The Multiple Dimensions of Action-Oriented Robotic Perception: Fission, Fusion, and Fashion

Ronald C. Arkin
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332

Abstract

Action-oriented perception provides an alternative to traditional high-level image understanding for the roboticist. By channeling sensory perception directly to motor behaviors (sensor fission) without mediating global representations rapid response is ensured. Forcing sensor fusion to be conducted within the context of motor needs reduces computational demand and enhances parallelism. Utilizing the correct visual algorithm to support motor action at the correct time (temporal coordination or sensor fashion) provides robust performance over wide ranges of activity. In this paper we describe the philosophy of action-oriented perception for the roboticist and discuss the different dimensions in which it may be effectively used.

1. Introduction

The traditional approach to computer vision and the processing of other related sensor modalities has more often than not ignored the fact that perceptual needs are predicated upon the motivational and behavioral requirements of the consuming agent. This has lead to a wealth of literature on perceptual processing which serves as little more than academic curiosities: i.e., solutions looking for problems. Although much of this work can serve as a useful pastime for researchers and provide for glitzy demonstrations to impress the uninitiated, the resulting contributions over the past 30 or so years have not yet produced robust intelligent robotic perception.

Certainly scapegoats can be found: computer architectures were too primitive, neuroscientists have not provided an adequate understanding of human vision, etc. It is our contention, however, that the *means* were not at fault, but rather the desired *ends*. The fundamental problem of treating perception as an isolated phenomena is a very serious one and brings into question whether the "general vision" problem is really a problem at all, but rather an artifact of misguided research.

What is needed is a shift towards a new (or rather rediscovered) paradigm: viewing perception as a supportive (and hence subservient) process of action. More accurately, a duality exists: where perception directly supports action we will use the term *action-oriented (or task-driven) perception*; and where action directly supports perception we will use the term *active perception*. Nonetheless, even active perception in principle is a subset of action-oriented perception, as its ultimate goals serve also to provide support for acting within the world. Although important, we will not describe the work in active perception here (the reader is referred to [8] - see also animate vision [7]). In either case, action and perception are inseparable.

We will discuss in this paper the inherent advantages of action-oriented perception as a means for robotic control and provide an analysis of the dimensions in which perception can be utilized to support motor action in the context of time and space. This discussion will center on three related yet different methods: sensory channeling (sensor fission), sensor fusion within the channels, and temporal coordination (sensor fashion). Behavior-based robotics (reactive control) provides a basis for the implementation of this paradigm within intelligent robotic systems.

We first review the work upon which this philosophy is built and the framework within which our ideas are tested. What sensor fission, fusion, and fashion have to offer the roboticist is presented in Section 3 with a discussion of the *appropriate* role of representational knowledge or models. A summary and conclusions completes the paper.

2. Background

2.1 Perception as Communication

Consider that the world is trying to tell us something, if only we knew how to listen. Sensing can thus be viewed as a form of communication - where information flows from the environment to the attending agent. Obviously, if we don't know what to attend to we will have a hard, if not impossible, time discerning the messages that the world is providing. The world is telling us something if only we pay attention. Where and how our attentional and perceptual resources are directed depends strongly upon our motivation or intentional state.

The ethological literature is replete with examples of sensed information providing cues for evoking behavior (e.g., [19,20]). Indeed, evolution has provided biological agents with highly tuned apparatae to efficiently "pick-up" the information necessary to carry out useful actions. Looming and prey detectors [13] for guiding visual response in the frog are good examples.

In recognition behavior, we find that some agents are capable of discerning things others simply cannot (e.g., intra-species recognition among birds [12]). Perceptual cues that are necessary for the survival and routine functioning of an organism are extracted cheaply and efficiently from the environment while irrelevant information is not processed at all (i.e., it's not even discarded - pick-up never occurs). In other words, these agents have evolved mechanisms that enable efficient communication with the salient features of the world (salient in the context of that agent's needs). The implications are that we need to have our robotic agents attend to what is necessary in the context of their (not our) needs. Depending on their internal conditions, motivational state (plans), and sensory limitations we can develop algorithms that provide useful and focussed information for these actors.

2.2 Action-oriented Perception

Action-oriented perception is not a new concept. It has roots in both cybernetics [2] and cognitive psychology [17]. The underlying principle is that perception is predicated upon the needs of action: only the information that is germane for a particular task needs to be extracted from the environment. The world is viewed in different ways based upon the agent's intentions. In short, perception is conducted on a *need to know* basis. (A more detailed discussion of action-oriented perception appears in [3]).

The ramifications of this philosophy should be readily apparent for the designer of a robotic system. In general, there is no point to constructing full scale scene interpretation and three dimensional world reconstruction. Why bother? All that needs to be done is to identify the necessary perceptual cues to support motor action and the job is done. Much of the inherent difficulty in the general vision problem vanishes, as there is no need to perform such a complex and arduous task.

In order to effectively use action-oriented perception we must have some technique that provides a tight coupling between motor control and perception. The relatively new discipline of behavior-based robotics affords us just such a mechanism.

2.3 Behavior-based Robotics

It is not the intent of this paper to present a wide-ranging overview of behavior-based robotic systems (reactive control). Suffice it to say there are many examples of these systems (e.g., [1,4,9,18]). This relatively new approach to robotics provides an effective means for coupling perception to action in the context of motor needs. Some of the characteristics of this approach include:

- Tasks are typically decomposed into a collection of primitive behaviors.
- Global representations are generally avoided.
- Sensor decoupling is preferred over global sensor fusion.
- It is well situated for dynamic changes in the world.
- It has produced effective robotic demonstrations in dynamic worlds [4,10].

The divisions within this school of thought are largely based upon the use of world knowledge, control strategies (concurrent or arbitration), and other related issues. What is important to note, however, is that they all share a context for constraining perceptual input by providing perceptual specifications that are contingent upon specific motor requirements (the behavior itself). This framework is highly appropriate for developing action-oriented perception techniques along the lines developed in Section 3 below.

In our work, we have used motor and perceptual schemas as a means for encapsulating and representing motor actions and perceptual strategies. This approach has been strongly motivated by neuroscientific, cognitive, and ethological studies and is discussed in detail in [4]. For the purposes of this paper, we must recognize that motor schemas provide the specifications for a perceptual process: i.e., what must be discerned from environmental sensing and perhaps constraining where it may be located, often bootstrapped by expectations from either *a priori* models or previous sensing. Focus of attention mechanisms play a role in directing the perceptual process as to *where* to look while expectations provide clues (e.g., models) as to *what* the appearance of the event being sought is. The *how* aspect is captured by the perceptual schema, a perceptual process geared to provide sensory information that is relevant to a particular task, not some arbitrary modeling or interpretation process. The particular configuration of motor and perceptual schemas is configured prior to actual execution by a higher level planner [5].

3. Dimensions of Action-oriented Perception

The ways in which perceptual strategies can be arrayed relative to the motor behaviors constitutes the dimensions of action-oriented perception (Figure 1). The three approaches we are exploring are described below.

3.1 Connecting Relevant Perception to Appropriate Action: Sensor Fission

Rodney Brooks was the first to coin the term *sensor fission*. By this we mean the channeling of sensory information directly to an appropriate motor action without the construction of an intervening representation. Sensory channeling (fission) is in strong agreement with the principles of action-oriented perception, tying together in an immediate way the perception of the environment and a response. This form of sensor management is closely related to the stimulus-response (behaviorist) view of the world. No abstractions (percepts) are required and information is used locally.

In our research we have developed many different types of sensory algorithms that fit into this category. They include:

- Ultrasonic and visual obstacle avoidance

Figure 1: Dimensions of Action-Oriented Perception
a) Sensor fission - Multiple independent motor schemas, each with its own perceptual schema(s).
b) Sensor fusion - Multiple perceptual subschemas support a perceptual schema within the context of a single motor schema.
c) Sensor fashion - Multiple perceptual schemas are sequenced in the context of a single motor schema.

- Visual Road following
- Model-based target recognition
- Motion detection
- Others

The partitioning of the objects in the world into equivalence classes is based upon the needs of a motor action: obstacle or non-obstacle, road or non-road; landmark or anything else; moving object or stationary. The perceptual task directs the correct sensory processing to the consuming motor behavior. By splitting (fissioning) these tasks we obtain the ability to execute them in parallel on separate processors thus enhancing the computational performance.

Additionally there is inherent flexibility by configuring the perceptual tasks to meet the robot's current motor needs. Instead of creating an omniscient universal representation that can support any conceivable robotic requirement, the task is greatly simplified by eliminating any common intervening representations.

We actually soften the seemingly strict avoidance of representational constructs over some other approaches (e.g., [11]). We do not advocate the use of global representations (e.g., those that provide information without regard for motor needs), but do encourage the formation of local percepts defined within the context of a motor behavior. This approach is discussed below for sensor fusion within action-oriented perception and is implemented through the use of perceptual schemas and subschemas.

3.2 Sensor Fusion in Action-oriented Perception

The traditional role of sensor fusion has been to take multiple sources of information, fuse them into a single global representation, and then reason over that representation for the purposes of acting. As described previously, this methodology is inherently inefficient. Of course we cannot ignore the fact that multiple sources of information can significantly enhance the way in which an agent acts within the world. We advocate, however, that sensory reports be fused only within the context of motor action and not into some abstract all-purpose global representation.

In our schema-based methodology, perceptual schemas that are used for fusion purposes yield percepts that are directly related to the needs of a motor behavior. These perceptual schemas are supported by perceptual subschemas which feed their parent the direct sensor information from each source. The parent perceptual schema combines this information using uncertainty management techniques [15] to produce a percept and a measure of its belief to be used within the motor schema itself.

These perceptual schema/subschema arrays are configured prior to execution, based on the needs of the motor action, during the investigatory phase of the fusion process. The performatory phase of sensor fusion is analogous to the execution aspects of reactive control and proceeds without hierarchical supervision. A unique control scheme [14] has been developed to provide error recovery capabilities in light of potential sensor failures or uncertain readings. The approach and results of this work are described in more detail in a companion paper submitted to this conference [16].

3.3 Temporal Coordination of Perceptual Algorithms - Sensor Fashion

It is entirely possible, and in many instances highly desirable, to have more than one perceptual algorithm associated with a single motor behavior at different stages during its activation. We are currently developing a methodology that facilitates the control of these perceptual algorithms over time. In a sense, we are addressing the issue of when it is time appropriate, or fashionable, to use a particular algorithm to support a given behavior, hence the term sensor fashion. The choice of algorithm or algorithm sequence will of course be dependent upon environmental conditions and context (e.g., lighting) and the robot of necessity will have to dress out its sensor algorithms to fit the current occasion. Sensor fission is concerned with multiple independent concurrent algorithms; sensor fusion is concerned with combining multiple related concurrent algorithms, and sensor fashion is concerned with sequencing multiple related perceptual algorithms.

Some issues regarding how these perceptual algorithms can be coordinated over time are evidenced in:

- Long range versus close range perception.
- Feedforward versus feedback perceptual control.
- Model-based recognition versus adaptive tracking.

One example we explore in our research lies in the context of docking with a manufacturing workstation. The motor behavior is of one form, the **docking** motor schema. In our current instantiation, it involves as many as four distinct perceptual schemas:

1. Long-range visual detection using either a light-seeking strategy or a temporal activity detection algorithm.
2. A Hough transform model-based recognition strategy.
3. Adaptive tracking of a passive landmark using fast region segmentation.
4. Final positioning using texture-based vision or ultrasound.

A description of several of these algorithms appears in [6].

The primary issues in temporal coordination are when to use each of these algorithms and how to determine the best time to switch over from one perceptual strategy to the next. We have tied this transition process to distinct aspects of the docking schema: while the robot is far from the dock undergoing ballistic motion control, we use the long-range detection algorithm; when within an expected recognition range, the model-based strategy is used to explicitly recognize the dock within the image, this exteroceptive cue triggering the controlled motion phase of the docking schema. As the range closes, a transition from the adaptive tracking algorithm to the final positioning algorithm occurs. We have completed an integrated demonstration of all of these algorithms operating in tandem to provide docking support even in the presence of a cluttered and dynamic environment.

3.4 What about models?

A priori knowledge of the world can and should play a role during perceptual processing. This view is in contrast to several of those in the reactive robotics community who eschew representational knowledge at all times. We, however, embrace its *appropriate* use. These uses include:

- Providing expectations for perceptual processes (i.e., *what* to look for).
- Providing focus-of-attention mechanisms (i.e., *where* to look for it).
- To initially configure sensor fusion mechanisms (investigatory phase - i.e., *how* to look for it).
- To sequence perceptual algorithms correctly (i.e., *when* to look for it).

These uses, when framed within the context of motor behavior, are entirely proper. It is not so much in what form the representation appears but rather how it is utilized that distinguishes the action-oriented approach from other methods.

4. Summary and Conclusions

The old problems and their attendant solutions have not worked in providing robust intelligent robotic perception. It is our claim that new problems must be formulated to make significant inroads towards a solution. These problems include:

- How can we represent and control sensing if we view it as a form of communication?
- How can we take advantage of expectations, attention, and intention in providing intelligent agents information about the world?
- How can behavior-based robotics effectively serve as a means for integrating this strategy?

We forward action-oriented perception as an approach to the solution of these problems. In various ways we have demonstrated limited success towards achieving these goals, but much more remains to be done. If we step back and review the sensory aspects of robotic intelligence, we may find that a different and non-traditional approach may produce useful results in short-order. We continue our advocacy for studying psychological, neuroscientific, and ethological models of perception as a basis for attaining these results.

REFERENCES

[1] Anderson, T.L. and Donath, M., "Animal Behavior as a Paradigm for Developing Robot Autonomy", *Robotics and Autonomous Systems*, 6 (1990), pp. 145-168.

[2] Arbib, M., *The Metaphorical Brain*, Wiley, 1972.

[3] Arkin, R.C., "The Impact of Cybernetics on the Design of a Mobile Robot System: A Case Study", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 20, No. 6, Nov/Dec 1990, pp. 1245-1257.

[4] Arkin, R.C., "Motor Schema-Based Mobile Robot Navigation", *International Journal of Robotics Research*, Vol. 8, No. 4, August 1989.

[5] Arkin, R.C., "Integrating Behavioral, Perceptual, and World Knowledge in Reactive Navigation", *Robotics and Autonomous Systems*, 6 (1990), pp. 105-122.

[6] Arkin, R.C., Murphy, R.R., Pearson, M. and Vaughn, D., "Mobile Robot Docking Operations in a Manufacturing Environment: Progress in Visual Perceptual Strategies", *Proc. IEEE International Workshop on Intelligent Robots and Systems '89*, Tsukuba, Japan, 1989, pp. 147-154.

[7] Ballard, D., "Animate Vision"", *Artificial Intelligence*, Vol. 48, Feb. 1991, pp. 57-86..

[8] Bajcsy, R., "Active Perception", *Proceedings of the IEEE*, Vol. 76, No. 8, August 1988, pp.966-1005.

[9] Brooks, R.A., "A Robust Layered Control System for a Mobile Robot", *IEEE Jour. of Robotics and Auto.*, Vol. RA-2, No. 1, 1986 pp. 14-23.

[10] Brooks, R.A., "Elephants Don't Play Chess", *Robotics and Autonomous Systems*, 6 (1990), pp. 3-15.

[11] Brooks, R.A., "Intelligence without Representation", *Artificial Intelligence*, Vol. 47, No. 1-3, Jan. 1991, pp. 139-159.

[12] Colgan, P., *Comparative Social Recognition*, J. Wiley, New York, 1983.

[13] Ewert, J-P., *Neuroethology: an introduction to the neurophysiological fundamentals of behavior*, Springer-Verlag, Berlin, 1980.

[14] Murphy, R.R., "A Control Scheme for Sensor Fusion for Navigation of Autonomous Mobile Robots", Proceedings of *SPIE Sensor Fusion III*, Cambridge, MA, Nov. 1990.

[15] Murphy, R.R., "An Application of Dempster-Shafer Theory to a Novel Control Scheme for Sensor Fusion", to be presented at *SPIE Stochastic Methods in Signal Processing, Image Processing, and Computer Vision*, San Diego, CA, July 21-26, 1991.

[16] Murphy, R.R., "Towards a Foundation for Sensor Fusion: Grammar and Control", working notes, AAAI Fall Workshop on Aspects of Sensory Intelligence, 1991.

[17] Neisser, U., *Cognition and Reality: Principles and Implications of Cognitive Psychology*, Freeman, 1976.

[18] Payton, D., "An Architecture for Reflexive Autonomous Vehicle Control", *IEEE Conf. on Robotics and Auto.*, pp. 1838-1845, 1986.

[19] Smith, W.J., *The Behavior of Communicating: An Ethological Approach*, Harvard University Press, Cambridge, MA, 1977.

[20] Tinbergen, N., *Social Behavior in Animals*, Methuen, London, 1953.