

On the Relative Complexity of Active vs. Passive Visual Search

JOHN K. TSOTSOS

Department of Computer Science, University of Toronto, and Canadian Institute for Advanced Research, Toronto, Ontario M5S 1A1

Abstract

Here, this author attempts to tie the concept of active perception to attentive processing in general and to the complexity level analysis of visual search described previously; the aspects of active vision as they have been currently described form a subset of the full spectrum of attentional capabilities. Our approach is motivated by the search requirements of vision tasks and thus we cast the problem as one of search preceding the application of methods for shape-from-X, optical flow, etc., and recognition in general. This perspective permits a dimension of analysis not found in current formulations of the active perception problem, that of computational complexity. This article describes where the active perception paradigm does and does not provide computational benefits along this dimension. A formalization of the search component of active perception is presented in order to accomplish this. The link to attentional mechanisms is through the control of data acquisition and processing by the active process. It should be noted that the analysis performed here applies to the general hypothesize-and-test search strategy, to time-varying scenes as well as to the general problem of integration of successive fixations. Finally, an argument is presented as to why this framework is an extension of the behaviorist approaches to active vision.

1 Introduction

In 1985, Bajcsy presented a view of perception that she termed *active perception* [Bajcsy 1985]. She proposed that a passive sensor be used in an active fashion, purposefully changing the sensor's state parameters according to sensing strategies. Basically, it is a problem of intelligent control applied to the data-acquisition process that depends on the current state of data interpretation including recognition. It is interesting to note that this use of the term "active" in vision appeared in the dissertation of Freuder in 1976. He used what he called "active knowledge" to assist in determining where to look next in an image based on the current state of interpretation [Freuder 1976]. He did not include considerations of camera movements, but the general idea is the same. Bajcsy detailed a proposal for an experimental environment for conducting research on this topic. Cameras should have focus, zoom, and aperture control: the camera mounts should permit vergence angle control, pan/tilt, up/down, and right/left motions. The observer should know the camera's position, orientation, focus, zoom, and in general all the external param-

eters of the system. Internal parameter adjustment was not involved in the proposal (but was in Freuder's system).

Since then, many elaborations of the idea have appeared [Bandyopadhyay et al. 1986; Krotkov 1987; Clark & Ferrier 1988; Ballard 1987, 1989; Ballard & Ozcandarli 1988; Aliomonos et al. 1987; Abbott & Ahuja 1988; Paul et al. 1987; and others]. A particularly good example of the application of active perception is given by Aloimonos et al. [1987]. They assume that the observer is active and the observer's purpose is to control the geometric parameters of the sensory apparatus. The idea is applied to the computations of shape-from shading, contour, texture, motion, and optical flow. They show how the additional data provided by several views in time can convert an ill-posed problem into one that is well posed. Moreover, this often allows one to use more general assumptions and increases robustness to noise. However, in general, the system needs accurate knowledge of the viewing transformation up to second derivatives.

It is important to note that although the concept of active perception is relatively new to computer vision,

it is not new in the psychology of perception. In 1874, Franz Brentano introduced the concept of *act psychology*. He raised the possibility that a subject's actions play a role in perception and that perception and indeed all conscious acts must be grounded by real objects. G.E. Müller defined his *Komplextheorie* of collective attention in 1904 where perception was based on actions. See Metzger [1974] for an overview of these and other early ideas. To add further perspective to the current work on active vision, we should note that the approach seems to represent a marriage between two historically important points of view. Helmholtz believed in perceptual hypotheses, in the derivation of the best interpretation given the evidence and in attentional mechanisms that guide processing even without eye movements [Helmholtz 1910].¹ Gibson, on the other hand, abandoned the usual experimental paradigms that used 2D images and argued for an eye that moves freely in a natural 3D world [Gibson 1979]. Of course, both scientists included much more in their theories; these two particular viewpoints however characterize each of them and, together, form a reasonable view of active perception as currently popular.

Active perception necessarily must address the problem of the integration of successive fixations of a scene into a coherent whole. This has been a subject of study within a psychology community, and is closely tied in with eye-movement research.² Computational proposals for the general problem have not previously appeared; this article makes a preliminary step in that direction.

The particular formalization that will be presented in successive sections addresses not only the problem of active perception; with no changes whatsoever, the same formalism and results apply to long sequences of time-varying scenes, whether the sensor is moving or not. The results will therefore be undistinguished along this line.

2 Attention and Search

The concept of attention can be found in even the earliest writings on perception (for a good collection of recent papers, see Parasuraman & Davies [1984]; a short review appears in LaBerge [1990]). Helmholtz believed that a conscious or voluntary effort may focus attention on a particular spot in the visual field [Helmholtz 1910], and this led to the attentional spotlight idea that is widespread in models of perception. It is well known that human perceptual systems do not

completely analyze all incoming stimuli to the same degree, and that certain elements are attended to and others are not. This is how the psychology community has recognized, in an informal way, that the task of visual perception is computationally intractable from a search perspective, and attention represents their mechanism for achieving a solution. The immediate computational counterpart is simply to include in any theory the ability to select which stimuli are to be processed and to what degree.

Visual search is a common if not ubiquitous subtask of vision, in both man and machine. A basic visual search task is defined as follows: given a target and test image, is there an instance of the target in the test image? [Rabbitt 1978]. Typically, experiments measure the time taken to reach a correct response. Region growing, shape matching, structure from motion, the general alignment problem, and connectionist recognition procedures are specialized versions of visual search in that the algorithms must determine which subset of pixels is the correct match to a given prototype or description. The basic visual search task is precisely what any model-based computer vision system has as its goal: given a target or set of targets (models), is there an instance of a target in the test display? Even basic vision operations such as edge finding are also in this category: given a model of an edge, is there an instance of this edge in the test image? It is difficult to imagine any vision system that does not involve similar operations. It is clear that these types of operations appear from the earliest levels of vision systems to the highest.

In Tsotsos [1989], a computational definition of the visual search task was presented, and the unbounded case was distinguished from the bounded case.³ An equivalence was drawn between unbounded search and bottom-up processes, and bounded search and task-directed visual processes. Then, a proof was given showing that the unbounded case is NP-Complete in the size of the image, while the bounded case has linear time complexity in the same variable. The NP-Completeness of the unbounded case is due solely to the inability to predict in a nonexponential manner which pixels of a test image correspond to objects. It is claimed without proof that problems such as those listed above are therefore NP-Complete.

These results provide the strongest possible evidence for the abandonment of purely bottom-up schemes that address the full generality of vision. It is thus necessary to sacrifice generality in order to reshape the vision problem and to optimize the resources dedicated to

visual information processing so that a tractable problem is addressed. Attention is one important ingredient of the optimization process.

Figure 1 shows the spectrum of attentional mechanisms—it is not complete. The spectrum is organized by size of the “space” selected by attention, where space does not refer only to the three-dimensional world. The largest selection is that of task, then of the world model within which the task is to be solved, then the selection of 3D visual space that is relevant, then the selection of subsets of visual space, then selection of subsets of computing units to apply, and finally, the selection of operating parameters of each unit. Adaptation is the lowest form of attentional manifestation in this categorization, the attentional beam mechanism is next and actions of the oculomotor system are above that. In Tsotsos [1990], the concept of an attentional beam was derived using complexity constraints in visual search tasks. In visual search, the task, world model, and visual space are preselected as part of the experimental conditions. Adaptation may be relevant for the experiment as a whole rather than on a case-by-case basis. Once the subject has adapted to the experimental conditions, adaptation may play a lesser role. Selection of subunits however is relevant during the course of the experiment, and for this reason the beam falls out

naturally as an important attention mechanism. Of course the others have impact, but they seem to be already fixed by the time the experiment is well underway. Active perception as commonly formulated seems to be linked with the third and fourth layers of the attentional spectrum.

3 Active Vision: Why?

Several reasons for the need for active approaches to perception have been put forward. Summarizing, active vision is useful in at least the following ways:

- to see a portion of the visual field otherwise hidden
- to compensate for spatial nonuniformity of a processing mechanism
- to increase spatial resolution
- to disambiguate aspects of the visual world (through induced motion, or lighting changes for example)
- to enable better mathematical formulations for a particular problem.

All of the above (except perhaps the last point) tacitly assume that some hypothesize-and-test mechanism is at work. Only if hypotheses are available, can a particular action due to an active perception mechanism

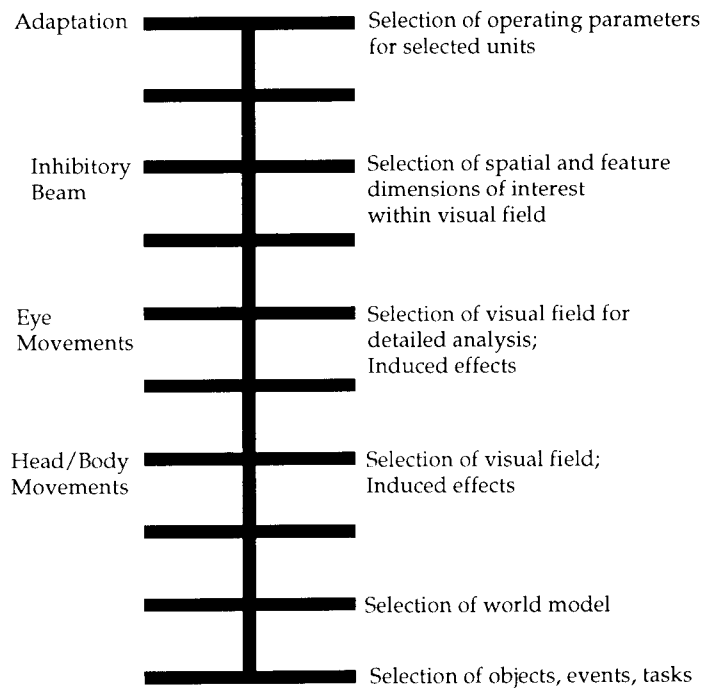


Fig. 1. The spectrum of attentional mechanisms.

actually yield benefits. Otherwise the search space is simply too large; that is, attention is needed. This is not to say that reasoning and sophisticated knowledge representation schemes are necessary; quite the contrary, as the next section will show. Thus, this article proposes a much broader view of active vision so that it encompasses all levels of the attentional spectrum, that is, all levels of visual information processing that require hypothesize-and-test.

4 Complexity of Passive vs. Active Visual Search

The issue of computational efficiency for active approaches is explored here with the goal of showing that, in some cases, there is another basic computational reason for active perception, namely, to gain improvements in computation time. This is a consideration orthogonal to the above list. In the following sections, comparative complexity functions will be presented for various types of visual search tasks in both passive and active formulations. The specific situations described above are modeled only with the number of views that they require regardless of the reason for more than one view of the scene.

In most discussions of active perception, little consideration is given to the additional cost imposed on a perceptual system if it must manage an active sensor. The costs fall into at least the following computational categories:

1. deciding that a change of visual field is needed and why (why will help determine which change is best)
2. deciding on which change is best (priority sequence: eye movements, head movements, body movement)
3. execute change
4. adapt system to new viewpoint (focus, light levels, etc.)
5. correspondence of objects and events between old and new viewpoints.

These computations and actions may be relatively expensive in terms of overall time to understand a scene, or to perform some task. Deployment of an active strategy versus a passive one must therefore depend on the associated costs as well as the benefits.

It is important to note that in the remainder of this article, only an abstract framework suitable for complexity analysis is presented. This discussion does not address how the framework may be implemented. As such, the discussion may be likened to the “in principle” solutions of Marr’s computational level [Marr 1982].

4.1 Setting Up the Framework

Two key definitions are required to begin the specification of a computational formalism for visual search: (1) unbounded visual search in which either the target is explicitly unknown in advance or it is somehow not used in the execution of the search; and (2) bounded visual search, in which the target is explicitly known in advance in some form that enables explicit bounds to be determined that can be used to limit the search process. These bounds may be in the form of spatial extent of the target, feature dimensions that are involved or specific feature values.

A test image containing an instance of the target is created by placing an instance of the target, which may have undergone translating, rotating, and/or scaling, in the test image. We assume for the moment that such spatial transformations are not present. The test image may also contain confounding information, such as other items, noise, and occluding objects; or other processes may distort or corrupt the target.

The typical visual search experiments found in the literature all deal with 2D images of 2D targets. An active approach would yield no difference in the image acquisition or understanding process since all data is available to the sensors at all times.⁴ One must consider the broader forms of visual search possible in the 3D world, where the targets:

- may be occluded by another object or may be self-occluded thus requiring sensor motion or object manipulation to eliminate the occlusion;
- may be outside the fovea or field of view, thus requiring sensor or observer motion to foveate the target;
- may be far away, thus requiring sensor zoom or observer motion to bring details into view;
- may be out of focus if it is out of the camera depth of field (or out of the stereo-vergence plane in a binocular system), thus requiring focusing actions.

These sensor motions may for some problems be necessary in order to acquire sufficient data to complete the task. There is a different class of motions as well that seem to not be required in principle, but rather can make problem solution simpler. They include:

- induced lighting changes (photometric stereo, for example)
- induced motion effects (kinetic depth, for example)

In order to permit a more formal analysis of active vision, a formalization of the visual search problem,

both passive and active, must first be presented. The formalization first appeared in Tsotsos [1989], but is extended here to the active version of the problem.

4.2 The Role of Worst-Case Analysis

Elsewhere [Tsotsos 1990b; 1991], the relevance of a worst-case analysis was briefly discussed; here this discussion is expanded. A worst-case analysis provides an upper bound on the amount of computation that must be performed as a function of problem size. If one knows the maximum problem size, then the analysis places an upper bound on computation for the whole problem as well. Thus, one may then claim, given an appropriate implementation of the problem solution, that processors must run at a speed dependent on this maximum in order to ensure real-time performance for all inputs in the world. Worst cases do not occur only for the largest possible problem size; rather, the worst-case time-complexity function for a problem states that for any problem size the worst-case number of computations may be required simply because of unfortunate ordering of computations (for example, a linear search through a list of items would take a worst-case number of comparisons if the item sought is the last one). Thus, worst-case situations in the real world may happen frequently for any given problem size.

Many argue that worst-case analysis is inappropriate for the problem at hand for one of the following reasons:

1. Relying on worst-case analysis and drawing the link to biological vision implies that biological vision handles the worst-case scenarios.
2. Biological vision systems are designed around average or perhaps best-case assumptions.
3. Expected case analysis more correctly reflects the world seen by biological vision systems.
4. The definitions used for bounded and unbounded search are artificial and do not relate to real systems.

Each of these criticisms will be addressed in turn.

1. This kind of inference is quite incorrect. As was shown in Tsotsos [1990a; 1990b]:

— Bounded search corresponded exactly to the known target visual search experiments commonly seen in the psychology literature (Treisman [1988] for example). The predicted performance agrees completely with the observed performance.

— Unbounded search, since it is NP-Complete, points to the need for major optimizations and abstractions in vision systems regardless of whether they are biological or not. It cannot be the case that the brain is solving problems that have exponential time complexity where the exponentials are large. These optimizations and abstractions included the mechanisms of parallelism, visual-model hierarchical organization, input hierarchical abstraction, visual maps, and spatio-temporal receptive fields. When used together, they lead to an architecture with biologically plausible time- and space-complexity relationships [Tsotsos, 1988; 1990a]. These mechanisms are common throughout most early vision proposals.

Thus, there is no implication that human vision handles all worst-case scenarios at all. The whole argument exists only to prove that all worst-case scenarios cannot be handled by human vision in a bottom-up fashion.

2. It is far from obvious what kind of assumptions (if any) went into the design of biological vision systems. Vision systems emerged as a result of a complex interaction of many factors including a changing environment, random genetic mutations and competitive behavior. It is probably the case that the best we will ever be able to do under such circumstances is to place an upper bound on the complexity of the problem, and this is all worst-case analysis will provide.

3. Analyses performed by other authors (Grimson [1988], for example) based on expected or average cases, depend critically on having a well-circumscribed domain and an algorithm. Thus the complexity measures derived reflect algorithmic complexity and not problem complexity as is the goal of the present paper.⁵ Only under those conditions can average or expected case analyses be performed. In general, it is not possible to define what the average or expected input is for a vision system in the world. Furthermore, the result of the analysis will be valid only for the average input, and does not place a bound on the complexity of the vision process as a whole. This also would not provide any guidance in the determination of required processor for real-time performance. See also Uhr [1990].

4. The definitions of the two search types are not intended as a basis for implementing a vision system. They are abstractions of the problem, descriptions of “in principle” solutions at what Marr called the “computational level” [Marr 1982]. As abstractions, one can begin to analyze certain (but not all) of the properties

of such solutions. One of the properties is the required time complexity as a function of problem size. This is a very common kind of analysis within computer science (see Garey & Johnson [1979]). How these definitions are to be used in implementations must involve further investigation of their robustness to noise, and other properties; but their time complexity would not change problem class by such further elaboration (say, from NP to P).

4.3 Specifying an Instance of Visual Search

An instance of the “bounded visual-search” problem is specified as follows:

A test image I

A target image T

A difference function $\text{diff}(a)$ for $a \in I$, $\text{diff}(a) \in R_\rho^0$, (R_ρ^0 is the set of nonnegative real numbers of fixed precision ρ)

A correlation function $\text{corr}(a)$ for $a \in I$, $\text{corr}(a) \in R_\rho^0$

Two thresholds, θ and ϕ , both positive integers

The unbounded visual-search problem is specified in the same manner but without the target image. Details on how this particular collection of data may represent the visual-search problem follow.

4.3.1. A test image I is the set of pixel/measurement quadruples (x, y, j, m_j) . x, y specify a location in a Euclidean coordinate system, with a given origin. There are p unique image locations (or x, y pairs). Note that the locations are not necessarily spatially contiguous. M_I is the set of measurement types in the image, such as color, motion, depth, etc., each type coded as a distinct positive integer. m_j is a measurement token of type j , represents scene parameters, and is a nonnegative real number of fixed precision, that is with positive error due to possible truncation of at most ρ . (Only a finite number of bits may be stored.) $I' \subseteq I$ is a sub-image of I , that is, an arbitrary subset of quadruples. It is not necessary that all pixel locations contain measurements of all types. For ease of notation, $a_{x,y,j}$ refers to a particular test image tuple and has value m_j . If $j \notin M_I$ or if the x, y values are outside the image array, then $a_{x,y,j} = 0$.

4.3.2. A target image T is a set of pixel/measurement quadruples defined in the same way as I . The set of locations is not necessarily spatially contiguous. There are q unique locations in the target image. M_T is the set of measurement types in the target image. The types correspond between I and T , that is, type 3 in one image is the same type in the other. The two sets of measurement types, however, are not necessarily the same. The coordinate system of the target image is the same as for the test image and the origin of the target image coincides with the origin of the test image. $t_{w,z,j}$ refers to a particular target image tuple and has value m_j . If $j \notin M_T$ or if the w, z values are outside the image array, $t_{w,z,j} = 0$.

4.3.3. The processing proceeds as follows. A subset of I, I' , is chosen, a coordinate system is identified, and the functions diff and corr are computed with respect to that subset and the target. The diff function will be the sum of the absolute values of the point-wise differences of the measurements of a subset of the test image with the target image. It is expressed as follows for an arbitrary subset I' of the test image:

$$\sum_{a \in I'} \text{diff}(a) = \sum_{a \in I'} |t_{w,z,j} - a_{x,y,j}| \leq \theta$$

This sum of differences must be less than a given threshold θ in order for a match to be potentially acceptable. Note that other specific functions that could find small enough values of some other property may be as suitable. The threshold is a positive integer.

4.3.4. Since a null I' satisfies any threshold in the above constraint, we must enforce the constraint that as many figure matches must be included in I' as possible. 2D spatial transforms that do not align the target properly with the test items must also be eliminated because they would lead to many background-to-background matches. One way to do this is to find large enough values of the point-wise product of the target and image. This is also the cross correlation commonly used in computer vision to measure similarity between a given signal and a template. A second threshold, ϕ , provides a constraint on the acceptable size of the match. Therefore,

$$\sum_{a \in I'} \text{corr}(a) = \sum_{a \in I'} (t_{w,z,j} \times a_{x,y,j}) \geq \phi$$

The instance of visual search described here applies exactly as given for scenes where motion is represented

by an instantaneous velocity field. If a long sequence of images is considered, the only change is to redefine an image element as a quintuple (x, y, t, j, m_j) where t spans the image sequence. In general, an image element may be an arbitrary n -tuple and all proofs and results given below hold.

4.4 Passive Unbounded Visual Search

Given a test image, a difference function, and a correlation function, is there a subset of pixels of the test image such that the difference between that subset and the corresponding subset of pixels in the target image is less than a given threshold and such that the correlation between the two is at least as large as another specified threshold? In other words, is there a set $I' \subseteq I$ such that it simultaneously satisfies $\sum_{a \in I'} \text{diff}(a) \leq \theta$ and $\sum_{a \in I'} \text{corr}(a) \geq \phi$?

One point regarding the above specification of visual search must be emphasized. The target image is not permitted to provide direction to any aspect of the computation other than in the computation of the diff and corr functions.⁶ The constraints given must be satisfied with subsets of the input image. This definition involves two constraints to be simultaneously satisfied and that the two constraints represent error and size satisfaction criteria. Note that this definition does not force acceptance of only the best match, but accepts any sufficiently good match. This is very similar to many other kinds of recognition definitions. For example, the standard definition of region growing involves the maximal contiguous subset of pixels that satisfies a given property. Moreover, this definition should not be interpreted as a template-matching operation. Although template matching may be posed in the above manner, there is nothing inherent in the definition to exclude other matching forms—the notions of image points, measurements, and constraints representing large enough cover and low enough error are ubiquitous in matching definitions.

THEOREM 1. *Unbounded visual search is NP-Complete.*

Proof: This was proved in Tsotsos [1989] by reduction to Knapsack.

On the assumption that no prior information nor assumptions are available for this problem,⁷ the number of arithmetic operations for the unbounded visual-search problem is given by

$$O(|I| \cdot 2^{|I|})$$

where $|I|$ is the number of pixel/value tuples in the test image.⁸ If all measurements are represented at all pixel locations in the test image, then the expression becomes

$$O(p \cdot |M_I| \cdot 2^{|M_I|p})$$

$|I|$ is a product of $|M_I|$, the number of measurement types in the test image; and p is the number of pixel locations in the test image.

A concrete example of such kinds of search problems is provided by the well-known “Dalmatian sniffing at fallen leaves” image, or the “Horseman” of Verville and Cameron [1946], or the incomplete figures of Leeper [1935]. In each, the observer is not given any cues as to how to group image blobs, and must try arbitrary groupings until one of the grouping “clicks” onto a form that is known. This has exactly the character of the grouping operations required for the above computations; the computations involved in the “aha!” experienced when the correct grouping is found is the correlate of the diff and corr functions.

4.5 Passive Bounded Visual Search

If we consider the use of the target item, it is easy to show that the problem has linear-time complexity. The key is to direct the computation of the difference and correlation functions using the target rather than the test image. This kind of search task is common in computer vision and corresponds to any model-based vision task. However, we still seek the appropriate subset of the test image. If there is a match that satisfies the constraints, then its extent can be predicted in the test image; all locations are possible. The bounded visual-search task is stated as follows:

Given a test image, a target image, a difference function, and a correlation function, is there a subset of pixels of the test image such that the difference between that subset and the corresponding subset of pixels in the target image is less than a given threshold and such that the correlation between the two is as large as possible?

In other words, is there a set $I' \subseteq I$ such that it simultaneously satisfies

$$\sum_{t \in I'} \text{diff}(t) \leq \theta \quad \text{and} \quad \sum_{t \in I'} \text{corr}(t) \geq \phi$$

where

$$\sum_{t \in I'} \text{diff}(t) = \sum_{t \in I'} |t_{w,z,j} - a_{x,y,j}| \leq \theta$$

and

$$\sum_{t \in T} \text{corr}(t) = \sum_{t \in T} (t_{w,z,j} \times a_{x,y,j}) \geq \phi$$

Note that the definitions of *diff* and *corr* have changed slightly in that the pixel locations and measurement set used are those of the target rather than the test image. In other words, the hypothesizing of target-test subset correspondence is based on the target.

THEOREM 2. *Bounded visual search has time complexity linear in the number of test-image pixel locations.*

Proof. The computation of the *diff* and *corr* functions is driven by the target image and the measurements present in the target. A simple algorithm is apparent. Center the target item over each pixel of the test image; compute the *diff* and *corr* measures between the test and target image at that position; among all the positions possible, choose the first solution that satisfies the constraints. The worst-case number of arithmetic operations for bounded visual search would be $O(|T|p)$ or $O(q|M_T|p)$.

4.6 Active Visual Search

A different formulation of the visual search problem is needed for an active strategy. If an active approach is to yield efficiency benefits, then it must in some way reduce the search space required for problem solution when compared to passive approaches. This implicitly assumes that for a given problem active and passive strategies are equally able to solve the problem. Careful preplanning may eliminate the need for active methods in some cases. Consider the recognition of a 3D object: one passive strategy would be to simply take N views of the object spaced by a fixed amount, and reconstruct it. After all, this is the basis for 3D computed tomography, a very successful technology. The motion of the camera is not dynamically determined, but is rather preprogrammed. Contrast this with an active approach which may decide dynamically which are the best views to use for recognition depending on the context and task. Both schemes can yield a correct solution. The complexity issue is orthogonal. For purposed of comparison, it is assumed that both passive and active strategies are on equal footing. It should be pointed out that the set of problems referred to (i.e., the intersection of solvable passive and solvable active approaches) is

neither small nor unimportant. This is because the use of the term “active” in this paper is very broad. Included in the comparison are all time-varying vision tasks, as well as any approach that involves attentional processes, hypothesize-and-test, adaptive schemes; any process that fits into one or more of the levels of the spectrum of attentional mechanisms shown in figure 1 and involves sampling input over time. The use of the term is not restricted to strategies that move cameras only.

A critical feature of active approaches is that time is required, that is, a sequence of perceptual signals must be acquired over time, where the active strategy controls what signals are to be acquired and how they are to be processed. Given the efficiency consideration and the processing control possible using an active approach, we can reformulate the visual-search problem as follows:

Given a test image sequence in time I_t , is there a sequence of sets \mathcal{G}_t for $t = 1$ to τ , where \mathcal{G}_t is the union of all sets $I'_t \subseteq I_t$, such that each element I'_t of \mathcal{G}_t satisfies,

$$\sum_{a \in I'_t} \text{diff}(a) \leq \theta_t \quad \text{and} \quad \sum_{a \in I'_t} \text{corr}(a) \geq \phi_t$$

where

$$\theta_1 \geq \theta_2 \geq \dots \geq \theta_\tau$$

$$\phi_1 \leq \phi_2 \leq \dots \leq \phi_\tau$$

and

$$\theta_\tau = \theta \quad \text{and} \quad \phi_\tau = \phi$$

of the passive definition.

The sequence of thresholds may be trivially set to $\theta_{t+1} = \theta_t - 1$ and $\phi_{t+1} = \phi_t + 1$. θ_1 and ϕ_1 are positive integers. Additional analysis is needed to determine if better settings exist. The thresholds act as hypothesis-pruning filters tuned more and more tightly as time progresses. Therefore, this strategy and formalism works equally well for the general case of hypothesize-and-test with or without sensor motions, and for static or time-varying images.

THEOREM 3. *Active unbounded visual search is NP-Complete*

Proof: The active problem consists of a sequence of passive problems. Earlier, it was shown that the passive problem is NP-Complete. Thus, the active version is also NP-Complete.

The complexity of the unbounded version of this problem is still exponential, because it has the passive problem described earlier as a subproblem. Its worst-case complexity would be

$$O \left(\sum_{t=1}^{\tau} (|I_t| \cdot 2^{|I_t|}) \right)$$

because all viable hypotheses must be carried over from one image to the next, and only if all image subsets are considered will all viable hypotheses be found. If all measurements are present for all locations then the expression becomes

$$O \left(\sum_{t=1}^{\tau} (p_t \cdot |M_t| \cdot 2^{|M_t|p_t}) \right).$$

p_t is the total number of pixel locations from which to select candidate hypotheses at time t .

THEOREM 4. *Active bounded visual search has time complexity linear in the number of test-image pixel locations.*

Proof: The active bounded problem is a sequence of bounded passive problems, and therefore, it too has behavior linear in the number of pixels, and its complexity is given by

$$O \left(\sum_{t=1}^{\tau} (|T| \cdot p_t) \right)$$

This formulation is guaranteed to work correctly because the solution subset(s) are not discarded during the iterations. The thresholds are set up so that the correct solution, which would satisfy the thresholds defined in the passive case, will satisfy the thresholds of each iteration up to the last in the active case.

In essence, this formulation describes a hypothesize-and-test search framework, where at each time interval a number of hypotheses are discarded. This active strategy applies to both unbounded and bounded visual-search problems. This formulation enables partial solutions to be inspected midstream through an image sequence for the active strategy. The passive strategy would require waiting until all the images are acquired and then operating on a spatiotemporal block of data.

Note that \mathcal{H}_t represents that set of active hypotheses at each time step. In the unbounded version of the problem, hypotheses are arbitrary pixel/value groupings,

while in the bounded problem, hypotheses are pixel/value groups which all have the same size and configuration as the target. Although the constraints on the diff and corr functions can effectively eliminate hypotheses, the elimination of the pixel/value tuples which make up a hypothesis is a bit more difficult. A given pixel/value tuple may participate in a large number of hypotheses. If one of those hypotheses is eliminated, it does not mean that the pixel/value tuple may also be eliminated. All of the hypotheses in which a pixel/value tuple participates must be eliminated before the tuple is removed from further consideration.

Consider 2D projections in time of a 3D scene. From image to image, the same location may be present in the images, but with differing viewpoint. Thus, a given pixel of information in one image may be foreshortened in size, and may appear differently due to lighting changes. Even if its corresponding pixel is found, it is not necessarily the case that the pixel may be eliminated or saved. The different information may be important (as in a photometric stereo scheme). Further, the information may be confounded due to the foreshortening and may be sampled by the image sensor together with neighboring locations, all depending on the sensor resolution. Locations eliminated in one image remain eliminated for future images; duplicates are discarded and new measurements at viable locations are kept as long as the location is viable. So, the best we can do in characterizing the number of pixel/value tuples to consider at time t is to simply sum the pixel/value tuples carried over with those acquired and to subtract from this sum the number of tuples that can be eliminated as a result of hypothesis pruning. This is given by

$$H_t = H_{t-1} - e_{t-1} + h - e_{t-1} \quad (1)$$

where H_t is the number of pixel/value tuples at time t , h is the number of pixel/value tuples acquired each time sample, e_{t-1} is the number of pixel/value tuples eliminated from time sample $t - 1$ and found and eliminated at time t , ($e_0 = 0$), e_{t-1} which represents the number of pixel/value tuples eliminated from the image at time $t - 1$, ($e_0 = 0$). It should be clear that correspondence in general is difficult; it may be assisted by accurate knowledge of sensor motions. The value of H_t may grow monotonically as images are acquired for nontrivial scenes.

In an implemented system, the amount of memory that may be allocated to the storage of these hypotheses would be bounded by H_{\max} at any time t . In effect, the

storage of hypotheses forms a representation of the salient aspects of the visual world that are under consideration for solution of the problem currently being addressed by the system. This bound places an additional constraint on the selection of thresholds since the values of e_{t-1} and e_{t-1} , depend in part on the values of the thresholds. h is fixed by the imaging-system parameters. Therefore,

$$H_{\max} \geq H_t \geq H_{t-1} - e_{t-1} + h - e_{t-1}$$

$$= \sum_{a=1}^t (h - e_{a-1} - e_{a-1}) \quad (2)$$

For the passive analysis of a spatiotemporal block of data, the value of H_{\max} is necessarily $h\tau$, for τ images of size h . H_{\max} is smaller on average in the active case since hypotheses may be eliminated from one time instant to the next. If one knows H_{\max} and h in advance, then one may be able to select values of the thresholds θ and ϕ in such a way that

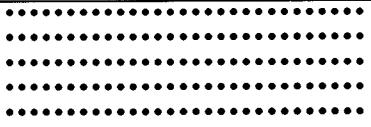
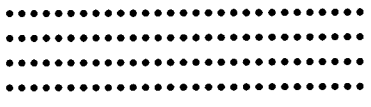
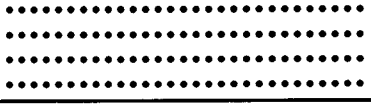
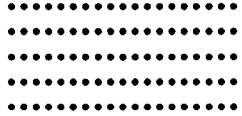
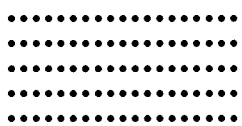
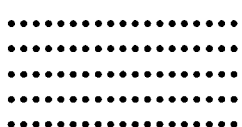
$$ht - H_{\max} \leq \sum_{a=1}^t (e_{a-1} + e_{a-1}) \quad (3)$$

is satisfied for all t . A greedy algorithm may suffice for this, choosing thresholds very conservatively to begin with, and increasing ϕ while decreasing θ until the constraint is just satisfied. The greedy algorithm may be guided by the fact that on average, $h - H_{\max}/\tau$ tuples must be eliminated for each time sample.

4.7 Worst-Case Time Complexity for Passive vs. Active Visual Search

Table 1 shows the worst-case time-complexity functions for passive and active visual search, both bounded and unbounded. A comparison is made between input representations of data as a spatiotemporal block (n -D plus time) versus a temporal sequence of spatial images. The derivation of these functions follows those given earlier in a straightforward manner. The expressions in the table assume with no loss of generality that the number of pixel/value tuples at each time t are given by $p_t |M_t|$. The additional variables used in table 1 below are: τ is the total number of time samples in data set; K_t is the overhead per time sample for active computations; and, C_t is the cost for correspondence computations in the active case for each time sample.

Table 1. Worst case time complexity for passive vs. active visual search.

		spatio-temporal block	temporal sequence
passive	A unbounded	$O(\tau p \cdot M_t \cdot 2^{ M_t \tau p})$	
	B unbounded optimized	$O(M_t \cdot 2^{ M_t } \cdot (\tau p)^{2.5})$	
	C bounded	$O(q \cdot M_T \cdot \tau p)$	
active	A unbounded		$O\left(\sum_{t=1}^{\tau} (p_t \cdot M_t \cdot 2^{ M_t p_t} + K_t + C_t)\right)$
	B unbounded optimized		$O\left(\sum_{t=1}^{\tau} (M_t \cdot 2^{ M_t } \cdot p_t^{2.5} + K_t + C_t)\right)$
	C bounded		$O\left(\sum_{t=1}^{\tau} (q \cdot M_T \cdot p_t + K_t + C_t)\right)$

4.8 When Does Active Approach Win?

What insights does the comparison in table 1 yield?

A. Active strategies are always more efficient for unbounded visual search problems. The sum of smaller exponentials of the active case will always be less than the single large exponential of the passive case for reasonable cost implementations. This also shows the power of hypothesize-and-test frameworks for unbounded perceptual tasks.

B. In Tsotsos [1988, 1990a], it was noted that even though the brute-force approach to the problem of visual search is intractable if the target and task are unknown, human perceptual systems still solve the task. It must be that the problem is reshaped and the perceptual processing machinery is optimized in order to do this. These same approximations and optimizations can be applied to the brute-force strategy described above for the active unbounded problem. The mechanisms of parallelism, visual model hierarchical organization, input hierarchical abstraction, visual maps and spatio-temporal receptive fields when used together, lead to biologically plausible time and space complexity relationships [Tsotsos 1988, 1990a] reducing the complexity of unbounded search tasks to $O(2^{|M_I|} \cdot p^{1.5})$. The bounded case is still linear and is not qualitatively improved by these optimizations. The passive unbounded problem complexity then becomes

$$O(|M_I| \cdot 2^{|M_I|} \cdot (\tau p)^{2.5})$$

and the comparable expression for the active unbounded problem is

$$O\left(\sum_{t=1}^{\tau} (|M_I| \cdot 2^{|M_I|} \cdot p_t^{2.5} + K_t + C_t)\right)$$

If the hypothesis pruning is good (p_t sufficiently less than p so as to offset the overhead cost), then the active approach may be more efficient.

C. If enough data is available in one image to solve the bounded problem, the active approach (over two or more images) is always less efficient than a passive single image solution. If the data must be found over a sequence of images, active strategies may be more efficient than passive ones for bounded problems only under certain conditions. In general the following constraint must be satisfied for the active approach to be more efficient (recalling that $|T| = q |M_T|$):

$$A(|T| \cdot \tau p) > B \left(\sum_{t=1}^{\tau} (|T| \cdot p_t + K_t + C_t) \right) \quad (4)$$

where A and B are constants. If the τ images are available over a period of S seconds, then each side of the above inequality is constrained by S . This can be restated in order to provide a constraint on the relative efficiency of the active overhead computations:

$$A(|T| \cdot \tau p) - B \left(\sum_{t=1}^{\tau} (|T| \cdot p_t) \right) > B \sum_{t=1}^{\tau} (K_t + C_t) \quad (5)$$

Assuming $A = B$, $|M_I| = 1$, and for ease of notation the number of pixel/value tuples = the number of pixels, thus using equations 1, 2, and 3 equating H_t with p_t , the relationship is

$$p\tau - \sum_{t=1}^{\tau} (p_{t-1} - e_{t-1} + p - e_{t-1}) > \frac{1}{|T|} \sum_{t=1}^{\tau} (K_t + C_t) \quad (6)$$

simplifying to

$$\sum_{t=1}^{\tau-1} (e_t + e_t) > \frac{1}{|T|} \sum_{t=1}^{\tau} (K_t + C_t) \quad (7)$$

This may be unsatisfiable if the number of pixel/value tuples eliminated in each consecutive pair of images is not large enough. This constraint implies that camera motion must be sufficiently slow to permit lots of overlap and further, the thresholds set for the diff and corr functions must be tight in order for the active approach to be more efficient than the passive one for bounded problems. Camera motion must be carefully planned so as to maximize the number of eliminated pixels in the next image and to minimize the number of carried-over pixels in the new image. In other words, large overlap and slow camera motion have a positive effect. This of course must be balanced with the total number of images because the down side of large overlap is more images.

Note that the left-hand side of equation (7) is the same as the right-hand side of equation (3). Combining the two, an overall constraint can be stated as

$$h\tau - H_{\max} \leq \sum_{t=1}^{\tau-1} (e_t + e_t) > \frac{1}{|T|} \sum_{t=1}^{\tau} (K_t + C_t) \quad (8)$$

A greedy algorithm that eliminates hypotheses until the constraint is satisfied will suffice. Correct solution is guaranteed provided the strict ordering of thresholds as outlined in section 4.6 is maintained.

It should be clear that the practical application of the above constraints is far from obvious; it requires a careful evaluation of the costs of specific computations for a given active strategy. Worst-case time-complexity estimates must be made for specific algorithms so that the constraints may be evaluated.

5 Where Reactive Vision?

The above discussion points to the need for intermediate representations (the space of hypotheses) in an active vision paradigm. This may appear to go against the reactive philosophy currently a major foundation of active vision approaches (for example, see Ballard [1989]). Can the two positions be reconciled?

It is easy to place the hypothesize-and-test idea into a reactive framework: assume (as is done in such frameworks) that the set of choices of stimulus-action pairings is given. The behavior specification of the device can easily provide this. The standard hypothesize-and-test paradigm operates as follows:

- propose a particular explanation as the correct one for the current input (explanation may involve perceptions, actions, etc.)
- devise a test in order to verify that it is indeed the correct explanation
- if the hypothesis passes the test, proceed with that explanation and its consequences
- if the hypothesis fails the test, select another explanation and try again. The selection mechanism may be quite complex depending on the size of the hypothesis space.

Suppose now that all hypotheses can be tested in parallel. Connell [1989], for example, defines nodes exactly in the above form. A parallel realization is feasible only for relatively small stimulus-action pair spaces, such as the ones currently implemented in various reactive devices (Brooks [1986], for example).⁹ This would lead to a configuration such as shown in figure 2; this is not unlike the kinds of circuits derived using the subsumption ideas of Brooks. This circuit is reactive—it reacts to stimuli as they enter the system, and it appears to have many behaviors since it has many stimulus-action pairs to choose from and it resolves

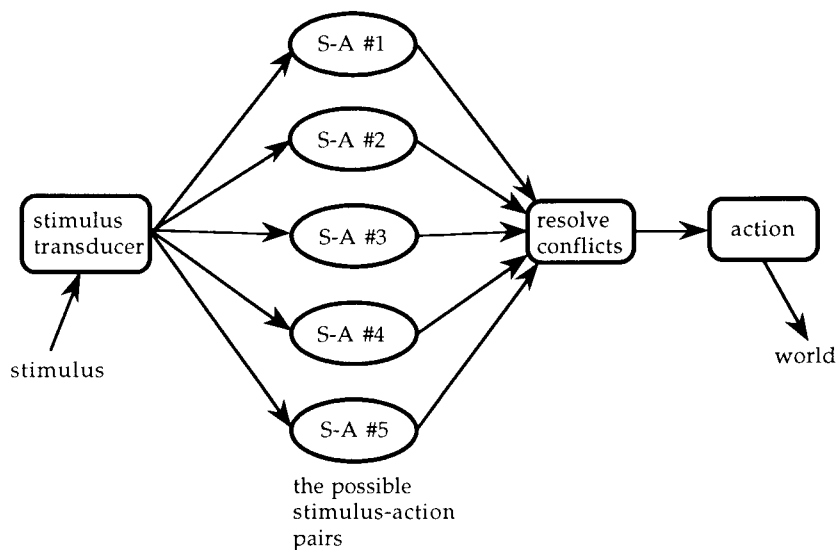


Fig. 2. A "reactive" viewpoint on hypothesize-and-test.

conflicts. But most importantly, it has exactly the same hypothesize-and-test mechanism that has been common in AI. The key difference is the parallel implementation.

Figure 2 is an abstraction of the several control diagrams given in Connell [1989]. The stimulus transducer passes to each stimulus-action test the subset of input data that is relevant for that action. All “hypotheses” are thus represented and tested individually. Actions are executed that move the sensors or do whatever else is appropriate depending on the best hypotheses and the scheme repeats. The maximum number of stimulus-action (S-A) pairs (or hypotheses) may be bounded by H_{\max} , as described earlier. If the hypotheses are filled by the sensing system dynamically such that the set of pixel/value tuples is the *stimulus* and the *action* is a vote for a camera motion that improves the values of the diff and corr functions,¹⁰ then the scheme in figure 2 is isomorphic to that proposed in section 4.6. The “resolve conflicts” stage would look at all of the camera motion votes and decide which to actually execute.

Now, expand the stimulus input to full-size images rather than simple signals. If one strictly adheres to the belief that no intermediate representations are required then the scheme will be limited because: (1) only very primitive sensory (internal as well as external) processing can be permitted (“small images or signals”); (2) only very primitive degrees of spatial and temporal context relations are possible; (3) only a small set of behaviors can be included. Here is the reason why.

The paradigm in figure 2 is strongly related to visual search as defined earlier. Recall that in the visual search task, a subject is presented with a target (or targets) and a test image and asked to determine whether or not that target is present in a test image. This involves 2 behaviors (or stimulus-action pairs): (1) if target present, press button A; (2) if target absent, press button B. The choices for processing the input in a hard-wired reactive framework are

- i. each S-A processes only one fixed subset of input in a uniform manner regardless of task or image characteristics;
- ii. each S-A processes the whole image in a uniform manner regardless of task or image characteristics;
- iii. each S-A processes the appropriate subset of the input anywhere in the image in a manner appropriate to that subset and the task.

Neither i or ii will exhibit “intelligence” or the observed human visual search behavior unless the world is very cooperative and the task is very vague. Choice

iii on the other hand yields a problem with the same structure as the formal visual-search problem. Visual-search definitions are not limited to visual images: the problem and conclusions extend naturally to any type of signal input.

The implications are far reaching: visual search can be viewed within the behaviorist paradigm. More importantly, unless the visual world or visual behavior is trivialized, any behaviorist approach necessarily must solve the visual-search problem as outlined earlier in this paper. Behaviorism with embedded bounded visual-search problems requires no more than linear time for the signal matching tasks. If a target is known, (target or goal is explicitly represented and/or realized by the circuit), the behaviorist paradigm can be very fast even if targets can be rotated and scaled. On the other hand, behaviorism with embedded unbounded visual-search problems may require exponential time for the signal-matching tasks since the unbounded problem is NP-Complete regardless of the implementation (both passive and active problems).

If it is assumed that any vision system must begin with a set of pixel measurements, and that one of its first steps is to create an edge representation (actually any extracted representation will do for this argument—depth, color, etc.), then the obvious question that must be answered is: what are the physical structures responsible for the edges? this is an unbounded visual search problem: there are no constraints on the size, extent, shape, etc. of the structures sought. It is exactly like the problem of interpreting the image of the dalmatian sniffing at leaves mentioned earlier. The model base of possible objects in the world is of no help since it would be very large and varied in general and thus not provide useful constraints. This kind of question is at the heart of visual processing, and thus it follows that unbounded visual-search tasks can be found easily in most vision problems. Thus, if the target is given as a set of constraints or is known only implicitly in some way (the spatial extent or configuration, or specific appearance are not known nor constrained), behaviorism will fail due to the computational load for realistic non-trivial images unless optimizations and approximations of the kind described in Tsotsos [1990a] are also included: intermediate representations, attention, hierarchical organization, spatial abstraction, logically segregated visual maps. But such mechanisms are exactly the kinds of things that behaviorists typically claim are not needed: it appears that this claim is unjustified. If the targets are always specified so that a behaviorist

approach can dispense with intermediate representations, then the space of stimulus-action pairs will necessarily become so large that the scheme will collapse under the weight of the computational demands: each possible world must be explicitly represented.

It should now be apparent that no reconciliation is necessary. Any behaviorist approach to vision or robotics must deal with the inherent computational complexity of the perception problem; otherwise the claim that those approaches scale up to human-like behavior is easily refuted.

Conclusions

Aspects of attention have been linked to the concept of active perception. Active vision has been placed into the same formalism as previous analysis for the purpose of complexity analysis; thus, they are now compatible and complementary mechanisms. The formalization of the active search problem is based on the key principle suggested by Bajcsy in her original paper: "active perception is a problem of intelligent control applied to the data acquisition process that depends on the current state of data interpretation." This leads directly to search within a hypothesize-and-test framework and thus to attentional processing in general. Analysis yielded an efficiency reason for the use of active vision strategies for some specific situations and those situations where active strategy is less efficient than passive schemes were discovered. A constraint on processing time was developed that could be used to determine the efficiency of an active strategy and which may guide selection of camera motions in a real-time system. The conclusion is that vision systems should be able to dynamically decide whether to employ an active or passive strategy based on a number of decision dimensions, one of them being efficiency. It will not always be the case that an active strategy is more efficient than a passive one. The formalism presented for this analysis was linked to hypothesize-and-test strategies in general, static or time-varying images, with or without moving sensors; the same results are applicable directly. Finally, it was argued that this analysis is an extension of the behaviorist paradigm and that it addresses the scaling problem directly.

Acknowledgments

I wish to thank Ruzena Bajcsy for suggesting this problem to me. The author is the Canadian Pacific Fellow

of the Canadian Institute for Advanced Research. This research was conducted with the financial support of the Natural Sciences and Engineering Research Council of Canada and the Information Technology Research Center, a Province of Ontario Center of Excellence.

Notes

1. See the passage translated by and appearing in Nakayama and Mackeben (1981).
2. See Steinman [1986] for a 25-year review of eye movement research; also see Ballard [1989] for a corresponding computational viewpoint.
3. The definitions of bounded and unbounded search are presented in section 4.
4. The experiments typically require that the images be foveated and subjects maintain fixation at a given point. Trials with eye movements are discarded. If either of these conditions are changed, active strategies would have an effect even for 2D images of 2D targets.
5. Recall that algorithmic complexity is defined as the time complexity of a specific algorithm (a step-by-step procedure for solving a problem), while problem complexity is given by a function that is an upper bound on all possible algorithms for a given problem independent of implementation (a problem is a general question with parameters and a statement of what properties a solution must satisfy) [Garey & Johnson 1979].
6. The diff and corr functions need not be specified algebraically; rather they may also be provided as a table whose values are fetched via look-up.
7. It is well known that Knapsack has a pseudo-polynomial solution. However, it is argued in Tsotsos [1990b; 1991] that this is not a biologically plausible solution for vision. Moreover, this does not prove that the problem is not exponential given no further assumptions.
8. There are $2^{|I|}$ test image tuples and their average size is given by

$$\frac{\sum_{k=0}^{|I|} \binom{|I|}{k} \cdot k}{2^{|I|}} = \frac{|I|2^{|I|-1}}{2^{|I|}} = \frac{|I|}{2}$$

and there are two arithmetic operations for each element.

9. It was shown in Tsotsos [1989, 1990a] that parallelism alone cannot sufficiently account for human perceptual abilities. The number of hypotheses possible quickly outstrips the size of the brain or of any computer.
10. It is not obvious how this may be accomplished efficiently at this time. In principle, however, it is clear that it can be accomplished as follows: associate a centroid with each hypothesis, i.e., a location in the image that corresponds to the center of mass of the pixel locations that make up the hypothesis; then try different values of the thresholds to improve the values of the diff and corr functions, noting which pixel/value tuples would be eliminated; compute a new centroid for the proposed modified hypothesis set; vote for a camera motion toward this new location. Verification of this approach is currently under investigation.

References

- Abbott, A., and Ahuja, N. 1988. Surface reconstruction by dynamic integration of focus, camera vergence and stereo. *Proc. 2nd Intern. Conf. Comput. Vision*, December, Tarpon Springs FL, pp. 532–543.
- Aloimonos, J., Weiss, I., and Bandyopadhyay, A. 1987. Active vision. *Proc. 1st Intern. Conf. Comput. Vision*, London, pp. 35–54.
- Bajcsy, R. 1985. Active perception vs. passive perception. *Proc. IEEE Workshop on Computer Vision: Representation and Control*, October, Bellaire MI, pp. 55–62.
- Ballard, D. 1985. Task frames in visuo-motor coordination. *Proc. IEEE Workshop on Computer Vision: Representation and Control*, October, Bellaire MI, pp. 3–10.
- Ballard, D. 1987. Eye movements and visual cognition. *Proc. IEEE Workshop Spatial Reasoning and Multi-sensor Fusion*, St. Charles IL, October, pp., 188–200.
- Ballard, D. 1989. Animate vision. *Proc. 11th Intern. Joint Conf. Artif. Intell.*, Detroit
- Ballard, D., and Ozcanarli, A. 1988. Eye fixation and early vision: Kinetic depth. *Proc. 2nd Intern. Conf. Comput. Vision*, Tarpon Springs, FL, pp. 524–531.
- Bandyopadhyay, A., Chandra, B., and Ballard, D. 1986. Active navigation: Tracking an environmental point considered beneficial. *Proc. IEEE Workshop on Motion: Representation and Analysis*, Charleston, SC, May, pp. 23–28.
- Brooks, R. 1986. A layered intelligent control system for a mobile robot. *IEEE J. Robot. Autom.* RA-2: 14–23, April.
- Clark, J.J., and Ferrier, N. 1988. Modal control of an attentive vision system. *Proc. 2nd Intern. Conf. Comput. Vision*, December, Tarpon Springs, FL, pp. 514–523.
- Connell, J. 1989. "A Colony Architecture for an Artificial Creature." MIT AI Lab, Ph.D. Thesis, AI-TR1151.
- Freuder, E. 1976. "A Computer System for Visual Recognition Using Active Knowledge." Ph.D. thesis; also AI-TR-345, MIT AI Lab, June.
- Garey, M., and Johnson, D. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman: New York.
- Gibson, J.J. 1979. *The Ecological Approach to Visual Perception*. Houghton Mifflin: Boston.
- Grimson, E. 1988. The combinatorics of object recognition in cluttered environments using constrained search. *Proc. 2nd Intern. Conf. Comput. Vision*, Tampa, FL, pp. 218–227.
- Helmholtz, H. von. 1910. *Hanbuch der physiologischen detik*. English translation, J. Southall, Dover: New York, 1925.
- Krotkov, E. 1987. Focusing. *Intern. J. Comput. Vision* 1:223–237.
- LaBerge, D. 1990. Attention. *Psychological Science* 1–3, 156–162.
- Leeper, R. 1935. A study of a neglected portion of the field of learning: The development of sensory organization. *J. Genet. Psychol.* 46:41–75.
- Marr, D. 1982. *Vision*. W.H. Freeman: New York.
- Metzger, W. 1974. Consciousness, Perception and Action. *Handbook of Perception, vol. 1, Historical and Philosophical Roots of Perception*. Academic Press, San Diego, CA, pp. 109–125.
- Nakayama, K., and Mackeben, M. 1989. Sustained and transient components of focal visual attention. *Vision Research* 29(11):1631–1647.
- Parasuraman, R., and Davies, D., (editors). 1984. *Varieties of Attention*. Academic Press: Orlando, FL.
- Paul, R., Bajcsy, R., et al. 1987. The integration of sensing with actuation to form a robust intelligent control system. GRASP LAB 97, Dept. of Computer and Information Science, University of Pennsylvania, March.
- Rabbitt, P. 1978. Sorting, categorization and visual search. In *The Handbook of Perception: Perceptual Processing*, Vol. IX, edited by E. Carterette and M. Friedman. Academic Press: New York.
- Steinman, R. 1986. Eye movement. *Vision Research* 26(9):1389–1400.
- Treisman, A. 1988. Features and objects: The Fourteenth Bartlett Memorial Lecture. *Quart. J. Exp. Psychol.* 40A(2):201/237.
- Tsotsos, J.K. 1988. A 'complexity level' analysis of immediate vision. *Intern. J. Comp. Vision* 1(4):303–320.
- Tsotsos, J.K. 1989. The complexity of perceptual search tasks. *Proc. 11th Intern. Conf. Artif. Intell.*, pp. 1571–1577, Detroit.
- Tsotsos, J.K. 1990a. A complexity level analysis of vision. *Behav. Brain Sci.* 13(3):423–455.
- Tsotsos, J.K. 1990b. A little complexity analysis goes a long way. *Behav. Brain Sci.* 13(3):459–469.
- Tsotsos, J.K. 1991. Is complexity analysis appropriate for analyzing biological systems? *Behav. Brain Sci.*, 14(4):770–773.
- Uhr, L. 1990. Some important constraints on complexity. *Behav. Brain Sci.* 13(3):455–456.
- Verville, E., and Cameron, N. 1946. Age and sex differences in the perception of incomplete pictures in adults. *Pedagogical Seminary* 68:149–157.