CVGIP: IMAGE UNDERSTANDING
Vol. 60, No. 1, July, pp. 65–73, 1994

DIALOGUE

A Computational and Evolutionary Perspective on the Role of Representation in Vision

MICHAEL J. TARR*

Department of Psychology, Yale University, New Haven, Connecticut

AND

MICHAEL J. BLACK

Department of Computer Science, Yale University, New Haven, Connecticut

Received July 6, 1992; revised November 2, 1993

Recently, the assumed goal of computer vision, reconstructing a representation of the scene, has been critcized as unproductive and impractical. Critics have suggested that the reconstructive approach should be supplanted by a new purposive approach that emphasizes functionality and task driven perception at the cost of general vision. In response to these arguments, we claim that the recovery paradigm central to the reconstructive approach is viable, and, moreover, provides a promising framework for understanding and modeling general purpose vision in humans and machines. An examination of the goals of vision from an evolutionary perspective and a case study involving the recovery of optic flow support this hypothesis. In particular, while we acknowledge that there are instances where the purposive approach may be appropriate, these are insufficient for implementing the wide range of visual tasks exhibited by humans (the kind of flexible vision system presumed to be an end-goal of artificial intelligence). Furthermore, there are instances, such as recent work on the estimation of optic flow, where the recovery paradigm may yield useful and robust results. Thus, contrary to certain claims, the purposive approach does not obviate the need for recovery and reconstruction of flexible representations of the world.

1. INTRODUCTION

Young disciplines often experience moments of doubt: "Are we doing the right thing?" or "Is this approach viable?" [1]. Nowhere is this better exemplified than in the study of computer vision [2]. While progress has been made, the goal of general vision, on the order of human visual perception, remains elusive. Recently, this has led to the suggestion that the entire endeavor is flawed, that we should discard the dominant paradigm, and that it should be replaced with a new, more practical alternative.[1] While this position may not qualify as a "paradigm shift" [3], it certainly advocates a substantial change in direction. To justify this radical deviation, proponents of the new, so-called *purposive approach* muster three lines of support: first, that machines fall far short of the visual capabilities of humans; second, that current computer vision systems cannot actually do very much that is useful in the way of visual perception; and, third, that the purposive approach is consistent with the notion that biological organisms have evolved brain machinery composed of independent processes, each devoted to solving a particular visual task [4].

Contrary to these arguments, we take an entirely conservative posture, suggesting that the presently dominant reconstructive approach is viable, and, moreover, that there are well-grounded computational and evolutionary reasons for its current, as well as possible future, successes. In support of these claims, two kinds of evidence are presented: first, a general examination of the goals of vision in both artificial and biological systems and, second, a case study of current trends in the recovery of optic flow that illustrates the continuing viability of the reconstructive approach. This evidence leads us to conclude that the reconstructive approach does provide a framework for understanding both human and machine vision and that, in particular, there are already instances

* Please address all correspondence to Michael J. Tarr, P. O. Box 208205, New Haven, CT 06520-8205, E-mail address: tarr@cs.yale.edu.

[1] This is reminiscent of proposals that the entire field of artificial intelligence should be scrapped, for instance, see [53] or [54], or that cognitive phenomena should be reduced to neural explanations, for instance, see [55] or [56].

1049-9660/94 $6.00

where successes have led to useful and robust vision systems. Moreover, this is not at the exclusion of the purposive approach, but rather suggests a common ground that we believe to be more fertile than either approach alone.

2. UNDERSTANDING VISION

In understanding vision, one must begin with the fact that many problems in visual perception are considered to be "ill-posed" [5] and that in order to find a solution, inference and constraint must be introduced. While these constraints may be phrased in general terms, for instance the "rigidity" constraint of Ullman's [6] structure-from-motion algorithm, it is also possible to narrow the domain, positing *specialized* constraints dependent upon the visual task at hand. This constitutes the crux of the purposive approach. Moreover, there are instances in the natural world where the notion of narrow constraint is obviously applied. In a now classic (but once ridiculed) paper, Lettvin *et al.* [7] demonstrated that a frog's retina contains special purpose hardware sensitive to small moving black spots, commonly understood to be "bug detectors." Clearly, this adaptation provides great advantages to the amphibian so equipped: frogs mostly eat bugs and their effective detection, even at the cost of an occasional false alarm, is presumably a key to frog survival. However, what is good for frogs is not necessarily good for humans or otherwise "intelligent" machines. In particular, it may be impossible to identify the specific tasks appropriate to this level of constraint in such complex systems. Marr [8] raises precisely this point with regard to the level at which knowledge or hypothesis is brought to bear on visual perception. For while such constraints may not be "general but particular and true only of the scene in question," this suggests that "any very general vision system must command a very large number of such hypotheses and be able to find and deploy just the one or two demanded by the particular situation." Moreover, "this prospect casts a whole (new) complexion on the vision problem" [8, p. 271]. We contend that, couched in terms of the purposive approach, the number of functionally independent human visual behaviors, as well as their consequential constraints, is too large a number to represent efficiently. Indeed, we doubt whether human visual behavior, or for that matter the operation of any general purpose vision system, can be understood within such a narrow context.

Even when one considers higher primates such as vervet monkeys, there is little evidence that their mental representations are generally as abstract and as flexible as our own [9]. While it is true that "many animals are specialists, performing skills with much greater sophistication in some contexts than in others" [9, p. 310], this is not the hallmark of human cognition. For example, many species of birds acquire knowledge of bird song through domain specific "tunable blueprints," an innate special-purpose acquisition mechanism, while human children seem to acquire domain specific knowledge through the operation of general acquisition mechanisms that are rooted in flexible representational structures [10]. Without such representations, knowledge will remain compartmentalized and inaccessible, leaving the mental system without the capacity to extend knowledge from one context to another [9]. In particular, it is this ability that distinguishes human information processing from that of other species.

3. RELIGIOUS RECONSTRUCTIONISM[2] AND FANATICAL PURPOSIVISM

It is flexibility that distinguishes this reconstructive approach from the purposive approach. Reconstruction, or the recovery paradigm, focuses on deriving functional descriptions of the visible world including its geometric properties and the physical properties of the visible surfaces. The goal then is to build a symbolic, possibly nonspatial, description of the scene. Once derived, symbolic descriptions may be used in a variety of "cognitive" operations, such as visual reasoning, planning, or propositional thought. Stated succinctly, "the goal of a perception system, whether biological or machine, is to create a model of the real world and to use this model for interacting with the real world" [2, p. 116].

To a large degree, what the reconstructive paradigm has meant, from the artificial intelligence perspective, is that perception can be effectively ignored; that vision is a self-contained problem which will produce symbolic input to AI programs [11]. Hence, expert systems, production systems, and other traditional AI domains have assumed an entirely input-driven perspective on visual reconstruction—one that disdains any interaction between a perceiver's knowledge base and perceptual mechanisms. We contend that this viewpoint is unfortunate, first because both vision algorithms and AI programs may benefit from some degree of interaction, and second because it places an unnecessary burden upon the reconstructive approach. For there is nothing inherent in the goal of reconstructing the scene that suggests that "higher level" task-specific and contextual knowledge should not be utilized in recovering information from the environment. Indeed, there are numerous examples from human psychophysics where so-called "top-down" processing influences our perception. For instance, it has been demonstrated that humans recover shape from shading through the interaction of sensory information with high-level knowledge, for example, by assuming a single overhead light source [12]. Similar approaches, many specifi-

[2] We thank Jitendra Malik for suggesting this wonderful phrase.

cally motivated by biological models, may also be found in computer vision, where there has been a recent trend toward "active vision" [13, 14]—a computational approach in which vision and higher level control processes are closely coupled.

Interestingly, because both the active and the purposive approaches emphasize the use of task-specific constraints, the two have often been associated. However, the underlying theoretical axioms of these two approaches are quite dissimilar. Specifically, the goal of the active approach is to reformulate traditionally static vision systems in terms of the dynamic exploration of the environment (e.g., by using a mobile camera and/or multiframe algorithms). Not only does the introduction of an active perceiver facilitate the application of previously acquired information to relevant ensuing contexts, but some seemingly intractable vision problems may be solved by the imposition of such constraints and by the availability of richer visual input [13]. Thus, active vision is an appealing and promising technique for developing robust vision algorithms (as well as understanding biological vision [15]). But is should be emphasized that these advantages *do not* dictate a necessary shift in the end-goals of the perceptual system. Rather, active vision algorithms are fundamentally independent of whether the computational objective is scene reconstruction (with ensuing task performance) or simply task completion. Therefore, reformulating current algorithms in terms of active vision does not entail the reduction of vision to a set of task-specific problems.

In contrast, the goal of the purposive approach is to build systems that will accomplish particular domain-specific tasks, the output of which is successful task completion. The study of vision in general is reduced to the study of the "tasks that organisms possessing vision can accomplish" [4, p. 349]; independent of such tasks, the study of the general problem of vision is not even thought to be possible. To present a concrete example, the standard goal of model-based object recognition is supplanted by a framework in which objects are viewed in terms of their roles, functions, or purposes. It is these properties, not the object's geometry, that serve as the basis for its visual recognition [16]. Specifically, "a chair is an object on which a person can sit. . . . To recognize a chair, we should check for the presence of the functional primitive (the surface patch) just defined" [16, p. 124] (this example is reminiscent of Gibson's [15] idea of objects "affording" their functions). But these arguments belie the nature of complex visual information processing and the stated goals of computer vision/artificial intelligence.[3] Moreover, there are no *a priori* reasons for supposing that general purpose vision is impossible: unquestionably, the evolution of vision in humans offers an existence proof for the development of precisely the kind of flexible, perceptual system to which the discipline of artificial intelligence often aspires.

Of course, in specifying end goals that entail the reconstruction of the scene one could argue that computer vision is barking up the wrong tree altogether.[4] Some proponents of the purposive approach have done just that, asserting that the goal of computer vision should not be to build systems that mimic human vision or to serve as general purpose perceptual systems, but to provide answers to the question, "What am I going to use this visual ability for?" [4, p. 348]. Yet this conception is at odds with one of the major tenets of the purposive approach, that machine perception is not up to snuff with human visual capabilities. For while there is no denying that this is the current state of affairs, this comparison leads to a different research agenda than does the notion that we should give up trying to build "intelligent" vision systems and instead concentrate on simpler domain specific problems.[5] The commonly understood goal of the reconstructive approach is both explaining and implementing complex visual behaviors [17]; changing the goal does not solve this problem, it simply avoids it. Taken together, these arguments suggest that the purposive approach does not really provide an alternative explanation to the reconstructive approach, but rather simply offers an alternative goal for computer vision—one that cannot hope to explain or accomplish many of the commonly held objectives of artificial intelligence, cognitive psychology, or neuroscience.

4. EVOLUTIONARY PERSPECTIVES

It would be iniquitous to suggest that advocates of the purposive approach completely ignore evolutionary considerations. Indeed, Aloimonos [4] suggests that the purposive approach is "consistent with evolution" (p. 348) in that individual visual abilities, such as avoiding danger, locating food, and recognizing kin, would seem to have been selected for independently of each other. Brooks [11] makes a similar argument in favor of simple *physically grounded* systems, speculating that the relatively recent arrival of *Homo sapiens* on the evolutionary scene indicates that complex problem solving behavior, language, and other uniquely human traits are essentially outgrowths of more fundamental sensing and reactive abili-

[3] Any attempt to understand recognition at this level is also plagued by the fact that even seemingly well-defined concepts such as "even number" appear to be neither definitionally nor prototypically represented [57]. Thus, for more complex concepts, such as "chair" or "fruit," it may be difficult, if not impossible, to operationalize their core functions or purposes.

[4] And concurrently much of human psychophysics; indeed, Gibson [15] would most likely concur with this argument.

[5] Of course, this ignores the less extreme viewpoint that "intelligent" behavior may be understood as an emergent property of simpler processes (e.g., [17, 51]).

ties common to many species. These points are reiterated by Cheney and Seyfarth [9] when they state that "natural selection, it appears, has acted not on general skills but on behavior in more narrowly defined ecological domains" (p. 310). In this instance, we concur and set forth the hypothesis that the characterization of natural selection as a "tinkerer" (e.g., [18]) provides strong reasons to believe in some version of information-processing modularity [19] in the evolution of complex systems. However, this conception of independent mechanisms should not be confused with the purposive agenda of decomposing visual problems into continually simpler tasks. Bequeathing modularity upon a particular subsystem in no way entails that it is in any way purposive, but rather that it may be generally characterized as modality specific, innately specified, hard-wired, autonomous, and not assembled [19]. Note that all of these properties are orthogonal to the information-processing goal of the module. Computational objectives need to be specified independently and may take almost any form, including the recovery of scene attributes or the purposeful execution of a specific visual task.

Recent studies on the evolution of complex information-processing mechanisms in humans underscore this point [20]. For instance, Pinker and Bloom [21] have argued that natural language is the result of traditional Darwinian selective pressures. In particular, they suggest that human language satisfies two important criteria for when a trait should be attributed to natural selection: first, complex design for some function (e.g., the computational objective), in this instance "the communication of propositional structures over a serial channel" [21, p. 712]; and, second, the absence of alternative explanations for such complexity. Similarly, we surmise that when human vision is judged by the same two criteria it too should be considered the product of selective evolutionary pressures.

First, it seems incontrovertible that the human visual system exhibits complex design. But for what functions? It is here that we believe the traditional goals of the reconstructive approach come into play, setting forth two clear objectives: the reconstruction of the scene and the recognition of objects within the scene.[6] There are numerous lines of admittedly introspective evidence that human vision is adapted for fulfilling precisely these functions: for example, the perception of object properties for recognition has implications for kin recognition, social interaction, visual communication, predator avoidance, tool making, and food identification; likewise, the reconstruction of a symbolic representation of the visual scene has implications for danger avoidance, navigation, food location, tool use, and visual reasoning.

Second, we are dubious as to whether the purposive approach provides an alternative explanation for the evolution of such complex behaviors. Essentially, the purposive approach offers a "divide and conquer" explanation in which "the machinery of the brain devoted to vision consists of various independent processes that are devoted to the solution of specific visual tasks" [4, p. 348].[7] Although these abilities may be based on common principles, they are hypothesized to have evolved at separate times and within different neural hardware. While this independence reduces the complexity of the behaviors that an organism must acquire (or develop) during its lifetime, it *increases* the complexity of evolving such behaviors—necessitating the repeated derivation of common design principles for many related adaptive tasks. In contrast, the reconstructive approach assumes that each independent mechanism contributes to a common representation that suffices for general purpose vision. Consequently, a solution to a particular computational problem need only arise once in order to be employed across a wide range of visual processes. Of course, reconstruction mechanisms may have likewise originated in particular task-specific abilities (indeed, they almost certainly did), but as *generally adaptive visual principles* they were coopted for many purposes.[8]

Note that human visual cognition clearly displays some of the attributes that one would expect to find in a flexible, general purpose vision system. For instance, it is well documented that humans use mental imagery—a sophisticated subsystem for performing visual reasoning via symbol manipulation over spatial representations [22]. The use of mental imagery has been implicated in a variety of problem solving domains; not only is it useful for solving the piano mover's problem, but there is evidence that it is used in scientific reasoning, creative discovery, and navigation [23]. Moreover, the flexibility of these processes facilitate their extension to many other visual domains: for instance, there is behavioral evidence that the mental imagery mechanism referred to as "mental rotation" is also sometimes used in object recognition [24]. Anthropologists have also speculated that an "increased ability to think in—and communicate by means of—spe-

[6] Contrary to the view espoused by several prominent theories of object recognition, e.g., [58], recognition does not entail reconstruction, see [59]; the converse is also true. Aloimonos [4] points out that reconstruction does not entail recognition, and moreover, that the tasks accomplished by each may be considered independently.

[7] This element of the purposive approach is quite similar to Brooks' "subsumption architecture" [11], e.g., a collection of *independent behaviors* that connect "perception to action."

[8] It may be argued that certain procedures are optimal and therefore will be likely to reoccur. However, this appears to be true only at the most general level, for instance in the homology between human and octopus eyes or between bird and bat wings. The specific underpinnings of these homologous structures are not based upon the same evolutionary history and therefore vary along many dimensions.

cific visual images" and "an emerging consideration of design possibilities by way of two- and three-dimensional images" may have helped to spur the rapid development of new tools and weapons in human evolution. Furthermore, while the actual initiation of image-based representations may not be attributable to "the crossing of a neurological threshold," there is no doubt that certain types of neural hardware (and coincident information-processing capabilities) are a prerequisite [25, pp. 98–99]. Thus, without the presence of such flexible visual representations, one of the most distinctive signatures of human behavior, tool use, might have been impossible.

5. COMPUTATIONAL CONSIDERATIONS

Where then does the necessary inference and constraint arise if not from the purposive approach? In fact, the seeds for solving this problem may be found in Marr's [8] formulation of the problem of vision. This perspective has been reiterated by Marr's collaborator Whitman Richards who states, "The success of the perceptual act is intimately coupled with the observer's ability to build internal representations whose assumptions reflect the proper structure and regularities present in the world. . . . Fundamental to perception is thus the notion that there is indeed structure in the world" [26, p. 11]. The reconstructive approach is no more mired in the ill-posed nature of visual perception than is the purposive approach! Constraints are introduced, but at the level of the physical world rather than at the level of a specific task.

Indeed, some of the assumptions of the purposive approach appear to be restatements of constraints already found in recovery algorithms. For instance, structure-from-motion algorithms have often introduced the concept of multiple frames [27], a constraint that is often construed in a manner akin to the idea of "active" vision [13]. In particular, using the types of assumptions found in active vision, structure-from-motion algorithms are more robust. Exactly this point has often been raised in support of the purposive approach, for example, stating that "one can get more constraints on the motion parameters using many frames" [4, p. 351]. Again we wish to emphasize that this application of active vision in no way entails purposiveness. In fact, many current instantiations of the reconstruction paradigm implicitly make use of active vision and may benefit further by making this explicit. However, this is not the same level at which "purposivists" have sometimes proposed constraints be applied. For positing a multiple or even a many view constraint is entirely consistent with the approach as espoused by even the most ardent neo-Marrian reconstructionists.

6. DIRECTIONS IN RECONSTRUCTION: A CASE STUDY

We now turn to some of the specific criticisms leveled against the reconstruction paradigm. In order to do this, we have selected the problem of recovering optical flow, a frequently cited example of the apparent failures of reconstruction [4]. In particular, in a brief case study, we attempt to address two of the most widespread concerns that have been raised about current recovery algorithms: (1) that they are not robust in the presence of noise and (2) that they are irremediably inefficient. More generally, both of these criticisms are rooted in the belief that the reconstruction paradigm has failed to take into account the computational demands of a real world perceiver. Indeed, given the current state of the art, these charges may be somewhat valid. However, in light of promising advances in the field, we maintain that this does not render the framework itself beyond repair, but rather suggests new directions for future research.

We begin by considering the specific problem of recovering a dense optical flow field. In addition to being the subject of significant scrutiny by critics of the reconstruction paradigm, it provides an appropriate topic for study because it traditionally has been a popular problem in computer vision and is considered to be an important mechanism in biological vision (e. g., [15]). From a computational standpoint, results to date indicate that the estimation of optical flow is too inefficient for robotic applications and too unreliable to be useful for problems such as structure-from-motion [4]. Yet based on human psychophysical results, the recovery of optical flow is a generally solvable problem. Therefore, robust formulations must exist.

As mentioned, one general problem has been that algorithms for optical flow have not taken into account the real-time demands of an active perceiver, for instance, a mobile robot. Indeed, this particular criticism has been raised frequently, specifically suggesting that because the computation of optical flow is ill-posed and requires regularization, algorithms for computing it are inherently iterative and ill-suited to real-time applications. The purposive paradigm's alternative to using optical flow is to find representations that are easier to compute, for example, normal flow or qualitative descriptions of the flow field.

Do the previous inadequacies of optical flow algorithms imply that the endeavor should be abandoned? We think not. First, motion provides important structural information about the world, and it should be exploited—as evidenced by examples from biological organisms. Second, recent work using robust and dynamic algorithms addresses the main criticisms lodged by advocates of the purposive paradigm. Therefore, we believe it is too soon to dismiss optic flow as unusable. In fact, there are indications that we are entering an exciting period in which robust approaches are being developed and the issues of incremental processing in a dynamic environment are being taken seriously. In the sections that follow, we consider how these new trends in optical flow research answer some of the criticisms of the purposivists.

6.1. *Robust Optical Flow*

First, we address the criticism that many current approaches to estimating optical flow are not robust. Typically, optical flow algorithms embody a set of idealized assumptions about the scene, camera motion, and image noise that, in practice, are frequently violated. Most current algorithms are non-robust in that they are sensitive to such violations and, moreover, their performance does not degrade gracefully. However, there are a number of promising techniques for increasing robustness, including the use of robust statistics, better models of optical flow, long image sequences, and analysis of the reliability of flow estimates.

One approach for dealing with the problems of robustness involves the use of robust statistical techniques [28] to deal with motion discontinuities and noise. For example, the work of Schunck [29] uses a robust clustering of constraint lines to determine optical flow. Black [30] has used robust techniques to develop a framework for robust estimation of optical flow and has shown significant improvements over standard least-squares formulations. Correlation-based approaches can also benefit from robust techniques which reduce the effects of outliers at motion discontinuities [31].

One of the most commonly violated assumptions effecting the robust recovery of optical flow is that of spatial smoothness. Traditional regularization schemes for recovering smooth flow fields have the detrimental effect of over-smoothing at motion boundaries. If the recovery of structure is our goal, then these boundaries are likely to be important. There are now many approaches that attempt to solve this problem. The most notable are the Markov random field (MRF) approaches [31–33] in which motion discontinuities are represented either explicitly using "line processes" [34] or implicitly using *weak continuity constraints* [35].

Recently, there has also been interest in achieving more robust and accurate motion estimates by adopting more sophisticated parametric models of image motion within a region. In particular, affine models of optical flow have been shown to be good local approximations to image motion in many common situations [36]. Such models constrain the local motion estimate more strongly than traditional regularization (non-parametric) schemes and allow the estimation of motion in regions with only sparse image structure.

Finally, in cases where the flow estimates are poor or uncertain, it is useful to have an estimate of the flow vector's certainty through the use of confidence measures [37], probabilistic estimates [38], or estimation theoretic techniques [39]. This work is an important contribution, for it may allow processes that use optic flow to ignore poor measurements and hence produce accurate results.

6.2. *Temporal Persistence*

The second main criticism of optical flow algorithms is that their computational expense prevents them from being used under real world conditions. Many previous approaches have only considered the two-frame estimation problem and those that have considered longer sequences typically have done so in a batch fashion [40, 41]. There have been recent advances on this front; in particular, there are now a number of incremental approaches that compute optic flow dynamically and refine the flow estimates incrementally over an image sequence.

For instance, Singh [39] uses a Kalman filter base approach [42] to estimate optical flow incrementally. There are a number of other analogous approaches for estimating depth from motion [43, 44]. While there are problems with the Kalman filter approach, it brings us closer to the objective of dynamic optical flow. An alternative *incremental minimization* approach [31] uses a robust formulation and solves the difficult minimization problem incrementally over a sequence of images. This approach is unique in that it explicitly incorporates a *temporal persistence* constraint in the formulation of the optical flow equation. Temporal persistence provides a powerful additional constraint, *at the level of the physical world,* on the interpretation of visual motion and results in increased robustness.

In a different direction, various researchers have been looking at the real-time computation of optical flow. Notable is the work on implementing optical flow equations using analog devices [45] which incorporates the idea of weak continuity for preserving motion discontinuities. Others have built hardware to perform real-time correlation for optical flow, depth map generation, and tracking [46, 47].

Computational models of optic flow and, in particular, the inclusion of the temporal persistence constraint, also have important implications for understanding biological vision. Tarr and Black [48] have demonstrated that temporal persistence produces systematic distortions in motion recovery very similar to the patterns of distortion observed in human behavioral studies of memories for moving objects [49]. While such results do not conclusively demonstrate that similar algorithms are being used in the human visual system, they do indicate that the level of constraint and general approach is appropriate for developing successful algorithms.

6.3. *Motion and Action*

If the above trends do, in fact, lead to robust and efficient algorithms for optical flow, one can still ask whether this is a reasonable goal. One of the most damning criticisms of the recovery paradigm is that it has often failed to ask this question and, as a result, has developed in a

partial vacuum, isolated from both high level problems in artificial intelligence and the requirements of a dynamic robot. The purposive approach has pointed out this omission and shown the importance of considering the relevance of representations to the tasks a robot must actually perform. As we have noted, the black box view of vision is in no way fundamental to the recovery paradigm, but merely reflects the simplifying choices necessary to attack a difficult problem.

With the purposive approach, one begins with a narrowly defined task and determines what information is both necessary and easily computable to achieve the task. This is a traditional engineering approach to robotics. In contrast, consider the approach of Nishihara [47] that on the surface appears very similar; as with the purposive approach he attempts to solve simple tasks using robust and efficient methods. Nishihara, however, begins with a biologically motivated representation based on the sign of Laplacian of Gaussian filtered images. He then defines a correlation operation that is a simple, general computational mechanism for exploiting the representation. These simple tools are general enough to support a wide variety of tasks, including the computation of optical flow, vergence, tracking, near/far discrimination, and stereo depth recovery. Not only can many of these tasks be implemented in real time, but the approach is appealing in that some of the results are compatible with human psychophysical data.

In a similar vein, Woodfill and Zabih [50] use an optical flow field, computed in real-time, for the tracking of non-rigid objects by an active camera. As hardware for the real-time computation of optical flow becomes more commonplace, we expect that we will see more applications of such representations to problems currently considered to be in the domain of the purposive approach, e.g., tracking, collision avoidance, and heading estimation, as well as to numerous other problems that are not so easily captured by a purposive analysis, e.g., structure-from-motion and motion-base segmentation.

7. FINAL THOUGHTS

We conclude by reiterating that we are not excluding a role for the purposive approach in the study of vision, but that we believe it is better suited for understanding and mimicking the overall visual behavior of frogs rather than humans. This is quite similar to the approach recently taken by Brooks [11, 17, 51] in his development of simple mobile robots that mimic insect behavior. There are also some aspects of human visual behavior, particularly those associated with "automatic" processing, that may warrant a purposive analysis. For instance, a purposive analysis of navigation or wayfinding problems may yield robust algorithms which rely only on qualitative or partial information about optical flow (for example, normal flow [52]). Likewise, qualitative information about surfaces may suffice for grasping an object. Indeed, to date much of the actual research done within the purposive framework has focused on similar problems [4]. Furthermore, some elements of these novel approaches, in particular the concept of an active observer, seem to hold great promise for the study of vision at all levels. Not only do such active techniques provide new approaches to solving many difficult and ill-posed problems in computer vision but they offer a new path for exploring the relationship between computer and biological vision systems.

However, it is also our position that it is crucial that the qualitative information provided by "purposive" modules be general enough that this same information may be utilized in the reconstruction of the scene.[9] Indeed, human psychophysical studies indicate that this routinely occurs. Therefore, if the purposive approach does have a role in understanding general purpose vision, it seems likely to be at the level of well-defined and narrowly constrained tasks, but *without* obviating the need for recovery and reconstruction. Moreover, the sometimes unstated goal of much of computer vision, developing complex visual processing systems capable of producing symbolic descriptions that interact with more traditional AI systems, is alive and well. This is true not only because of the present day successes of the reconstructive approach in computer vision (some of which we have discussed here), but because we believe such an approach holds out the best hope for ultimately understanding and duplicating the adaptive nature of human vision.

ACKNOWLEDGMENTS

The first author was supported by a grant from the Air Force Office of Scientific Research, Contract F49620-91-J-0169. The second author was supported by a grant from the National Aeronautics and Space Administration (NASA Training Grant NGT-50749). We thank P. Anandan, Chris Brown, Gareth Funka-Lea, Greg Hager, David Heeger, and David Kriegman for their helpful comments and advice.

REFERENCES

1. M. R. Banaji, The physical and mental bases of human thought and the impending death of dualism, *IEEE Expert* **6**, 1991.
2. R. C. Jain and T. O. Binford, Ignorance, myopia, and naivete in computer vision systems, *CVGIP: Image Understanding*, **53**(1), 1991, 112–117.
3. T. S. Kuhn, *The Structure of Scientific Revolutions, 2nd ed.*, Univ. of Chicago Press, Chicago, IL, 1970.
4. J. Aloimonos, Purposive and qualitative active vision. In *Proceed-*

[9] We retain some skepticism as to whether object recognition may be construed as purposive under any circumstances, since, unlike frogs, human recognition performance is not tied to any salient environmental conditions.

ings, 10th International Conference on Pattern Recognition, Atlantic City, NJ, June 1990, Vol. 1, pp. 346–360.

5. J. Marroquin, S. Mitter, and T. Poggio, Probabilistic solution of ill-posed problems in computational vision, *J. Amer. Statist. Assoc.* **82**(397), Mar. 1987, 76–89.
6. S. Ullman, *The Interpretation of Visual Motion,* The MIT Press, Cambridge, MA, 1979.
7. J. Y. Lettvin, H. R. Maturana, W. S. McCulloch, and W. H. Pitts, What the frog's eye tells the frog's brain, *Proc. Inst. Radio Eng.* **47,** 1959, 1940–1951.
8. D. Marr, *Vision,* Freeman, New York, 1982.
9. D. L. Cheney and R. M. Seyfarth, *How Monkeys See the World,* Univ. of Chicago Press, Chicago, IL, 1990.
10. S. Carey, *Conceptual Change in Childhood,* The MIT Press, Cambridge, MA, 1985.
11. R. A. Brooks, Elephants don't play chess, in *Robotics and Autonomous Systems,* Vol. 6, pp. 3–15, Elsevier, Amsterdam, 1990.
12. V. S. Ramachandran, Perceiving shape from shading, in *The Perceptual World,* (I. Rock, Ed.) pp. 127–138, Freeman, San Francisco, CA, 1990.
13. J. Aloimonos, I. Weiss, and A. Bandyopadhyay, Active vision, in *Proceedings, First International Conference on Computer Vision, London, England, June 1987,* pp. 35–54, IEEE, Los Alamitos, CA.
14. R. Bajcsy, Active perception, *Proc. IEEE* **76,** 1988, 996–1005.
15. J. J. Gibson, *The Ecological Approach to Visual Perception,* Houghton–Mifflin, Boston, 1979.
16. Y. Aloimonos and A. Rosenfeld, A response to "ignorance, myopia, and naivete in computer vision systems" by R. C. Jain and T. O. Binford, *CVGIP: Image Understanding* **53**(1), 1991, 120–124.
17. R. A. Brooks, Intelligence without representation, *Artif. Intell.* **47,** 1991, 139–159.
18. R. Dawkins, *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe Without Design,* Norton, New York, 1986.
19. J. A. Fodor, *Modularity of Mind,* The MIT Press, Cambridge, MA, 1983.
20. L. Cosmides and J. Tooby, From evolution to behavior: Evolutionary psychology as the missing link, in *The Latest on the Best: Essays on Evolution and Optimality,* (J. Dupre, Ed.) pp. 277–306, The MIT Press, Cambridge, MA, 1987.
21. S. Pinker and P. Bloom, Natural language and natural selection, *Behavioral Brain Sci.* **13**(4), 1990, 707–727.
22. S. M. Kosslyn, *Image and Mind,* Harvard Univ. Press, Cambridge, MA, 1980.
23. R. A. Finke, *Principles of Mental Imagery,* The MIT Press, Cambridge, MA, 1989.
24. M. J. Tarr and S. Pinker, Mental rotation and orientation-dependence in shape recognition, *Cognitive Psychology* **21**(2), 1989, 233–282.
25. R. White, Visual thinking in the ice age, *Scientific American* **261**(1), 1989, 92–99.
26. W. Richards, *Natural Computation,* The MIT Press, Cambridge, MA, 1988.
27. M. L. Braunstein, D. D. Hoffman, L. R. Shapiro, G. J. Andersen, and B. M. Bennett, Minimum points and views for the recovery of three-dimensional structure, *J. Experiment. Psych. Human Perception and Perform.* **13**(3), 1987, 335–343.
28. F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions,* Wiley, New York, 1986.
29. B. G. Schunck, Image flow segmentation and estimation by constraint line clustering, *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(10), Oct. 1989, 1010–1027.
30. M. J. Black, *A Robust Gradient Method for Determining Optical Flow,* Tech. Rep. YALEU/DCS/RR-891, Yale University, Feb. 1992.
31. M. J. Black and P. Anandan, Robust dynamic motion estimation over time, in *Proceedings, Computer Vision and Pattern Recognition, CVPR-91, Maui, Hawaii, June 1991,* pp. 296–302.
32. J. Konrad and E. Dubois, Miltigrid bayesian estimation of image motion fields using stochastic relaxation, *in International Conference on Computer Vision, 1988,* pp. 354–362.
33. D. W. Murray and B. F. Buxton, Scene segmentation from visual motion using global optimization, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-9**(2), Mar. 1987, 220–228.
34. S. Geman and D. Geman, Stochastic relaxation, gibbs distributions and bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-6**(6), Nov. 1984, 721–741.
35. A. Blake and A. Zisserman, *Visual Reconstruction,* The MIT Press, Cambridge, MA, 1987.
36. J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, Hierarchical model-based motion estimation, in *Proceedings, Second European Conference on May 1992* (G. Sandini, Ed.), Lecture Notes in Computer Science, Vol. 588, pp. 237–252, Springer-Verlag, New York/Berlin.
37. P. Anandan, Computing dense displacement fields with confidence measures in scenes containing occlusion, *SPIE Intell. Robots Comput. Vision* **521,** 1984, 184–194.
38. E. P. Simoncelli, E. H. Adelson, and D. J. Heeger. Probability distributions of optical flow. In *Proceedings Computer Vision and Pattern Recognition, CVPR-91, Maui, Hawaii, June 1991,* pp. 310–315.
39. A. Singh, Incremental estimation of image flow using a kalman filter, *J. Visual Commun. Image Represent.* **3**(1), Mar. 1992, 39–57.
40. R. C. Bolles, H. H. Baker, and D. H. Marimont, Epipolar-plane image analysis: An approach to determining structure from motion, *Internat. J. Comput. Vision* **1**(1), 1987, 7–57.
41. D. J. Heeger, Model for the extraction of image flow, *J. Opt. Soc. Amer.* **4**(8), Aug. 1987, 1455–1471.
42. A. Gelb (Ed.), *Applied Optimal Estimation,* The MIT Press, Cambridge, MA, 1974.
43. J. Heel, Temporal surface reconstruction, in *Proceedings, IEEE Computer Vision and Pattern Recognition, CVPR-91, Maui, Hawaii, June 1991,* pp. 607–612.
44. L. Matthies, R. Szeliski, and T. Kanade, Kalman filter-based algorithms for estimating depth from image sequences, *Int. J. Comput. Vision* **3**(3), Sep. 1989, 209–236.
45. C. Koch, J. Luo, and C. Mead, Computing motion using analog and binary resistive networks, *IEEE Computer,* Mar. 1988, 52–63.
46. H. Inoue, T. Tachikawa, and M. Inaba, Robot vision system with a correlation chip for real-time tracking, optical flow and depth map generation, in *Proc. IEEE International Conference on Robotics and Automation, May 1992,* Vol. 2, pp. 1621–1626.
47. H. K. Nishihara, Practical real-time imaging stereo matcher, *Opt. Engrg.* **23**(5), 1984, 536–545.
48. M. J. Tarr and M. J. Black, Psychophysical implications of temporal persistence in early vision: A computational account of representational momentum, in *Investigative Opthamology and Visual Science Supplement,* Vol. 33, p. 1050, May 1992.
49. J. J. Freyd, Dynamic mental representations, *Psych. Rev.* **94**(4), 1987, 427–438.
50. J. Woodfill and R. Zabih, An algorithm for real-time tracking of

non-rigid objects, in *Proceedings, National Conference on Artificial Intelligence (AAAI-91), July 1991.*

51. R. A. Brooks, *Intelligence Without Reason,* Tech. Rep. Memo 1293, MIT, Cambridge, MA, 1991.
52. Y. Aloimonos and Z. Duriç, Active egomotion estimation: A qualitative approach, in *Proceedings, Second European Conference on Computer Vision, ECCV-92, May 1992,* Lecture Notes in Computer Science, (G. Sandini, Ed.), Vol. 588, pp. 497-510, Springer-Verlag, New York/Berlin.
53. R. Penrose, *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics,* Oxford Univ. Press, New York, 1989.
54. J. R. Searle, Minds, brains, and programs, *Behav. and Brain Sci.* **3,** 1980, 417–457.
55. W. J. Freeman and C. A. Skarada, Representations: Who needs them?, in *Brain Organization and Memory: Cells, Systems, and Circuits* (J. L. McGaugh, N. M. Weinberger, and G. Lynch, Eds.), pp. 375–380, Oxford Univ. Press, New York, 1990.
56. J. R. Searle, Consciousness, explanatory inversion, and cognitive science. *Behav. and Brain Sci.* **13**(4), 1990, 585–596.
57. S. L. Armstrong, L. R. Gleitman, and H. Gleitman, What some concepts might not be, *Cognition* **12,** 1983, 263–308.
58. I. Biederman, Recognition-by-components: A theory of human image understanding, *Psych. Rev.* **94,** 1987, 115–147.
59. M. J. Tarr, *Orientation Dependence in Three-Dimensional Object Recognition,* Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 1989.