Residual Q-Learning Applied to Visual Attention

Cesar Bandera
Amherst Systems, Inc.
Machine Vision Dept.
30 Wilson Road
Buffalo, New York
14221-7082
cba@amherst.com

**Francisco J. Vico,
Jose M. Bravo**
Facultad de Psicologia
Universidad de Malaga
29017 Malaga (Spain)
fjv@eva.psi.uma.es,
jbm@eva.psi.uma.es

Mance E. Harmon
Wright Laboratory
WL/AAAT Bldg. 635
2185 Avionics Circle
Wright-Patterson AFB,
Ohio 45433-7301
harmonme@aa.wpafb.af.mil

Leemon C. Baird III
U.S.A.F. Academy
2354 Fairchild Dr.
Suite 6K41
USAFA, Colorado
80840-6234
baird@cs.usafa.af.mil

Abstract

Foveal vision features imagers with graded acuity coupled with context sensitive sensor gaze control, analogous to that prevalent throughout vertebrate vision. Foveal vision operates more efficiently than uniform acuity vision because resolution is treated as a dynamically allocatable resource, but requires a more refined visual attention mechanism. We demonstrate that reinforcement learning (RL) significantly improves the performance of foveal visual attention, and of the overall vision system, for the task of model based target recognition. A simulated foveal vision system is shown to classify targets with fewer fixations by learning strategies for the acquisition of visual information relevant to the task, and learning how to generalize these strategies in ambiguous and unexpected scenario conditions.

1 OVERVIEW OF FOVEAL VISION

In contrast to the uniform acuity of conventional machine vision, virtually all advanced biological vision systems sample the scene in a space-variant fashion. Retinal acuity varies by several orders of magnitude within the field of view (FOV). The region of the retina with notably high acuity, called the fovea, is typically a small percentage of the overall FOV (<3%), centered at the optical axis (Levine 1985). The wide FOV, supported by lower peripheral acuity, and the high acuity fovea impose a much smaller data set (frame size) and permit a much faster frame rate than supporting the entire FOV uniformly at high acuity. For example, the retinotopic regions of the human visual system would be 15,000 times larger if the retina supported maximum acuity throughout its FOV (Yeshurun 1989). Inherent with space-variant sampling is the context-sensitive articulation of the sensor's optical axis, whereby the fovea is aligned with relevant features in the scene (Yarbus 1967). These features can be targets (e.g., predators or prey), or classification features on the targets themselves. Space-variant sampling and intelligent gaze control together with multiresolutional image analysis are collectively called foveal vision.

Through variable acuity and gaze control, biological systems treat signal and computational bandwidth as a dynamically allocatable resource. Vision functions direct sensor gazing in a process called foveation. This process provides a feedback path whereby low-level functions operate in a context-sensitive fashion (Figure 1), i.e., the vision system attempts to neither oversample, which wastes system resources, nor undersample, which reduces
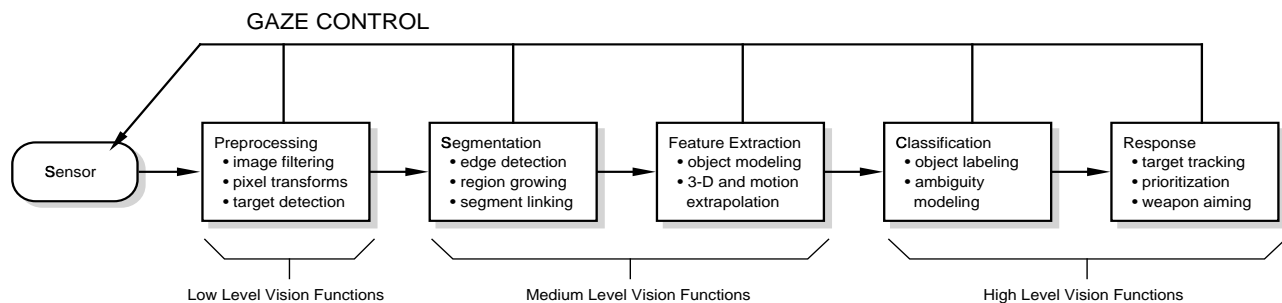

Figure 1: Data Path in Foveal Vision

system performance. The filtering of irrelevant data is performed at the earliest stage of the vision process, namely at the sensor itself, reducing the computational and data bandwidth over the entire vision data path.

Foveal vision is well-suited for applications where (1) scenario conditions, such as target range and kinematics, cannot be well controlled or anticipated, or (2) the system must simultaneously perform very different tasks, such as target tracking, object recognition, and navigation. It is often unfeasible to meet the FOV and spatiotemporal resolution requirements of all these tasks in all these conditions with a single uniform acuity vision system.

The premise of foveal vision is that the benefit from processing less (irrelevant) information per fixation is greater than the cost of making multiple refined fixations (i.e., *saccades*). Consequently, a fast intelligent robust visual attention mechanism that minimizes the number of fixations is necessary to acquire the aforementioned performance benefits of foveal vision.

The technical objective pursued by this research is the reduction in the overall number of saccades required to complete the recognition of a detected region of interest through the incorporation of reinforcement learning (RL) in the foveal vision attention mechanism. Target recognition is treated as a classification problem with a confidence threshold stop rule. This task requires an intelligent visual attention mechanism that can select fixation points whose interrogation yields as much visual information as possible relevant to class discrimination. In other words, the attention mechanism must be able to accurately predict the relevance of visual information expected from an interrogation, and use these predictions

to select a minimum length sequence of gazes whose acquired visual information, when integrated, permits the classification of the detected region of interest at threshold confidence (or better).

The foveal retinotopology used in this work is illustrated in Figure 2. It consists of r concentric rings about a fovea. Each ring is d receptive fields wide, the fovea is of size $4d \times 4d$, and the size of the receptive fields in the i'th ring from the fovea is $2^i \times 2^i$ (in terms of fovea pixels). The size of a receptive field is proportional to the L_∞ distance from its ring to the lattice center. Localized acuity is thus inversely proportional to distance from lattice center, and the acuity gradient is inversely proportional to d. Table 1 gives the ratio of central to peripheral acuity, and the bandwidth compression factor f_c (the ratio between the number of pixels in a uniform acuity image to the number of receptive fields in a foveal image with the same FOV and maximum resolution) for different ring counts. The rectilinear arrangement of receptive fields, the linear roll-off in acuity, and the power of two steps in acuity support computationally efficient multiresolution image processing. This approach to multiacuity sampling and processing is called hierarchical foveal machine vision (HFMV) (Bandera and Scott 1989, 1992).

The problem of visual attention is amenable to machine learning solutions. Classical or analytical solutions are very difficult due to unpredictable variability in scenario conditions (e.g., object range and orientation) and the model database (e.g., the addition of new classes). RL is a practical machine learning solution because it can train on the same performance feedback used by the visual
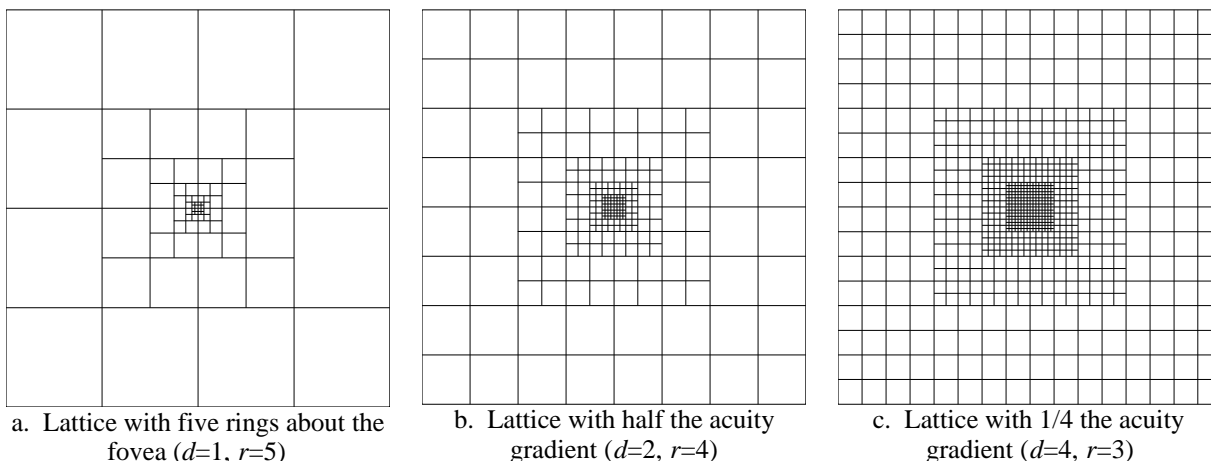

| a. Lattice with five rings about the fovea (d=1, r=5) | b. Lattice with half the acuity gradient (d=2, r=4) | c. Lattice with 1/4 the acuity gradient (d=4, r=3) |

Figure 2: Examples of HFMV Retinotopologies

Table 1: Bandwidth Compression and Central-to-Peripheral Acuity Ratio for Different Numbers of Acuity Steps

r	0	1	2	3	4	5	6	7	8	9
acuity	1:1	1:2	1:4	1:8	1:16	1:32	1:64	1:128	1:256	1:512
f_c	1	2.3	6.4	19.7	64	216	745	2621	9,362	33,825

process itself, namely the instantaneous recognition confidence. This simple scalar signal gauges the performance of the recognition process, serves as the process stop rule, and can serve as the reinforcement signal. It can be expressed in many different forms, such as a class likelihood ratio (to be maximized) or as the entropy of the class probabilities (to be minimized).

2 VISUAL ATTENTION AS AN OPTIMIZATION PROBLEM

Foveal object recognition can be posed as a task of sequentially interrogating the discriminant features of the object with high resolution. This is not to say that recognition is performed exclusively with the fovea; low acuity wide area measurements such as object aspect ratio and orientation can be performed efficiently in a single gaze with peripheral vision. Model-based foveal recognition must satisfy two important functional requirements:

1. A target model must be decomposed into salient features that are localized in scale-space such that they have a small spatial extent × spectral bandwidth product (e.g., detailed and small, like a corner or marking on the target, or large and coarse, like the target's aspect ratio). Visual features in naturally occurring fractal scenes typically meet these criteria.

2. Class likelihood must be expressed in terms of the probability of detection of target features. Different target features may be detected with different gazes, so a mechanism is required for the temporal integration of partial evidence.

The relevance of a fixation point to the task of classification is a function of the model database, which defines the different classes and hypothesizes the location of object features in the scene. Note that the model database features themselves have varying relevance to the classification task. An object model tends to be composed of salient features that describe that object. The features important to the task of classification are not those prominent in an object, but those that distinguish that object from the other objects in the model database. Relevance is also a function of accrued evidence, which may favor certain classes more heavily than others, and of the gaze history (no sense in revisiting a scene feature, unless in very noisy or kinematic conditions).

The multiresolution nature of foveal vision imposes a variable confidence in the detection of a feature that is a function of the localized acuity registering the feature (in addition to feature attributes). This accrued evidence can be represented as a list or topographic map of detected feature locations and the confidence associated with each feature detection.

The framework we employed for the foveal machine vision recognition is a post-binding model-based framework that assumes a user supplied model library that defines saliency in the different object classes, and a solution to the binding problem. In other words, the initial detection of an object is assumed to acquire sufficient information for the computation of a scale and pose for each model in the model database that best corresponds to the object detection. This post-detection-pre-recognition solution to the binding problem associates every feature of every model in the database with its maximum likelihood location in space, and permits visual attention to treat features from the model database as potential fixation points for interrogation.

The framework integrates the information from multiple saccades in order to classify a detection. After every saccade, the recognition process outputs an *a-posteriori* target class probability mass distribution function. This standard recognition process output permits using the entropy of integrated perception as a measure of visual information relevance. The entropy E is defined as

$$E = -\sum_{i=1}^{M} P(i) \log_2\big(P(i)\big) \qquad \text{(eq. 1)}$$

where M is the number of classes, and $P(i)$ is the probability that the detection belongs to the i'th class. By using log to the base two, entropy is expressed in bits.

Entropy is maximum when all probabilities are the same (perceptual ambiguity is maximum), and is minimum when one probability is one and all the rest are zero (perceptual certainty is maximum). Entropy lowers when information is acquired that helps in the classification of an object. The value of a saccade is defined as the decrease in entropy upon the processing of the foveal sensor frame. This value serves as a reinforcement signal to the RL system.

In this work, only spatially localized high bandwidth features such as corners will be considered as model features. Visual attention is less critical for the acquisition of lower acuity features, as these can be sufficiently registered by peripheral vision. For example, the model of each object class can be a semantic net of corners which is fit over a detected "blob" in the scene (Figure 3). As the blob is interrogated, a confidence measure is associated with each feature in the model database that indicates the presence or absence of that feature. From these feature confidence measures, the target probability vector P is derived. The collection of all confidence measures for all features is treated as the state of the recognition process. This state representation fulfills two key functions: it represents all acquired relevant visual information, and it can be used to compute the entropy of the temporally integrated perception.
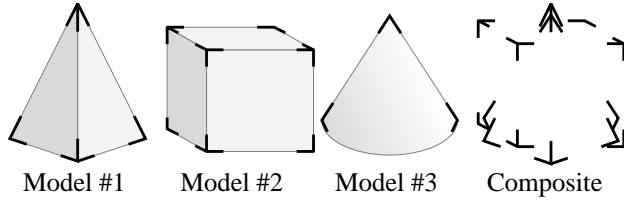
Figure 3: Decomposition of Objects into Vertices

Model #1   Model #2   Model #3   Composite

The visual attention mechanism selects a feature from the model database for interrogation given the state of the recognition process. This selection process is learned through the use of the residual form (Baird 1995, and Harmon, Baird, and Klopf 1995) of Q-learning (Watkins 1989), which we will refer to as *residual-Q*. The discounted cumulative entropy (i.e., the discounted cumulative reinforcement or utility) is defined as

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k} \qquad \text{(eq. 2)}$$

where r_t is the reduction in classification entropy after the transition from time t to time $t+1$, and $0 \le \gamma \le 1$ is the discount factor. Residual-Q attempts to learn a Q function that yields the optimal policy (i.e., a sequence of actions whose invoked reinforcement signals maximize R_t) when being greedy with respect to state-action pairs. The objective of the visual attention mechanism is to minimize its the discounted cumulative perception entropy.

The model database is implemented as a table describing the hypothesized location of each feature of each object in a 2-D space (binding normalized). Each of M models is described by N_i, i=1, ..., M, features, for a total of N features. Each feature has a distinct index (1, ... , N), and different models may have features in common. The discriminating power of a feature (i.e., the reduction in entropy caused by interrogating the feature) diminishes as more models incorporate the feature. None of the N_i features of a given model share the same location.

3 FOVEAL OBJECT RECOGNITION MODEL

3.1   FEATURE DETECTION

The foveal object recognition model is illustrated in Figure 4. The feature detection module accepts a scenario file that specifies the location in space of the visible features of the actual object, and the location in space of a fixation point. The output of the feature detection module is a vector V_f with N elements, each describing, with a scalar in the range of -1 to 1, the evidence detected in the current sensor frame corroborating the existence of each feature in the model database. The magnitude of the value indicates the level of confidence in the evidence of the feature, and the sign indicates whether the evidence is corroborative (positive) or not. A value of 1 in the i'th element of V_f indicates that the i'th feature of the model database has been unambiguously confirmed as present in the scene. A value of -1 indicates that feature has been unambiguously confirmed as absent in the scene. A value of 0 indicates no visual information in the current sensor frame substantiating or refuting the existence of the feature (e.g., the feature is outside the imager's FOV).
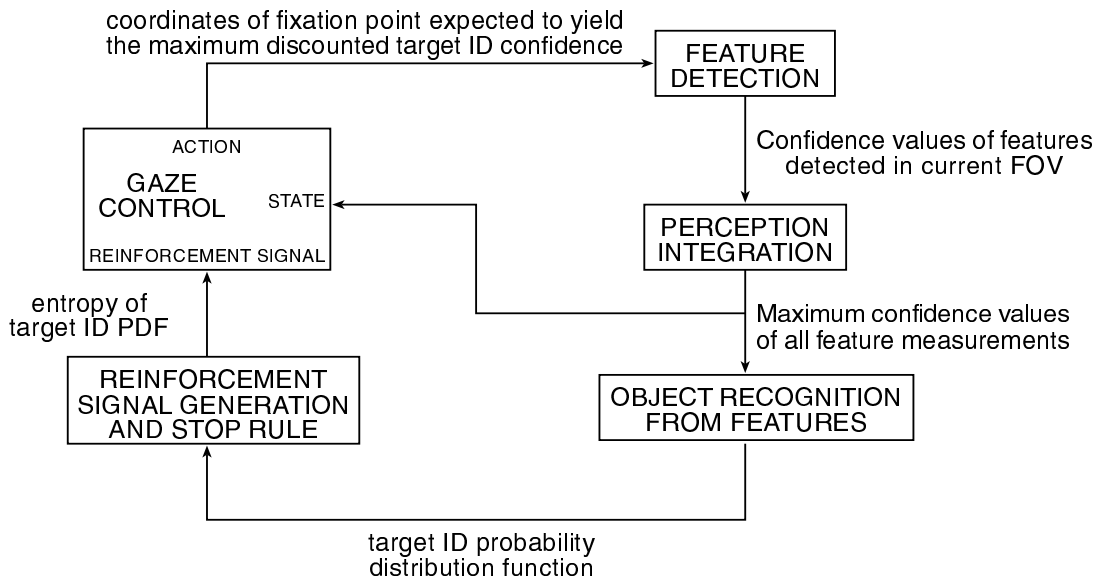

Figure 4: Block Diagram of Foveal Object Recognition Model

For the simulations presented in this paper, the ambiguity of a feature detection is computed as

$$V_f(i) = \begin{cases} s2^{-r_i} & r_i \leq r_{max} \\ 0 & r_i > r_{max} \end{cases} \qquad \text{(eq. 3)}$$

where r_i is the number of the ring (level of acuity in the FOV) covering the location of the i'th feature (the fovea is ring 0), $r_i > r_{max}$ represents the case where the feature location is outside the sensor's FOV, and s is 1 or -1 depending on the presence or absence of that feature in the scenario. The ring number is computed from the displacement $(\Delta_{i,x}, \Delta_{i,y})$ of the i'th feature position and the current sensor fixation point

$$r_i = \left\lfloor \log_2 \frac{\max(\Delta_{i,x}, \Delta_{i,y}, 1)}{d} \right\rfloor \qquad \text{(eq. 4)}$$

where $\lfloor x \rfloor$ is the truncation operator.

3.2 PERCEPTION INTEGRATION

The integrated perception builder integrates the information in V_f from the current frame with that from previous frames into vector V which represents the system's integrated perception and the state of the recognition process. The integrated perception simply retains the most confident evidence obtained on each feature:

$$V(i) = \begin{cases} V(i) & |V(i)| \geq |V_f(i)| \\ V_f(i) & |V(i)| < |V_f(i)| \end{cases} \qquad \text{(eq. 5)}$$

As with V_f, the elements of V range in value from -1 to 1, with magnitude representing feature measurement confidence and sign indicating presence or absence. The initial value of the integrated perception is $V(i)=0$, $i=1, ..., N$ (i.e., maximum system ambiguity).

3.3 OBJECT RECOGNITION FROM FEATURES

To generate the class likelihood vector (from which system entropy can be calculated) from the integrated perception vector, a backpropagation net (Rumelhart *et al.*, 1986) is trained on the potential states of the integrated perception. This approach consists of a net with N inputs, M outputs, and a hidden layer with $(M+N)/2$ nodes. A training data set is formed by compiling a list of possible integrated perception state vectors given some class as true, for all possible classes. The net is trained by driving it with these state values, and presenting it with the associated true class (+1 for the true class, -1 for the rest).

As an example, consider a simple two-class, three-feature scenario, with one class described by one feature and the other class described by the same feature plus another two features. The training data set is

Class	Input Pattern	Desired Output
1	[-1 -1 1]	[1 -1]
2	[1 1 1]	[-1 1]

The net response to different integrated perception states for this simple example is given in Table 2 below. The net output is treated as a class confidence vector C, from which heuristic class discrimination and entropy measures (after normalization into a probability distribution function) can be computed. The net generalizes for incomplete or ambiguous input patterns.

Table 2: Object Recognition Net Response to Perceptions

Ambiguous Perception (valid for both classes)	Class 1 Confidence	Class 2 Confidence
0 0 1	0.108919	0.130706

Perceptions from Class 1 Scenarios	Class 1 Confidence	Class 2 Confidence
-1 0 0	0.982249	-0.942880
0 -1 0	0.995989	-0.982058
-1 -1 0	0.998828	-0.995255
-1 0 1	0.989553	-0.969371
0 -1 1	0.997631	-0.990780
-1 -1 1	0.998990	-0.996125

Perceptions from Class 2 Scenarios	Class 1 Confidence	Class 2 Confidence
1 0 0	-0.996765	0.995901
0 1 0	-0.997766	0.997149
1 1 0	-0.993702	0.990032
1 0 1	-0.987956	0.981961
0 1 1	-0.998380	0.998181
1 1 1	-0.995156	0.993976
0 .25 0	-0.119307	0.518573
.25 0 0	-0.003283	0.418612
.75 0 .25	-0.808777	0.936406
0 .25 .25	0.059788	0.389670
.75 0 .25	-0.848022	0.947849
0 .75 0	-0.952402	0.988224

3.4 REINFORCEMENT AND STOP RULE

This module normalizes the class confidence vector C into a probability distribution function (PDF) vector P such that all the elements are in the range (0, 1) and sum to 1. From this PDF, the recognition entropy is computed (equation 1). The entropy is compared against a stop rule threshold. It is also subtracted from the previous entropy; the reduction in entropy is the reinforcement signal used to drive Q-learning.

3.5 GAZE CONTROL

A reinforcement learning system implementing the residual form of Q-learning serves as the gaze control mechanism. The Q function is implemented with a series of backpropagation networks, each computing the state-action pair utility for a particular action (Figure 5). The update equation for the residual-Q algorithm is given in equation 6, where ϕ is the weighting factor between the residual gradient and direct method update vectors.

$$\Delta w = -\alpha \left(R + \gamma \max_{u'} Q(x',u') - Q(x,u) \right)$$
$$\times \left(\phi\gamma \frac{\partial}{\partial w} \max_{u'} Q(x',u') - \frac{\partial}{\partial w} Q(x,u) \right)$$
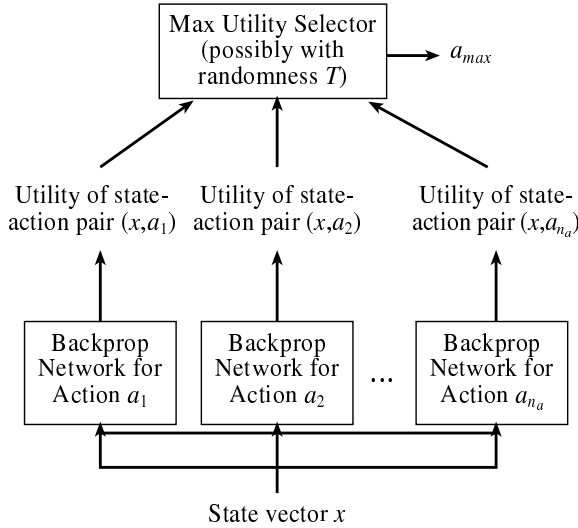
(eq. 6)

Figure 5: Parallel Q-Function Approximator

4 EXPERIMENTS AND RESULTS

The architecture described above was demonstrated in a simulation with a database consisting of ten classes and 20 distinctly situated features. Models were purposefully defined with considerable correlation to exercise the system's ability to gauge and react accordingly to differences in discriminating power among features (Table 3, where 1 indicates inclusion of a feature in the model). Each model has on average 10 features, and any one feature is common on average to five classes (the models were created with a random process that assigned any feature to any class with a probability of 50%). The 20 features are uniformly distributed in 2-D space such that for any fixation point the fovea can register no more than one feature (although lower acuity rings can detect more features with lower confidence).

The reduction in entropy ($E_{n-1}-E_n$) was used as the reinforcement signal. Other parameters follow: discount factor $\gamma=0.99$, learning constant $\alpha=0.9$, the residual coefficient $\phi=0.3$, entropy threshold (stop rule) $e_t=0.3$, and noise source $[-\beta T, +\beta T]$, $T=E_n-e_t$, $\beta=0.5$ added to the reinforcement signal to invoke exploration (Lin 1992).

Table 3 Model Database with 10 Classes and 20 Features

	1	2	3	4	5	6	7	8	9	10
1	1	1	0	1	0	0	0	1	0	1
2	1	1	0	0	0	0	1	1	1	1
3	1	0	1	0	1	1	1	0	0	1
4	0	0	1	1	0	0	1	1	1	1
5	0	1	1	0	0	1	0	0	0	0
6	0	0	1	1	0	1	1	1	0	1
7	0	1	0	1	0	1	0	0	1	1
8	0	1	0	1	1	1	0	0	1	1
9	1	1	1	1	1	0	0	0	1	1
10	1	0	1	0	0	1	0	0	0	0
11	1	1	0	1	0	1	1	1	1	1
12	1	0	0	1	0	0	0	0	0	0
13	1	1	0	0	0	1	1	0	1	0
14	0	0	0	1	1	1	1	1	0	1
15	0	1	1	0	0	0	1	0	0	0
16	1	1	1	1	1	0	0	0	0	0
17	1	0	0	1	0	0	1	0	0	0
18	0	0	0	1	0	0	0	1	1	1
19	0	1	1	0	1	0	1	1	1	1
20	0	1	1	0	1	1	1	1	1	1

Figure 6 illustrates how learning reduces the number of saccades required to classify targets of class 1, 3, and 6. Each trial consists of the recognition of a target of the given class, and all classes are equiprobable. The first two cases are representative of good performance improvement through learning, while the third case is representative of limited performance improvement.

The learned saccade sequences after 100 trials (each trial consisting of a complete sequence of saccades resulting in classification) are given in Table 4 for the 10 different classes. The table gives the indices of the model database features interrogated for each class. The system learns to first interrogate the 16th feature in the database. If that feature is detected, it proceeds to interrogate the 12th feature. Otherwise, if the 16th feature is confirmed absent, the system interrogates the 6th feature. The

system continues to implement the learned strategy in this fashion until the target is classified (entropy threshold is met). Some targets are classified in as little as four interrogations, while the 10th class requires 14 interrogations. This variability in saccade sequence is due more to the strategy than to the model, since statistically all the models are of the same complexity. Note how detecting the absence of a feature is just as significant as detecting its presence.
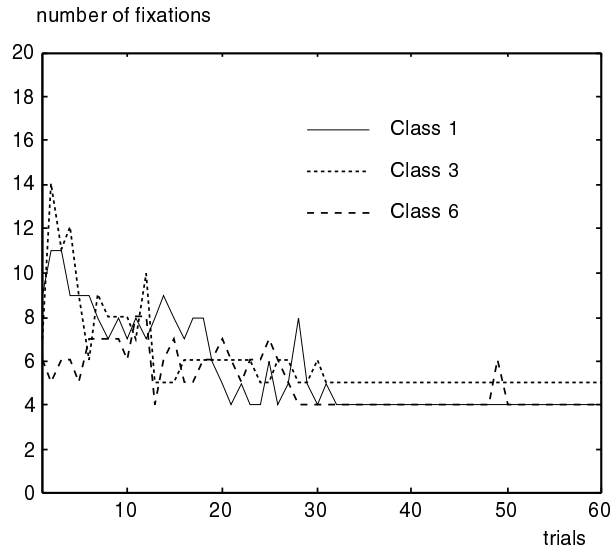
number of fixations

Figure 6: Fixations Required for Recognition

Table 4 Action Tree Learned Using Only the Fovea

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	16	12	6	3										
2	16	6	2	12	3	1	0							
3	16	6	3	12	2									
4	16	12	2	3	14	1	6							
5	16	6	3	12	2	1	14	0	5	19	8	9		
6	16	6	2	12										
7	16	12	6	3	14	2	1	0	5	19	8			
8	16	6	3	12	2	1	14							
9	16	6	2	12	3	14	1	0	19	5				
10	16	6	2	12	3	14	1	0	5	19	8	9	17	11

The above experiment was performed using only the fovea to impose the execution of many saccades and the optimization of these relatively long sequences. The experiment was repeated with the same RL parameters, but with a different foveal retinotopology. Three rings were added to implement graded acuity across the FOV. Neither the total number of receptive fields nor acuity along the optical axis were changed; fovea size was sacrificed for lower resolution, wider FOV perifoveal and peripheral vision.

The perifoveal and peripheral vision of this new retinotopology was able to detect additional features in any one glance, but with less acuity than the fovea. The RL based visual attention mechanism succeeded in using this partial, or ambiguous, information on the presence and absence of features in the scene to further reduce the number of saccades. The action tree formed in this experiment is presented in Table 5. Unlike the action tree of Table 4, which is traversed fixation at a time by the presence or absence of a single interrogated feature, the action tree of Table 5 is traversed fixation at a time by the complete and partial evidence of several features.

Randomness in the initial action in Table 5 (interrogating feature 18 versus feature 12) is from the temperature in the utility selector. The average number of interrogations was reduced from 8.1 for the uniform acuity retinotopology to 4.9 for the four acuity retinotopology.

Table 5 Action Tree Learned With Peripheral Vision

	1	2	3	4	5	6	7	8	9
1	18	12	6						
2	18	6	12	11	9	19	8		
3	18	12							
4	12	18	6	19					
5	18	12	6	19	11	9	8	0	14
6	12	6	8						
7	18	6	12	11	9				
8	18	12	6						
9	12	6	18	19	9	1			
10	18	12	6	19	9	11	8		

The uniform acuity experiment was repeated using a random gaze controller which selected features randomly and non-repeatedly during a trial. The average number of interrogations required to recognize a target (all targets equiprobable) was 13.2, which is 2.7 times greater than the number interrogations required by the foveal system.

Consistency in saccade sequences is an emergent behavior characteristic of human visual attention. This same behavior is exhibited by the RL simulations presented here. Specifically, the system seems to learn pieces of sequences, and performs the interrogations within these subsequences in a consistent order. However, the order in which the subsequences are connected to form the overall sequence is not necessarily consistent, and is (visual) data driven. Consistent behavior is more pronounced at the beginning of a trial than at the end, in part because entropy behaves as an exponentially decreasing function over time. The initial interrogations of a trial acquire information that is new to the system and which yields strong reinforcement. The final interrogations, particularly when the entropy threshold stop rule is low, receive little reinforcement to motivate the retention of sequence order.

5 FUTURE WORK

The experiments documented here assume equiprobable classes. One strong feature of RL is that it will learn different strategies for different *a-priori* class probabilities. We expect that once a strategy has been learned for a particular class distribution, RL will quickly adapt to changes in the scenario whereby some classes are encountered more frequently than others. This hypothesis is being tested.

The experiments performed to date also assume no penalty for gazing. However, in a real-time system with mechanically articulated cameras, any gazing action has an associated cost. This cost includes energy, and in the case of saccades, it also includes the duration of the saccade, during which the visual system is rendered ineffective. Cost should thus be a monotonically increasing function of saccade length (i.e., the amount of displacement of the optical axis). Experiments will be conducted with the utility of an action attenuated by the foveal displacement of the action. This technique is expected to further motivate consistent behavior by favoring contiguous short saccades over long saccades, and in effect reducing the strongly connected nature of action space.

The application of RL presented in this paper drives the active vision system into making confident classifications. This objective is separate from processing a reinforcement that gauges classification correctness (i.e., "right" or "wrong"). The probability of correct classification increases as the entropy threshold is lowered, but if the threshold is too low and the image quality is too poor, the stop rule may never be reached. The integration of classification correctness with classification confidence will also be investigated.

A long-term objective is to demonstrate the potential of RL in the context of a practical foveal machine vision prototype. This objective not only furthers the commercialization of RL, but also of hierarchical foveal machine vision, whose visual attention (and overall) performance is substantially improved through the use of RL. Future research will use a real-time platform operating in a practical nondeterministic scenario that typifies a commercial application of HFMV.

References

Baird, L. C. (1995). Residual Algorithms: Reinforcement Learning with Function Approximation. In Armand Prieditis & Stuart Russell, eds. Machine Learning: Proceedings of the Twelfth International Conference, 9-12 July, Morgan Kaufman Publishers, San Francisco, CA.

Bandera, C., Scott P. (1989). Foveal Machine Vision Systems. IEEE International Conference on Systems, Man, and Cybernetics, Cambridge, MA, November.

Bandera, C., Scott P. (1992). Multiacuity target recognition and tracking. Proceedings of the Second Automatic Target Recognizer Systems and Technology Conference, Fort Belvoir Center for Night Vision and Electro-Optics, March 17.

Harmon, M.E., Baird, L.C, & Klopf, A.H. (1995). Advantage updating applied to a differential game. In Tesauro, G., Touretzky, D.S., and Leen, T.K. (eds.), *Advances in Neural Information Processing Systems 7*. MIT Press, Cambridge MA.

Levine, M. D. (1985). *Vision in Man and Machine*, McGraw Hill.

Lin, L. J. (1992). Self-improving reactive agents based on reinforcement learning, planning and teaching. *Machine Learning*, no. 8, pp. 293-321, Kluer Academic.

Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning representations by back-propagating errors. *Nature*. **323**, 9 October, 533-536.

Watkins, C. J. C. H. (1989). Learning from delayed rewards. Doctoral thesis, Cambridge University, Cambridge, England.

Yarbus, A. L. (1967). *Eye Movements and Vision*, Plenum Press.

Yeshurun, E. L. Schwartz. (1989). Shape Description With a Space-Variant Sensor: Algorithms For Scan-path, and Convergence Over Multiple Scans. *IEEE Trans. PAMI*, vol. 11, no. 11, pp. 1217-1222, November.