LEARNING-BASED CONTROL OF PERCEPTION FOR MOBILITY

Matthew Barth, Subhodev Das, and Bir Bhanu
College of Engineering
University of California
Riverside, CA 92521

Abstract: Machine perception plays an important role in any intelligent system, and in particular, guiding an autonomous mobile agent. Machine perception techniques have progressed significantly in recent years, however perception systems are still plagued by a lack of flexibility and an inadequacy in performance speed for use in real-time tasks. To overcome these problems, we have applied *integrated learning techniques* to a perception system that is based on a *selective sensing paradigm*. The incorporation of multiple learning algorithms at different levels in our perception system provides a great deal of flexibility and robustness when performing different perceptual tasks. Making use of a selective sensing paradigm allows the system to eliminate a large amount of non-pertinent sensory data so that processing speed is greatly increased. We are implementing such a perception system to be used on an autonomous mobile agent. In this paper, we describe our methodology and give a preliminary example of learning within our perception system.

1. Introduction

Machine perception has been a fast growing research area throughout the past two decades. Much progress has been made in the low-level computer vision techniques, and these techniques have found their way into many different application areas. However, today's machine perception solutions (i.e., algorithms) tend to be very specific and rigid for any one particular application, such as target recognition or navigation. Machine vision solutions are also very sensitive to perceptual conditions, system environments, and specific application requirements. In order for machine perception to be useful, it will need to learn and adapt to various environments, changing conditions, and different tasks.

Another problem with today's machine perception algorithms is that they are often unable to perform tasks in real-time due to the large amount of time needed to process sensory information. In particular, perception systems for mobile robots must interpret a deluge of environmental information acquired by their numerous sensors, and are forced to operate very slowly due to lack of sensory processing power. In order to speed up the perception processing, researchers often 'tailor' their systems so that sensing and sensory processing operate rapidly enough for their particular application. In so doing, they eliminate the flexibility in their system and are not able to use their system for other types of applications.

In order to overcome these two problems of inflexibility and non-real-time performance, we have developed a sensing strategy that incorporates *machine learning* techniques for flexibility, and uses an intelligent *selective sensing paradigm* to increase operational speed. We are implementing this sensing strategy on a perception system to be used on an autonomous mobile agent (i.e., robot or vehicle). The goal of our autonomous mobile agent is to navigate successfully and interact within an unstructured and unknown environment. Using very little a priori knowledge, the agent must perceive the environment, perform scene understanding, move about while avoiding obstacles, and perform various tasks.

The perception system of an autonomous mobile agent is an ideal target for machine learning. Different learning techniques can be applied to the various tasks and subtasks the mobile agent must perform in order to navigate and interact in its environment. Depending on the task or subtask, we use learning techniques such as Explanation-Based Learning (EBL), learning-by-example, learning-by-doing, conceptual clustering, and genetic algorithm techniques.

The selective sensing paradigm that we use is based on the fact that in general, only a small percentage of information in the sensory data flow provided by various sensors is relevant to a particular task. This is especially true in computer vision, where a substantial fraction of image data can be ignored. It is possible to apply parallel processing hardware in a brute force fashion, however, the unstructured application of parallel processing is not very efficient and often leads to a waste of processing resources. By intelligently applying processing resources to only the pertinent data in the environment, sensory processing speed can be greatly increased.

We have developed our perception system to run on coarse parallel hardware. The sensing strategy that we use is similar to that of the human attention ability, where processing takes place only on the sensory information relevant to the task at hand. Indeed, if we attempt to perform detailed analysis and recognition algorithms on all of the sensory information, we would be quickly overwhelmed by the processing requirements. For this reason, a human places his attention only on what is pertinent in the environment.

The system also makes use of active vision hardware. Instead of analyzing passively sampled images, there is much to be gained by engaging in some kind of activity whose purpose is to control the geometrical parameters of the sensory device. Active vision consists of behaviors such as adjusting the lens aperture for the proper level of illumination, adjusting the focal length of the lens in order to bring images into sharp focus, converging and diverging a pair of binocular cameras, and moving the camera(s) independently in order to get a better view of an object. Another active behavior that we exploit is camera gaze control.

In this paper, we first provide a short background of other work in learning techniques applied to perception, and then briefly describe work in attentive sensing. We then describe the methodology of our perception system and the applied learning techniques. Finally we discuss in detail an example of the system's operation including some preliminary results, followed by a conclusion.

2. Background

2.1. Learning in Perception

Machine Learning (ML) is perhaps the most important capability of an intelligent system by the virtue of which such a system can acquire new information, adapt to changes in the operating environment, and improve its own performance over time [7]. Prior work in applying machine learning technology to machine perception and specifically the computer vision field has been limited to the training of statistical pattern classifiers. More recent computer vision systems that incorporate some type of learning component are relying on the learning-from-examples paradigm to obtain computer models for the objects of interest. The ANALOGY program created by Evans [13] which examined simple line drawings to acquire analogies among different drawings was the first system to incorporate a learning capability into a computer vision framework. The system by Connell and Brady [10] learns shape models from 2-D real images that can be used for recognizing subsequent instances of the learned concept. Perkins [22] utilizes machine learning to obtain models of object parts that may be used to inspect these parts at later instants. Segen [23] has used an automatic technique for learning descriptions from examples of real, complex, and nonrigid objects. Toriu et al [27] have used hierarchical clustering approach for automatic book discrimination. Kim et al [19] have used a multilevel classifier to recognize objects by viewing multiple instances of them without requiring explicit models or rules.

It is a positive step in the direction of a machine vision system where the visual memory is acquired directly from the environment. Lehrer and Reynolds [20] present a method for initial hypothesis formation using an automatically generated low-level knowledge base obtained from a set of training instances. Hutber and Sims [15] present a similar theoretical method for automatically generating rules in the knowledge base from a set of sample data. Bhanu et al [6] describe a genetic algorithm-based adaptive image segmentation technique. A multi-strategy learning technique that incorporates explanation-based learning and structured conceptual clustering is used for automatic object model acquisition and refinement in an aircraft recognition scenario [21].

There have been several applications of learning in the robot vision field. Shun-en and Calvert [25] use an approach to incrementally construct the 3-D models in an office or warehouse environment by matching planned multiple views. Tsai and Chen [29] apply machine learning to the task of adaptive navigation for automated vehicles using image analysis techniques. de Figueriredo and Wang [11] use an 'evolving frame' approach to learning with applications to adaptive navigation. Grefenstette and Pettey [14] make use of genetic algorithm for navigating a simulated robot. Aloimonos and Shulman [2] present a theoretical treatment of a neural network approach which can learn all the parameters that are involved in the 'shape from X' class of problems. Whitehead et al [31] uses 'markers' to temporarily record partial computational results in order to reduce the representational burden in learning. An approach to a class of machine learning that involves autonomous concept formation, using feedback from trial-and-error learning, in the context of an autonomous robot at a process control panel is described by Spelt et al [26].

2.2. Attentive Sensing

One of the most important characteristic of the active vision paradigm (introduced in [3]) is the ability for selective sensing or attention in space, resolution and time. Most past research involving this topic has been confined to the cognitive psychology and the neuroscience communities. Ullman's work [30] and the introduction of the active vision paradigm [3] have prompted a surge in activities in the computer vision community. The research in selective sensing or attention has progressed along two directions — explicit sequencing and implicit sequencing of attention. Explicit sequencing deals with time-ordered perception of features for the purpose of recognition. Thus, such a mode of attention is goal-driven. The earliest work in explicit sequencing is that of Yarbus [32]. Implicit sequencing evolves through a process of data interpretation. Hence, this mode of attention is data-driven. Implicit sequencing has received more attention in computer vision community than its explicit counterpart which traditionally has been of interest to the cognitive psychology community. The initial work in implicit representation of sequences is due to Didday and Arbib [12]. Criteria were identified for the selection of a fixation point which were motivated by known characteristics of fixation in human vision as well as computational considerations in [1]. Shmuel and Werman [24] have considered the related problem during surface map generation from multiple viewpoints; they use iterative Kalman-filtering techniques to predict a new camera pose for maximal reduction of uncertainty in depth information. Burt [9] has considered hierarchical approaches to the target selection process; pyramid-based implementation searches for information pertinent to a chosen task by processing images at multiple spatial resolutions. Barth [4] has developed an attentive sensing strategy for the inspection of integrated circuit wafers, using a multi-window vision architecture [5,16]. Some recent studies have considered higher-level criteria for fixation [8].

3. Methodology

3.1. System Description

The overall methodology of our system is illustrated in figure 1. We see that it is composed of several components, namely, Supervisory Control, Selective Perception, Dynamic Reasoning, Multi-Layered Action, and System Dynamic Knowledge. In general, environmental information is gathered through sensors which provide data to the selective perception component. The selective perception component interacts with the dynamic reasoner, which also deals with the multi-layered action component. The multi-layered action component performs any actuation in the environment, including the control of the active vision parameters.

In this paper we will only describe the issues associated with selective perception and its interaction with the dynamic reasoner. However, note that we have applied learning to all of the components in order to make the entire system flexible and robust.

3.2. Selective Perception

The selective perception component of our system is shown in figure 2. This component interacts primarily with the dynamic reasoner, both in control (indicated by dashed lines) and in information (indicated by solid lines). At the bottom of the figure, various sensors provide data along the sensory data flow path.

In order to perform selective perception, we divide the sensing tasks into two processing stages. The first stage of processing, called the *pre-attentive stage*, consists of **M** parallel processing units which rapidly and automatically extract salient data in the sensory data flow across the entire sensory field. These pre-attentive processing units are able to work rapidly since they are not concerned with the detailed quantitative measure of sensory data, but rather the qualitative question of whether there are items with unique features that deserve attention. They simply find items with salient features in the sensory data based on top-down and bottom-up focusing-of-attention methods and pass location information to the second stage of processing made up of **N** attentive processing units. Information gathered by the pre-attentive stage guides the processing of the *attentive stage*. The attentive processing units can then perform the detailed measurement or verification algorithms required by the task only on the pertinent sensory information. By conducting sensory processing in this manner, we are able to eliminate irrelevant sensory information and thus greatly speed up the sensory processing task.

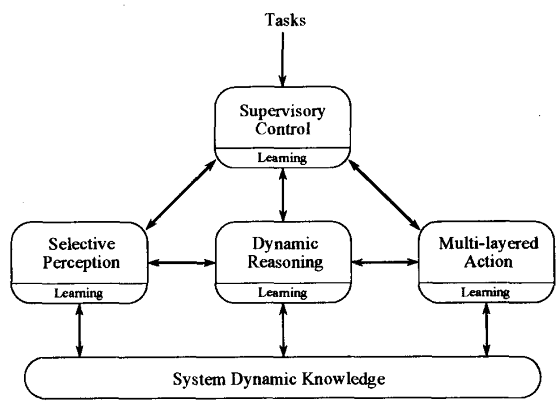

Figure 1: Overview of the functional components of the system.

The selective sensing strategy proceeds as follows. The pre-attentive processes are assigned to different sensory modalities (such as brightness, color, texture, motion, range, temperature, etc.) to extract, in parallel, salient information about the environment from the sensory data flow. Each pre-attentive process determines a degree of saliency for objects deserving attention based on its features within the sensory modality. Then, they place a minimum of information about the salient objects on a data structure called a 'saliency list' common to both the pre-attentive processes and the dynamic reasoner. Detailed data is not exchanged here, but rather just location information about each salient object and its intrinsic degree of saliency. The dynamic reasoner then can prioritize the saliency list and assign 'attentive' processes to each salient object for measurement or verification required by the perception task. The attentive processes are carried out in parallel and provide necessary information for the higher level tasks. The dynamic reasoner can also configure the active vision parameters through the multi-layered action component of the system. For example, if an attentive process required a different field of view or greater resolution, the active vision mechanisms would be set to comply.

The critical component of this selective sensing strategy is how salient items are selected in the pre-attentive stage. We can classify the determination of saliency in the sensing process into two general cases:

1) Goal-Driven Attention: Often a perception task is concerned with finding a particular object or set of objects in the environment with specific attributes. To detect these objects of interest, it is possible to set up pre-attentive sensory processes that specifically 'look' for a set of features that are characteristic of the objects. For example, regions can be identified via low-level segmentation algorithms. In the interest of simplicity and speed, only simple attributes are used in these pre-attentive processes, such as brightness, color, or size. When the object or region of interest does appear, it attracts attention by triggering the process that was set up to detect its attributes. Triggering can be accomplished by exceeding a threshold when matching measured attributes with known attributes. This method of 'target detection' in the pre-attentive stage is not limited to vision and can be extended to any sensing modality. 'Target recognition' then takes place after the object has been foveated in the attentive stage, using a larger set of more complicated features. When performing target detection in the pre-attentive stage, a degree of saliency can be determined based on the closeness of match with the known attributes. This degree of saliency is passed along with

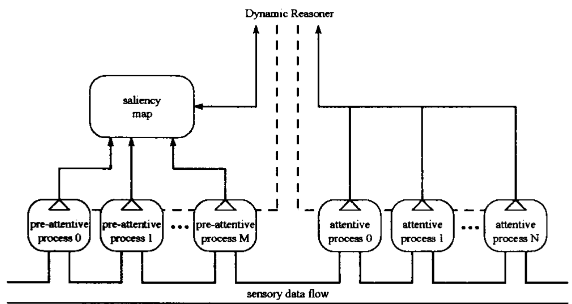

Figure 2: System selective perception component.

location information of the salient region. The dynamic reasoner uses this information when dispatching attentive processes. This pre-attentive process is a goal-driven attentive process, since we have prior knowledge of the attributes of the objects of interest. Further details on focusing-of-attention based on target detection can be found elsewhere [4].

Examples of 'target' items of interest for a mobile agent might be landmarks for route navigation. These landmarks may be items such as corners of a hallway in order to initiate a turn, or man-made landmarks that can be used for accurate position and orientation determination. Other targets a mobile robot may be seeking could be items of trash for a cleaning robot, or human beings for a service robot.

2) Data-Driven Attention: An object in the environment can also attract sensory attention if it 'different' from the other objects in the environment. How different an object is from the others is based on the object's general properties or features: e.g., if one or more features are distinct from the features of other objects, the object attracts attention. An object with different features is deemed peculiar. Most work in the psychological study of human attention has taken place in this area. It has been determined that there are many simple object features as well as spatial and temporal properties that can attract human attention if different from the corresponding features of other objects [28]. As an example, a red flower in a field of green grass quickly attracts attention based on the quality feature of color.

In this case, saliency is based only on the sensory data. No previous knowledge about the features or objects is required for an object to be salient, i.e., the saliency determination is data-driven. Data-driven saliency determination algorithms based on a concept of feature isolation have already been shown to be effective [4]. Data-driven saliency determination is useful when a task is to be carried out in an unknown domain, often the case for a mobile agent. For example, the detection of an object that has a motion field different from the surrounding environment should attract attention in order to identify the moving object. Further, obstacle detection and avoidance can be based on data-driven attention since obstacles unexpectedly appear along a previously smooth route. Lastly, landmarks can be autonomously detected during a training run of a mobile robot along a route using data-driven attention, and then subsequently detected later for route recognition when traversing the route again, using goal-driven attention [33].

Saliency location information flows up from the pre-attentive processes to the dynamic reasoner via a saliency list. The dynamic reasoner prioritizes this saliency list and then allocates and dispatches attentive processes to perform more detailed analysis of the sensory information. The results of the attentive processes are then passed up to the dynamic reasoner, forming the solutions to the perception tasks. In order to manage both the pre-attentive and attentive processes, there are control paths from the dynamic reasoner to the pre-attentive and attentive stages, indicated by dashed lines in figure 2. These control paths are an integral part of the learning processes described in the next section.

3.3. Learning in Selective Perception

In order to make the perception process more flexible and robust, we employ learning at three different levels in our selective perception paradigm.

1) Parameter Learning: The algorithms that run in both the pre-attentive and attentive processes have input parameters that are crucial to their overall effectiveness. Depending on any particular situation, a certain set of algorithm parameters will yield satisfactory results. However, our perception system has been designed so that it can operate robustly in a number of different situations, with different environmental conditions, and with different required tasks. Therefore, the algorithm parameters will need to vary according to these outside influences. We accomplish these parameter alterations using parameter learning algorithms.

Numerous parameter learning techniques exist, among them are gradient approaches, maximum likelihood estimation, clustering techniques, genetic algorithms, simulated annealing, rule-based systems, and surface response methodologies. We describe in detail a parameter learning technique based on clustering in the next section.

The parameter learning for both pre-attentive and attentive algorithms takes place in the dynamic reasoner. The dynamic reasoner is constantly tuning each pre-attentive and attentive process by evaluating the results of each process, and applying the best parameter set to each one. The evaluations are based on confirmation of results produced by other pre-attentive and attentive processes.

2) Utility Determination: Depending on the perceptual task at hand, the dynamic reasoner also determines a 'utility' for each pre-attentive and attentive process. Knowing the task, the dynamic reasoner loads appropriate routines as the pre-attentive and attentive processes. This is accomplished by learning through experience of which routines are more effective in certain situations. After the routines are loaded and are running, the effectiveness of each process is evaluated. The dynamic reasoner then assigns a weight to each process corresponding to the merit it has in helping to solve the perceptual task. In a sense, the dynamic reasoner 'attenuates' the saliency output of each pre-attentive routine through the dashed control lines seen in figure 2. Similarly, the dynamic reasoner provides a utility measure on the output of each attentive process.

As an example, suppose a perception task gains little information from a specific sensory modality, the pre-attentive processes which extract information about that sensory modality have their saliency output attenuated so that critical attentive processing time is not wasted. Along the same lines, if a particular pre-attentive routine has a high false alarm rate when passing saliency information to the saliency list, the dynamic reasoner would 'learn' that this pre-attentive routine was unreliable, and would therefore attenuate his saliency output. Thus in turn, it would be less likely that an attentive process would be assigned to evaluate the salient location hypothesized by the error-prone pre-attentive process.

In utility determination, the dynamic reasoner would base its learning on evidence accumulation. This is done through Bayesian formulation, or if we use both negative and positive evidence, through Dempster-Schafer formulation.

3) Cognitive Learning: At a higher level in our selective perception paradigm, we would need to accumulate knowledge that would provide the information required for our goal-driven focusing-of-attention used by some of the pre-attentive and attentive routines. Model information will need to be accumulated when determining landmarks for such things as route recognition. Numerous attentive processes would also make use of accumulated knowledge in order to perform recognition that was important to the perception tasks.

In this paper, we will not elaborate on cognitive learning, since we are primarily interested in attention control in the selective perception component of our system.

4. An Example: Parameter Learning for Obstacle Detection

In the scenario of an autonomous mobile agent with on-board active vision systems navigating in an unconstrained environment, we consider the task of detecting moving obstacles. Two of several of the pre-attentive processes which are significant for this task are those which can detect motion and find regions that may correspond to the obstacles. Once these two pre-attentive processes have produced supporting evidence, the appropriate attentive processes may be invoked to positively identify the obstacles and estimate their rate of closure to the vehicle platform.

Often a problem with both motion detection and region-finding algorithms is that their results are sensitive to varying environmental conditions, such as lighting. Therefore, we require that the algorithms *adapt* to the different conditions through a parameter learning technique based on clustering.

4.1. Algorithm Description

A simple motion detection algorithm that can run as one of the many pre-attentive processes works in the following way. It obtains differences in image intensities at the corresponding pixel locations in two consecutive frames I_k and I_{k+1} taken Δt time apart [17,18]. If the number of contiguous pixels where any change in intensity is observed is larger than some threshold τ_{m1} then the corresponding pixels are marked in the difference image I_d:

$$I_d(r,c) = 1, \text{ if } \sum_{\Delta r=-2}^{+2} \sum_{\Delta c=-2}^{+2} U\left(\left| I_k(r+\Delta r,c+\Delta c) - I_{k+1}(r+\Delta r,c+\Delta c)\right|\right) > \tau_{m1}$$

$= 0$, otherwise

where $U(.)$ is the unit step function. It is expected that the marked pixels form disjoint sets of connected regions in the difference image where each set of connected regions correspond to a object. The criterion that the motion algorithm uses to classify a set of connected regions to correspond to an object is that

$$\sum_{i=1}^{N_j} \left| I_k(r_i,c_i) - I_{k+1}(r_i,c_i)\right| > \tau_{m2},$$

where N_j is the total number of marked pixels in the j-th set of connected regions.

4.2. Learning Process

The three parameters of the motion detection algorithm – Δt, τ_{m1}, τ_{m2} – depend on image properties, particularly image contrast. Hence, these parameters need to adapt to the varying imaging conditions along the navigation route. The parameter adaptation constitutes the learning process. In our example, we apply a clustering technique for learning. We characterize the triplet (Δt, τ_{m1}, τ_{m2}) by the average image contrast. Thus, similar contrast values imply similar selection of triplets (after evaluating motion detection results) as belonging to an existing cluster, in which case the cluster center and variance are updated, or a new cluster is created with the current selection.

Every time a new frame is acquired its average contrast is computed to determine the relevance of using the existing triplet. In an unconstrained environment, it is unlikely that the same triplet will continue to give a high-level of performance after several frames. When this happens a new parameter triplet needs to be generated. If the new contrast value falls within an existing cluster then the new parameter values are interpolated from the existing members of the cluster. When the contrast value is located outside, the parameter values are extrapolated from the nearest cluster. In either case, the results of motion detection are evaluated. If the results are acceptable (based on the criterion given in section 4.3) then the new triplet is passed on to the parameter learning process. The triplet is discarded if the results are unacceptable and new parameter values need to be regenerated. In our example, we put upper and lower bounds on the value of each parameter. The regeneration process begins with the discarded values and follows a divide-and-conquer approach. For each parameter, the mean between the lower bound and discarded value and that between the upper bound and discarded value are computed. Thus, there are $2^3 = 8$ new triplets generated for each pass of the divide-and-conquer method. The motion detection result for each triplet is evaluated and the best is retained for subsequent passes of the divide-and-conquer mechanism. The process terminates when a pass fails to provide a triplet that enhances the results.

4.3. Evaluation Criterion

Evaluation of the motion detection results is important for new parameter generation and parameter learning. In our example, the evaluation is performed by verifying the support from the segmentation and other appropriate pre-attentive processes, e.g., edge detection, for the motion detection results. We note that because of frame differencing the pixels interior to the moving object boundary are less likely to be marked in the difference image than the boundary pixels. Thus, a good motion detection quality measure is the degree of overlap of the motion boundary, the segmented region boundary, and the detected edges:

$$Q = \frac{n(E_k \cap S_k \cap M_d) + n(E_{k+1} \cap S_{k+1} \cap M_d)}{n(M_d)}$$

where $n(\Omega)$ = number of elements in the set Ω, $E_i \equiv$ set of detected edgels in frame i, $S_i \equiv$ set of region border pixels in frame i, $M_d \equiv$ set of motion boundary pixels in the difference frame created from frames i and i+1.

4.4. Results

In our implementation of the moving obstacle detection algorithm, the successive frames are acquired at a fixed time interval of $\Delta t = 1$ second. The threshold parameters $\tau_{m1} \in [2, 8]$ and the parameter $\tau_{m2} \in [50, 150]$. The acceptable level of quality measure is 0.6.

At a certain instant during navigation, the successive frames of figure 3(a) and (b) are acquired. The images show an object (a mini-robot) in the lower right portion of the frames which is moving away from the mobile agent. The motion regions detected in the difference image after application of the thresholds τ_{m1} and τ_{m2} are shown in figure 3(c) with the thresholded difference superimposed on figure 3(b). The average contrast of figure 3(b) is 3.41; the values of the thresholds that result in acceptable quality of motion detection at this level of image contrast are $\tau_{m1} = 3$ and $\tau_{m2} = 71$. Several frames later, the ambient lighting condition changes abruptly. The two consecutive frames at this instant are shown in figures 4(a) and (b), with the average image contrast of figure 4(b) being 8.66. Because of the new imaging conditions, the thresholds τ_{m1} and τ_{m2} need to be adjusted.

The process of parameter learning begins by deriving a new set of values for τ_{m1} and τ_{m2} from the past experience of the agent under similar imaging conditions. Figure 5 shows the frequency of occurrences of different contrast values in the past. The threshold parameters corresponding to the mean of the largest distribution are $\tau_{m1} = 3$ and $\tau_{m2} = 73$. These values are selected for processing the frame difference of the figures 4(a) and (b) since the distribution is the closest to the current value of the contrast. The result of applying these threshold values to the difference image is shown in figure 4(d). Clearly, the low threshold values give rise to spurious motion regions that do not coincide with region boundaries and edge pixels, thus reducing the overall quality measure below the acceptable level. The successive generations of new parameter values through the divide-and-conquer method and the corresponding evaluation results are listed in table 1. Note that only the best of the $2^2 = 4$ combinations during each iteration is included in the table.

In the third iteration, the values of $\tau_{m1} = 6$ and $\tau_{m2} = 130$ produce motion detection results of acceptable quality. The difference image using these two threshold values is shown superimposed on the second frame of this sequence in figure 4(c). These parameter values are learned by the agent to update the distribution of figure 5. A comparison of figures 4(c) and (d) indicates the importance of adapting the threshold parameters in order to reduce the number of potential focus-of-attention for the attentive processes to identify the obstacle.

5. Conclusions

We have applied learning techniques to a perception system of an autonomous mobile agent. The learning techniques provide greater flexibility and robustness to the entire system, allowing the autonomous mobile agent to operate in a wider range of environments while performing a larger number of tasks. Furthermore, in order to greatly speed up the perception processing, we have based our perception system on an intelligent selective sensing paradigm. Through a two stage mechanism consisting of pre-attentive and attentive processes, this selective sensing strategy eliminates the need to process massive amounts of unnecessary sensory data. Thus, perceptual processing is performed in near real-time.

In the perception process, we have identified and applied learning at three different levels: 1) *parameter learning*, where the parameters that control the algorithms in the pre-attentive and attentive processes are dynamically modified as conditions change; 2) *utility determination*, where the dynamic reasoner determines a utility or 'usefulness' for the algorithms running in the pre-attentive and attentive processes; and 3) cognitive learning, where knowledge is accumulated in order to perform goal-driven algorithms in the pre-attentive and attentive routines.

We describe one example of parameter learning in our system, namely, that of a set of pre-attentive routines which detect obstacles. When using a clustering technique for learning, we are able to eliminate false alarms by altering the motion detection parameters based on the changing image properties. We plan to continue the development of our perceptual system along with the applied learning algorithms and processing routines.

number of iterations	τ_{m1}	τ_{m2}	Q
1	3	73	0.23
2	5	111	0.55
3	6	130	0.66

Table 1: Generation of threshold parameter values and evaluation quality measure.

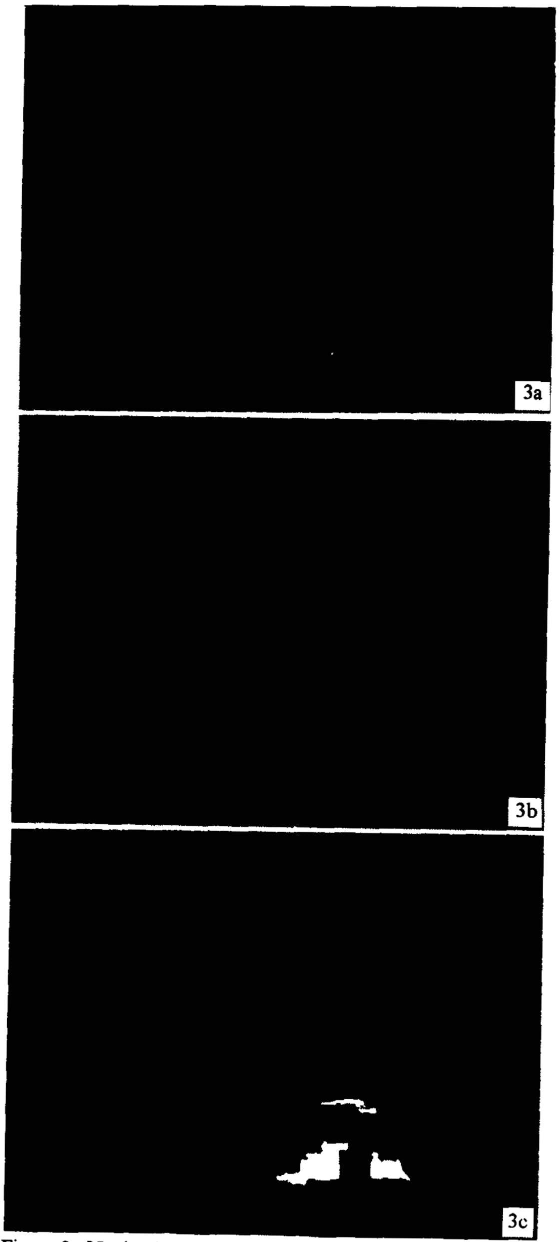

Figure 3: Navigation scene with object moving away under low contrast. a) Frame i; b) Frame i+1; c) Thresholded frame difference result superimposed on figure 3b).

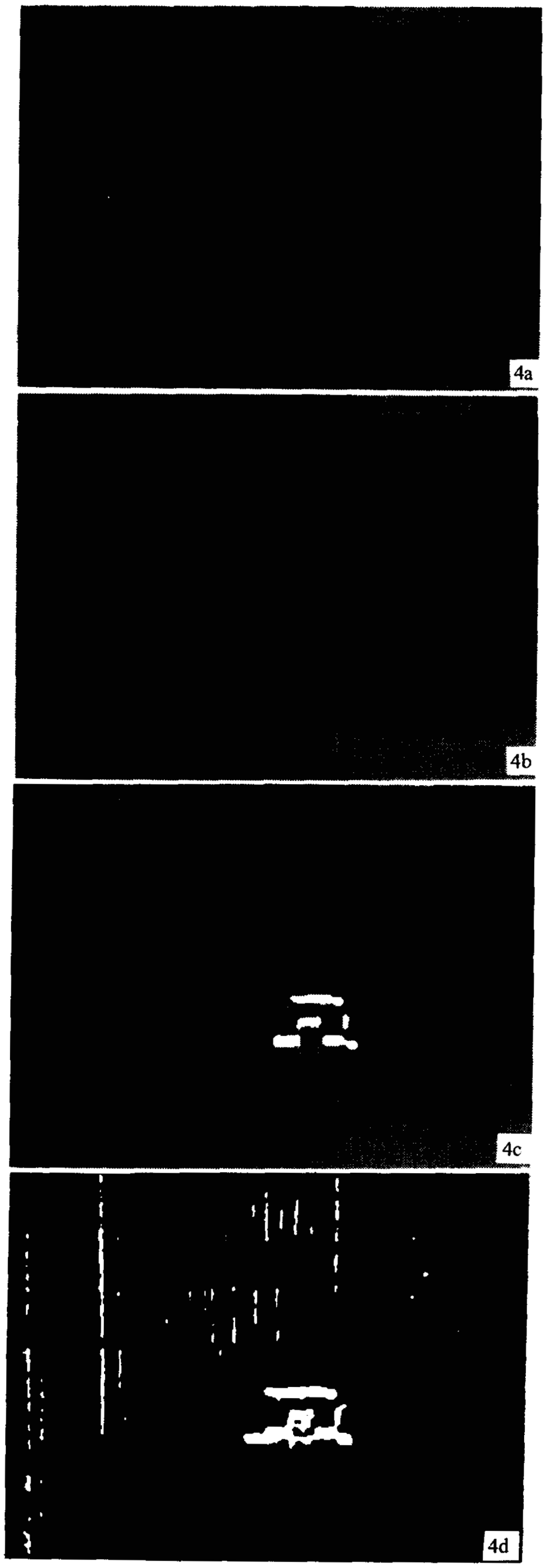

Figure 4: Navigation scene under high contrast. a) Frame i; b) Frame i+1; c) Thresholded frame difference result superimposed on figure 4b); d) Incorrectly thresholded difference image.

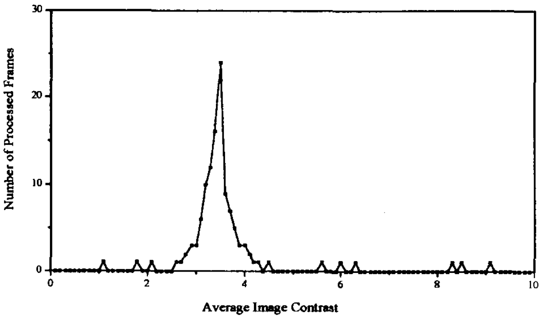

Figure 5: Contrast distribution of frame sequence.

6. References

[1] A. L. Abbott and N. Ahuja, "Surface Reconstruction by Dynamic Integration of Focus, Camera Vergence, and Stereo," *Proc. Second Intl. Conf. Computer Vision*, pp. 532-543, Tarpon Springs, Fla., Dec., 1988.

[2] J. Aloimonos and D. Shulman, "Learning Shape Computations," *Proc. of DARPA Image Understanding Workshop*, pp. 862-873, 1987.

[3] R. Bajcsy, "Active Perception vs. Passive Perception," *Proc. Workshop on Computer Vision*, pp. 55-59, Bellaire, Mich., Oct., 1985.

[4] M. Barth, "Robotic Attentive Sensing and its Application to Integrated Circuit Wafer Inspection," Ph.D. thesis, University of California, Santa Barbara, August, 1989.

[5] M. Barth and S. Hackwood, "A Flexible Small-Scaled Parallel Vision System," *Proc. IASTED Intl. Conf. Robotics and Automation*, pp. 7-11, Santa Barbara, Calif., May, 1988.

[6] B. Bhanu, S. Lee, and J.C. Ming, "Adaptive Image Segmentation," *Submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990.

[7] B. Bhanu, "Machine Learning in Computer Vision," Honeywell Systems & Research Center Technical Report, April 1988.

[8] R. M. Bolle, A. Califano, and R. Kjeldsen, "Data and Model Driven Focus of Attention," *Proc. 10th Intl. Conf. on Pattern Recognition*, pp. 1-7, Atlantic City, N.J., June, 1990.

[9] P. J. Burt, "Algorithms and Architectures for Smart Sensing," *Proc. DARPA Image Understanding Workshop*, pp. 139-153, Cambridge, Mass., April, 1988.

[10] J.H. Connell and M. Brady, "Learning Shape Descriptions," *Proc. of 9th International Joint Conference on Artificial Intelligence*, pp. 922-925, 1985.

[11] R.J.P. deFigueiredo and K.H. Wang, "An 'Evolving Frame' Approach to Learning with Applications to Adaptive Navigation," *Proc. of IEEE International Conference on Systems, Man and Cybernetics*, pp. 51-54, 1986.

[12] R. L. Didday and M. A. Arbib, "Eye Movements and Visual Perception: A Two Visual System Model," *Intl. J. Man-Machine Studies*, vol. 7, pp. 547-569, 1975.

[13] T.G. Evans, "A Program for the Solution of a Class of Geometric Analogy Intelligence Test Questions," *Semantic Information Processing*, M. Minsky (ed.), Cambridge, Mass.: MIT Press, 1968.

[14] J. Grefenstette and C.B. Pettey, "Approaches to Machine Learning with Genetic Algorithms," *Proc. of IEEE International Conference on Systems, Man and Cybernetics*, pp. 55-60, 1986.

[15] D. Hutber and P. Sims, "A Method for Learning Rules from Uncertain Data Applied to the Computer Vision Problem," *Proc. of Graphics Interface '86 and Vision Interface '86*, pp. 361-365, 1986.

[16] H. Inoue, H. Mizoguchi, Y. Murata, and M. Inaba, "A Robot Vision System with Flexible Multiple Attention Capability," *Proc. Intl. Conf. Advanced Robotics*, Tokyo, Japan, 1985.

[17] R. Jain, "Extraction of Motion Information from Peripheral Processes," *IEEE Trans. Pattern Analy. and Machine Intell.*, PAMI-3, no. 5, pp. 489-503, 1981.

[18] J. Kim and B. Bhanu, "Region Motion Detection Using the Color Difference Picture," Honeywell Systems & Research Center Technical Report, June, 1987.

[19] B.F. Kim, J. Bohandy, and V.G. Sigillito, "Multilevel Image Classifier," *Journal of Optical Society of America*, vol. 3, no. 9, pp. 1558-1560, 1986.

[20] M.B. Lehrer and G. Reynolds, "Initial Hypothesis Formation in Image Understanding Using an Automatically Generated Knowledge Base," *Proc. of DARPA Image Understanding Workshop*, pp. 521-537, 1987.

[21] J.C. Ming and B. Bhanu, "A MultiStrategy Learning Approach for Target Model Recognition, Acquisition, and Refinement," *Proc. DARPA Image Understanding Workshop*, pp. 742-756, Pittsburgh, PA, 1990.

[22] W.A. Perkins, "A Learning System that is Useful for Industrial Inspection Tasks," *IEEE Workshop on Industrial Applications of Machine Vision*, pp. 160-167, May 1982.

[23] J. Segen, "Learning Structural Descriptions of Shape," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 96-99, 1985.

[24] A. Shmuel and M. Werman, "Active Vision: 3D from an Image Sequence," *Proc. 10th Intl. Conf. Pattern Recognition*, pp. 48-54, Atlantic City, N.J., June, 1990.

[25] X. Shun-en and T.W. Calvert, "Incremental Construction of 3-D Models from a Sequence of Framed Views: Matching Partial Objects," *Proc. of Graphics Interface '86 and Vision Interface '86*, pp. 300-306, 1986.

[26] P. F. Spelt, G. de Saussure, E. Lyness, F. G. Pin, and C. R. Weisbin, "Learning by an Autonomous Robot at a Process Control Panel," *IEEE Expert*, vol. 4, no. 4, pp. 8-16, 1989.

[27] T. Toriu, T. Goto, and M. Yoshida, "An Investigation of Adaptable Vision Systems for Factory Automation," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 222-224, 1985.

[28] A. Treisman, "Preattentive Processing in Vision," *Computer Vision, Graphics, and Image Processing*, 31(2), 1985.

[29] W. Tsai and Y. Chen, "Adaptive Navigation of Automated Vehicles by Image Analysis Techniques," *IEEE Transactions on System, Man and Cybernetics*, SMC-16, no. 5, pp. 730-740, 1986.

[30] S. Ullman, "Visual Routines," *Cognition*, vol. 18, pp. 97-157, 1984.

[31] S. D. Whitehead, D.H. Ballard, "Learning to Perceive and Act by Trial and Error", *Machine Learning* 7, 1, 45–83, 1991.

[32] A. Yarbus, "Eye Movements and Vision," New York: Plenum Press, 1967.

[33] J. Y. Zheng, M. Barth, S. Tsuji, "Qualitative Route Description Using Autonomous Landmark Detection," *Proceedings of International Conference on Computer Vision*, Osaka, Japan, December, 1990.