Knowledge-Based Vision and Simple Visual Machines

Dave Cliff and Jason Noble
School of Cognitive and Computing Sciences, University of Sussex, Brighton BN1 9QH, U.K.
davec@cogs.susx.ac.uk, jasonn@cogs.susx.ac.uk

SUMMARY

The vast majority of work in machine vision emphasizes the representation of perceived objects and events: it is these internal representations that incorporate the 'knowledge' in knowledge-based vision or form the 'models' in model-based vision. In this paper, we discuss simple machine vision systems developed by artificial evolution rather than traditional engineering design techniques, and note that the task of identifying internal representations within such systems is made difficult by the lack of an operational definition of representation at the causal mechanistic level. Consequently, we question the nature and indeed the existence of representations posited to be used within natural vision systems (i.e., animals). We conclude that representations argued for on *a priori* grounds by external observers of a particular vision system may well be illusory, and are at best placeholders for yet-to-be-identified causal mechanistic interactions. That is, applying the knowledge-based vision approach in the understanding of evolved systems (machines or animals) may well lead to theories and models which are internally consistent, computationally plausible, and entirely wrong.

1 INTRODUCTION

The vast majority of work in machine vision emphasizes the representation of perceived objects and events: it is these internal representations that are the 'knowledge' in knowledge-based vision and the 'models' in model-based vision. In this paper, we argue that such notions of representation may have little use in explaining the operation of simple machine vision systems which have been developed by artificial evolution rather than traditional engineering design techniques; and hence are of questionable value in furthering our understanding of vision in animals, which are also the product of evolutionary processes.

This is not to say that representations do not exist or are not useful: there are many potential applications of machine vision, of practical engineering importance, where significant problems are alleviated or avoided altogether by use of appropriate structured representations. Examples include medical imaging, terrain mapping, and traffic monitoring (e.g., Taylor, Gross, Hogg, & Mason, 1986; Sullivan, 1992).

But the success of these engineering endeavours may encourage us to assume that similar representations are of use in explaining vision in animals. In this paper, we argue that such assumptions may be misleading. Yet the assumption that vision is fundamentally dependent on representations (and further assumptions involving the nature of those representations) is widespread. We seek only to highlight problems with these assumptions; problems which appear to stem from incautious use of the notion of 'representation'. We argue in particular that the notion of representation as the construction of an internal

hypothesis, intelligent action involves the receipt of symbols from symbol-generating sensory apparatus, the subsequent manipulation of those symbols (e.g. using techniques derived from mathematical logic, or algorithmic search), in order to produce an output symbol or symbol structure. Both the input and the output have meaning conferred on them by external observers, rather than the meaning being intrinsic to the symbol (Harnad, 1990).

In the field of artificial intelligence, Newell and Simon's hypothesis licensed a paradigm of research concentrating on intelligence as the manipulation of symbolic representations, and on perception as the generation of those symbols and symbol structures. Specialised symbol-manipulating and logic-based computer programming languages such as LISP (e.g., Winston & Horn, 1980) and PROLOG (e.g., Clocksin & Mellish, 1984) (from "LIST Processing" and "PROgramming in LOGic" respectively) were developed to ease the creation of 'knowledge-based systems' (e.g., Gonzalez & Dankel, 1993). In due course, undergraduate textbooks appeared that essentially treated the hypothesis as an axiomatic truth (e.g., Nilsson, 1982; Charniak & McDermott, 1985), paying little attention to criticisms of the approach (e.g., Dreyfus, 1979, 1981).

In the field of machine vision, the Physical Symbol System Hypothesis underwrites all research on knowledge-based vision, where it is assumed that the aim of vision is to deliver symbolic representations (or 'models') of the objects in a visual scene: in the words of Pentland (1986), to go "from pixels to predicates". This mappingtrtific3017602m9(lan8269-delids)]TJ-19.5212Td.7(represen)999.654(tations)-15999.6(w)999.3as(to)-13999stu

about the 2-d image such as the intensity changes and their geometrical distribution and organisation. Following this, the primal sketch is processed to form the "$2\frac{1}{2}$-d sketch", which represents orientation and rough depth of visible surfaces, and any contours of discontinuities in these quantities, still in a viewer-centred coordinate frame. Next, the $2\frac{1}{2}$-d sketch is processed to form an internal "3-d Model", which represents shapes and their spatial organisation in an object-centred coordinate frame; including information about volume. Hence, the 3-d model is an internal reconstruction of the external physical world.

Within Marr's framework, formation of the 3-d model is the end of the visual process, and the model is then passed to 'higher' processes, such as updating or

vidual genome encodes for a useful design. The final evolved design can then be implemented and analysed to determine how it functions.

In evolving sensorimotor controllers, a variety of possible 'building blocks' can be employed: for a comprehensive review and critique, see Matarić and Cliff (1995). In many of the systems discussed in the next section, continuous-time recurrent neural networks (CTRNNs) are employed: these are artificial neural networks composed of 'neuron' units with specified time-constants giving each neuron an intrinsic dynamics. The primary reason for employing such neural networks

ary approach with minimal pre-commitments concerning internal architecture or representations makes the question "*what types of representation do these machines use?*" an empirical one. That is, we must examine or analyse the evolved designs, generate hypotheses about the representations employed, and test those hypotheses in an appropriate manner. Possibly the

tinguished from a correlation is by noting that Harvey's argument implies that representations are essentially linguistic (i.e., form an *interlingua* between representation-using agents or entities). A representation should hence be normative: it should at least offer the opportunity to misrepresent; to more or less correctly capture some external state of affairs. In the simple visual machines discussed above, there is no representation because there is no possibility of misrepresentation. We, the external observers, can point to the activity patterns and refer to them as representations in explaining the system, and be right or wrong to varying degrees about what those patterns represent. But to talk of the agent *using* the representations is to confuse patterns of activity which represent something else, and patterns of activity which actually constitute the agent's perceptual or experiential world, a point forcefully made by Brooks and Stein:

> "There is an argument that certain components of stimulus-response systems are 'symbolic'. For example, if a particular neuron fires – or a particular wire carries a positive voltage – whenever something red is visible, that neuron – or wire – may be said to 'represent' the presence of something red. While this argument may be perfectly reasonable as an observer's explanation of the system, it should not be mistaken for an explanation of what the agent in question believes. In particular, the positive voltage on the

knowledge and representations into the gannet visual system. Presumably the 'knowledge' concerns the utility of τ as an indicator of time-to-contact, and the ease with which it can be derived from an appropriately sampled optic flow-field. But, in the absence of clear definitions of

be addressed by further research, but statistical arguments have been presented as powerful alternatives to representational accounts of lower-order visual processes (e.g., Srinivasan, Laughlin, & Dubs, 1982).

The examples we have given here, from studies of insects, amphibia, birds, and humans, are by no means conclusive proof of our arguments. However, we believe that they are significant and persuasive because, although all of the visually mediated tasks involved could be performed using

Lee, D. N., Lishman, J. R., & Thomson, J. A. (1982). Regulation of gait in long-jumping. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 448–459.

Lee, D. N., & Reddish, P. E. (1981). Plummeting gannets: a paradigm of ecological optics. *Nature*, *293*, 293–294.

Lee, D. N., Young, D. S., Reddish, P. E., Lough, S., & Clayton, T. M. (1983). Visual timing in hitting an accelerating ball. *Quarterly Journal of Experimental Psychology*, *35A*, 333–346.