To clean up the segmentation (in case some pixels where misclassified as outliers) a 9×9 median filter is passed over the segmentation mask to fill in missing pixels and remove misclassified outliers. Fig. 9 shows the segmented background (showing two objects as outliers) and the segmentation of the outlier pixels into the object of interest (a blue fish).

At the end of the motion segmentation stage, the segmented objects are matched to color models using the color histogram intersection method. If a match occurs, the current estimate of the ROF is set to zero thus accounting for the dynamic changes in the system, otherwise the ROF is used to initialize the optical flow at the next time step as shown in Fig. 6.

If the model object matches the peripheral segmented region, the animat localizes the recognized object using color histogram backprojection and foveates it to obtain a high-resolution view. It then engages in appropriate behavioral responses.

Behavioral Response to a Recognized Target

The behavioral center of the brain of the artificial animal assumes control after an object is recognized and fixated. If the object is classified as food the behavioral response would be to pursue the target in the fovea with maximum speed until the animat is close enough to open its mouth and eat the food. If the object is classified as a predator and the animat is a prey fish, then the behavioral response would be to turn in a direction opposite to that of the predator and swim with maximum speed. Alternatively, an object in the scene may serve as a visual frame of reference. When the animat recognizes a reference object (which may be another fish) in its visual periphery, it will fixate on it and track it in smooth pursuit at an intermediate speed. Thus, the fixation point acts as the origin of an object-centered reference frame allowing the animat to stabilize its visual world and explore its surroundings.

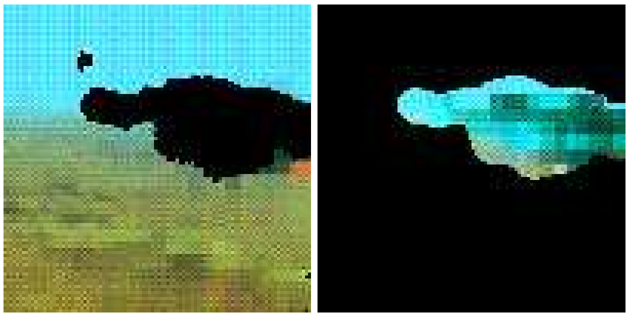
Fig. 10 shows a sequence of retinal images taken from the animat's left eye. The eyes are initially fixated on a red reference fish and thus the images are stabilized. In frame 283 to 286 a blue fish swims close by the animat's right side. The animat recognizes this as a reference fish and thus saccades the eyes to foveate the fish. It tracks the fish around, thereby exploring its environment. By foveating different reference objects, the animat can explore different parts of its world.

Fig. 11 shows a plot of the (θ_L, θ_R) gaze angles and turn angle between frames 200 and 400. It is clear from the figure that the animat was first fixated on the red fish which was to the left of the animat (negative gaze angles), and at frame 286 and subsequent frames the animat is foveated on the blue fish which is to its right (positive gaze angles).

Conclusion and Future Work

We have presented computer vision research carried out within an animat vision framework which employs a physics-based, virtual marine world inhabited by lifelike artificial fishes that emulate the appearance, motion, and behavior of real fishes in their natural habitats. We have successfully implemented a set of active vision algorithms for artificial fishes that integrate motion and color analysis to improve focus of attention and enable the animat to better understand and interact with its dynamic virtual environment.

In future work we will endeavor to increase the arsenal of active vision algorithms to support the whole behavioral repertoire of artificial fishes. The animat approach allows us to do this step by step without compromising the complete functionality of the artificial fish. It is our hope that the vision system that we are developing will also provide insights relevant to the design of active vision systems for physical robotics.



Segmented Background   Segmented Object

Figure 9: Results of incremental motion segmentation module.

Acknowledgements

We thank Xiaoyuan Tu for developing and implementing the artificial fish model, which made this research possible. The research described herein was supported by grants from the Natural Sciences and Engineering Research Council of Canada.

References

Adler, H. E. 1975. *Fish Behavior: Why Fishes do What They Do*. Neptune City, NJ: T.F.H Publications.

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

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

Black, M., and Anandan, P. 1990. A model for the detection of motion over time. In *Proc. Inter. Conf. Computer Vision (ICCV'90)*, 33–37.

Black, M., and Anandan, P. 1993. A framework for the robust estimation of optical flow. In *Proc. Inter. Conf. Computer Vision (ICCV'93)*, 231–236.

Black, M. 1992. Robust incremental optical flow. Technical Report YALEU/DCS/RR-923, Yale University, Dept. of Computer Science.

Blake, A., and Zisserman, A., eds. 1987. *Visual Reconstruction*. Cambridge, Massachusetts: The MIT Press.

Burt, P., and Adelson, E. 1983. The laplacian pyramid as a compact image code. *IEEE Trans. on Communications* 31(4):532–540.

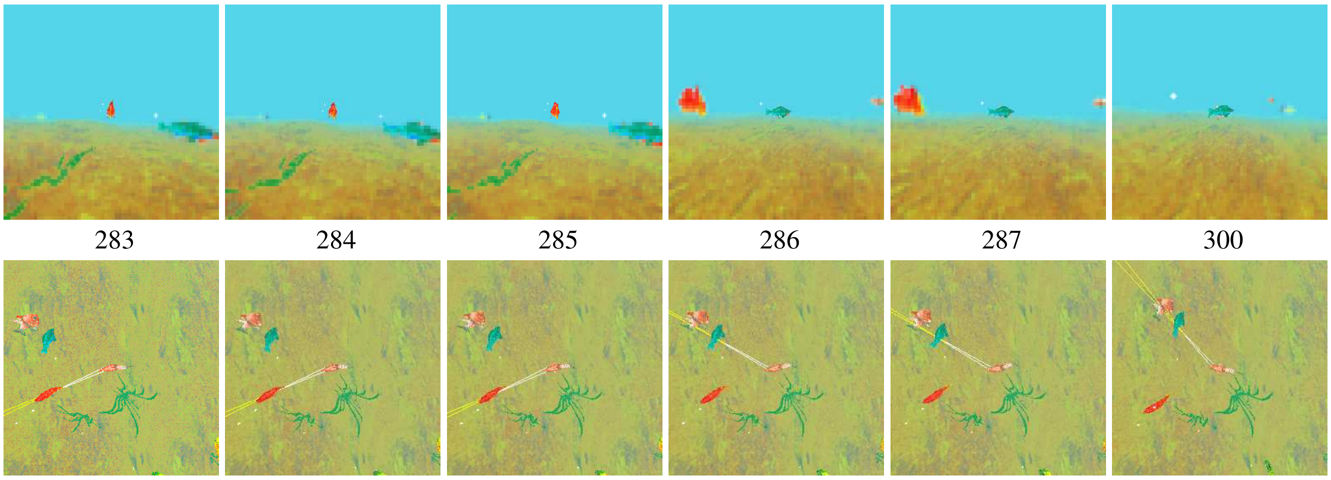


Figure 10: Retinal image sequence from the left eye of the predator (top) and overhead view (bottom) of the predator as it pursues a red reference fish (frames 283–285). A blue reference fish appears in the predator's right periphery and is recognized, fixated, and tracked (frames 286–300). The white lines indicate the gaze direction.

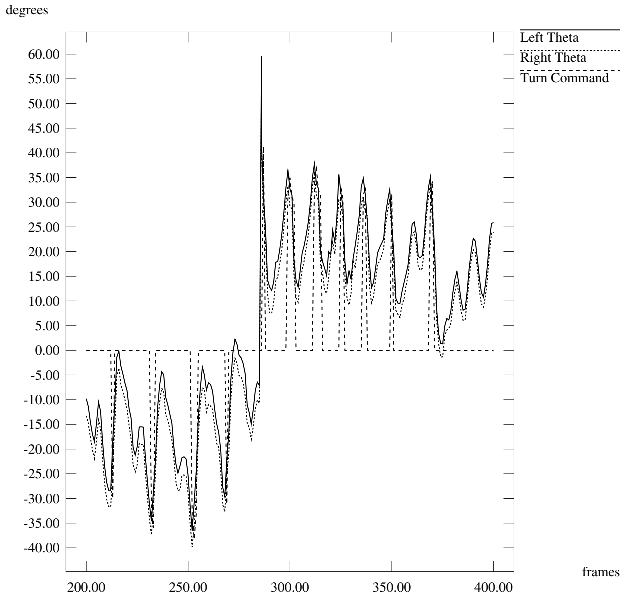


Figure 11: Gaze angles as the animat changes reference points at frame 286 from left (negative angles) to right (positive angles).

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

Campani, M.; Giachetti, A.; and Torre, V. 1995. Optic flow and autonomous navigation. *Perception* 24:253–267.

Cedras, C., and Shah, M. 1995. Motion-based recognition: A survey. *Image and Vision Computing* 13(2):129–155.

Dubuisson, M., and Jain, A. 1993. Object contour extraction using color and motion. In *Proc. Computer Vision and Pattern Recognition Conf. (CVPR'93)*, 471–476.

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

Gong, Y., and Sakauchi, M. 1992. An object-oriented method for color video image classification using the color and motion features of video images. In *2nd. Inter. Conf. on Automation, Robotics and Computer Vision*.

Hampel, F. 1974. The influence curve and its role in robust estimation. *J. Amer. Statistical Association* 69(346):383–393.

Horn, B. K. P. 1986. *Robot Vision*. Cambridge, MA: MIT Press.

Jagersand, M. 1995. Saliency maps and attention selection in scale and spatial coordinates: An information theoretic approach. In *Proc. Inter. Conf. Computer Vision (ICCV'95)*, 195–202.

Maver, J., and Bajcsy, R. 1990. How to decide from from the first view where to look next. In *Proc. DARPA Image Understanding Workshop*.

Rimey, R., and Brown, C. 1992. Where to look next using a bayes net: Incorporating geometric relations. In *Proc. Euro. Conf. Computer Vision (ECCV'92)*, 542–550.

Swain, M., and Ballard, D. 1991. Color indexing. *Inter. J. Computer Vision* 7:11 – 32.

Swain, M., and Stricker, M. 1993. Promising directions in active vision. *Inter. J. Computer Vision* 11(2):109 – 126.

Swain, M.; Kahn, R.; and Ballard, D. 1992. Low resolution cues for guiding saccadic eye movements. In *Proc. Inter. Conf. Computer Vision (ICCV'92)*, 737–740.

Terzopoulos, D., and Rabie, T. 1995. Animat vision: Active vision in artificial animals. In *Proc. Fifth Inter. Conf. Computer Vision (ICCV'95)*, 801 – 808.

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

Terzopoulos, D. 1995. Modeling living systems for computer vision. In *Proc. Int. Joint Conf. Artificial Intelligence (IJCAI'95)*, 1003–1013.

Tsotsos, J.; Culhane, S.; Wai, W.; Lai, Y.; Davis, N.; and Nuflo, F. 1995. Modeling visual attention via selective tuning. *Artificial Intelligence* 78:507–545.