Chapter 6

Computer Vision — Craft, Engineering, and Science

K. Dawson[1], D. Furlong[1], M. Jones[1], N. Murphy[2], and D. Vernon[3]

[1]Trinity College, Dublin, Ireland
[2]Dublin City University, Dublin, Ireland
[3]Commission of the European Communities, Brussels, Belgium

6.1 Introduction

We come now to the final chapter. The emphasis throughout the book so far has been on the issues which influence the design of vision systems and the considerations which affect the evolution of the paradigms upon which vision systems are based. This chapter will continue in this vein, drawing out the issues which appear to play a pivotal role in the development of computer vision sytems. We will begin with the most general, and indeed the most fundamental, issue: the paradigms for vision which underpin the way we conceive of a vision system and the role we see it playing: as a truly autonomous system or as a utilitarian control system. Leading on from there, we will proceed to discuss the possible taxonomies of vision sytems, whether it is more, or less, useful to classify them on a functional basis (what they do) or on the basis of their constituent algorithms (how they do it). The use of knowledge, explicit or implicit, *a priori* or learned, raises its head here too, for the chosen taxonomy dictates not only the methodology for the design of vision systems but also the manner in which knowledge is incorporated into the system.

Both of these discussions lead us naturally to contemplate the state of play in the field of computer vision and its maturity as a discipline. It is in this context that the title of this book arose — *Computer Vision; Craft, Engineering, and Science* — for it reflects the mutually-relevant but distinct evolutionary stages of development of all disciplines.

Finally, we conclude the chapter , and the book, by asking two questions which are simple to pose, but difficult to answer: What is the role of a vision system and what is the best way to build one? Definitive responses would surely ring hollow but, in the light of the material presented here, we will venture some tentative answers. Their verity is something that only experience will decide.

6.2 The Paradigms for Vision

One of the purposes of this book is to assess and examine the accepted paradigms which are inherent in the field of vision system design. To initiate proceedings it is therefore appropriate to query how vision systems are constructed. Are they input/output systems in the normal engineering black box or transfer function schema? Are they autonomous systems, i.e. systems embodying closure in the form of a number of sophisticated interacting feedback loops? Or are they fundamentally process systems, either representational or connectionist? We need to examine the schema which we inherently adopt in approaching vision system design.

As we saw in chapter 2, an exciting area of development has been the advent of *active vision* systems which support interaction and necessarily require a view of the vision system as an on-going process involving a nexus of competing sub-systems rather than as an interpretative system, i.e. as a process of getting as much information as possible from a static image or a series of static images. Such an approach demands that we formally deal with attention, which issue does not even arise when dealing with interpretative vision systems. Indeed, the more traditional approaches to vision system design have focussed almost entirely on computational problems in the sense of seeking to provide interpretation of static scenes. When we impose real-time constraints, the *real* problem of the purpose of a vision system emerges. It could be argued that Marr, for example, completely avoided such questions because his focus was on computational efficacy. However, when freed of the shackles of the computational Formula 1 Grand Prix, vision systems can be seen in a different light. Vision as a process is not just the effective recovery of the shape of a giraffe from a given sequence of scenes, for example, but becomes a problem of *dynamic purpose tracking*, where the purpose of the system has become maleable. This kind of purposive dynamism raises questions of representation, i.e. if our vision system is to dynamically interact with a given environment then how does it maintain its representation of that scene? Does it involve some kind of representational updating? And if so what is the form of the representation? Is it a geometric or symbolic representation? Symbolic storage seems to be required in that there is a richness and freedom allowed by dealing with a more abstract level than geometric representation. However, this in turn raises the issue as to whether or not symbolic storage is hallucination or representation — a road map or the road. If it is the former then it might be true to say that vision as process is a question of controlled hallucination rather than updated representation, and that the ability of the vision system to expect or predict would be based on internal modelling or simulation rather than scene interpretation and representational modification. As such, a symbolic representation must involve *controlled hallucination*. Getting robust invariant image descriptions is far from easy and requires much work for we cannot as yet say whether they can be general purpose in nature, or whether they must be application-specific grouping algorithms. Only through laboratory experiment is this question likely to be answered, i.e., the proof will be in the doing.

This view of the vision system as a controlled hallucinatory environment is a move away from a representational paradigm toward a constructionist stance where, rather than adopting a paradigm of visual input/output, the system is viewed as dealing with sensor surface perturbations from which the symbolic representation is manufactured. But, in such constructionist systems how does it know what to do? From where does its purpose derive from? Typically, the answer involves a rule-based supervisor of some sort who selects perceptual tasks based on a pre-defined menu of options. Only through moving more towards autonomous systems does the possibility of evicting the homunculus agent become available. Thus, a continuous spectrum of vision system paradigms begins to emerge: from representational systems, through constructionist systems, to autonomous systems, where many hybrid shades are feasible (see figure 6.1).

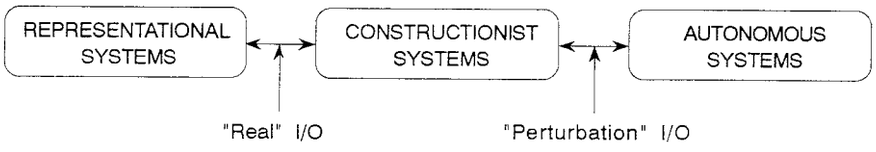


Figure 6.1: A spectrum of vision paradigms

At the representational system end of the spectrum, there is the inherent requirement of isomorphic mapping of some external world; at the autonomous end, this isomorpism is specifically denied and consequently the issue of isomorphic mapping does not even arise. In the middle ground, constructionist systems can be viewed as *almost* autonomous agents which are given purpose by some pre-determined rule-based system which will use sensor pertubations as a ground for its decisions but which will require isomorphic mapping. Are all these vision system paradigm options equally valid? Just what constitutes a *real* vision system? Is an array of photo sensors a vision system? Or would tacking on an expert system transform such an array into a vision system? It would appear that the answer depends very much on the context and that therefore *the choice of vision system paradigm is application dependent.* A vision system can mean different users. It may well be that there is no one all-embracing paradigm which characterises a vision system.

For many applications (e.g. inspection systems) relatively simple vision systems are pragmatic, but if the quest is *vision understanding* then we are of necessity required to consider issues of autonomy and intelligence. In this respect, the move toward vision as process is opening up new test-beds in that the functionality of the vision system is being relegated to a subservient layer with the focus being put on the development of intelligent homuncular agents. However, although this approach can be seen as a move along the spectrum of vision systems toward autonomy, these systems are not truly autonomous in that there is still reliance on pre-programmed,

rule-based task allocation. If a vision system cannot define its own goals, can we call it truly intelligent? Does such *homuncular autonomy* advance matters much? Developments in the field are of the nature of application specific solutions and are certainly not proving to be a panacea for general-purpose vision systems. To expand the relevance of our relevance of our research efforts requires that we expand our contextual perspective with the attendent that we specifically address issues of true autonomy. In this light, a very fundamental observation seems to be that there is not, as yet, any well developed *science* of autonomy. Much work remains in the definition and development of such a science which, from a research point of view, is good news indeed.

It might well be asked if effort expended in the development of solutions for specific problems in intelligent vision system design is energy well spent, for when we view such efforts from a more removed perspective it seems that we are attempting to erect magnificently ornate structures on very shaky foundations. If the same efforts were to be invested in the investigation of the principles of a science of autonomy then the long term rewards might be of far greater significance. There is an apparent reluctance on the part of vision system designers to invest in autonomy and, with regard to vision systems, the related topics of perception and intelligence. This is not frequently made explicit as advances in the form of pragmatic results (computational efficiency increases, or superior edge detection algorithms, for example) are far more gratifying and more publishable than any acknowledgement of profound ignorance when faced with problems of autonomy, perception and intelligence. In effect, it would appear that the vision system research community is phase-locked to a constricted domain by the mutual lack of admission of ignorance on the part of those involved.

Herein lies the crux of the matter, for if we cannot apprehend (and, worse, cannot even admit that we cannot apprehend) what natural vision is, then it is most unlikely that we can develop artificial vision systems. In effect, all we are doing is groping in the dark and gauging our efforts by the effectiveness of the results achieved. This might indeed be a pragmatic approach but is hardly intelligent, for the number of interacting variables involved and the range of strategies possible are so large as to mitigate against the achievement of general purpose results within several lifetimes, if at all. However, if we were to get the foundation right, and in the context of vision systems this means grasping what vision is, then the efforts to develop general purpose vision understanding systems are likely to be more efficient and to yield results of greater value and of more universal application than those which have been achieved to date. To extend the analogy, it might be said that many of the pre-fabricated functional elements — the cladding — that might go into the constitution of general purpose vision understanding systems are in existence, but that what is lacking is an approach to stable foundation and framework design onto which these functional elements might be attached. There is nothing to suggest that these fundamental issues are beyond the scope of the ingenuity of those working in vision system design. It is simply a matter of lack of attention to date. Therefore, the responsibility now lies

firmly with vision system designers to address these foundational issues - what is it to see, what is it to be intelligent, what is it to be autonomous.

6.3 Taxonomies of Vision Systems

There are two chief conclusions arising from the arguments in the previous section: first, that the long-term goal of robust general-purpose computer vision will necessarily involve a well-founded understanding of autonomy; and second, that much can be achieved, in the short term, by adopting pragmatic, application-specific, approaches to vision which exploit what is known and what has been established about visual sensing.

Let us now set aside the elusive, if important, goal of understanding autonomy and concentrate on the pragmetic issues; the issues which will facilitate short-term benefit, if not long-term truth. We begin by considering the manner in which we delimit, or classify, pragmatic vision systems. We do this in the expectation that, if we choose the appropriate classification criteria, and hence the attendant taxonomy, we might have greater success in designing vision systems which match well the chosen application.

The table of contents of most computer vision textbooks provides the reader with a list of algorithms and by deploying these algorithms, an expert is supposed to be able to construct a vision system. This sounds plausible. However, something is clearly missing: this is the knowledge and experience which is required to select and combine the algorithms in a manner which is most appropriate for a given task. Alternatively, one might address vision systems on the basis of their functionality; for example, by describing how one might go about addressing tasks such as gauging, pose estimation, free-space mapping, or terrain modelling. As a functional taxonomy is a higher level of abstraction than an algorithmic one, it seems that a mapping from the functional to the algorithmic taxonomy is required. The difficulty is that there can be considerable choice in the selection of the algorithms which best suit the task and it is here that vision becomes somewhat more *ad hoc* since there is no formal and well-grounded method for the selection of the algorithms. On the other hand, algorithmic problems are typically well specified and are mapped directly into computational solutions. Functional problems in vision, though, are often poorly specified (usually just through the provision of 'typical' example instances) and at present little has been done to formally relate functional problems to particular algorithmic solutions. Nonetheless, a functional taxomony seems to be the more useful of the two, although it necessitates the existence of the mapping between functionality on the one hand and algorithms and representations on the other.

Almost all vision systems use knowledge to a lesser or greater extent. There appear to be, at least, two forms of knowledge: there is domain knowledge and there is the knowledge base, which is accessible to the vision system, and which is abstracted, or filtered, from that domain knowledge. This knowledge base can comprise procedural knowledge or declarative knowledge, with implicit or explicit representations

of information. The decision as to which knowledge is 'important' from the entirety of the domain knowledge is what determines the processing model which is adopted at the algorithmic level of the vision system. It seems that, quite often, when a vision system is being designed, you don't initially decide on a particular processing model, i.e. a particular type of algorithm, but instead you look at what knowledge is important, and based on that you decide on the processing model. It is the knowledge which is important rather than an initial decision about which algorithm to choose. Of course, these two are to an extent mutually-dependent, but the issue is the order in which the choice is made. In essence, there is first the transformation from the application domain to the knowledge base and this, then, implicitly defines what type of techniques can be used. For example, the industrial inspection system which was described in the first chapter uses correlation-based rules to make the necessary quality assurance decisions, but the signatures which represent the 'useful' information about the application are what are primary and were developed first — the correlation paradigm followed. This observation about the ordering of choices in the design of a computer vision system — first knowledge, then algorithms — seems to suggest that indeed a functional taxonomy would be the most appropriate. However, the critical mapping from the functional and the knowledge to the algorithms still remains to be made explicit and formalized.

6.4 Vision in a Broader Context

Consideration of the possible means of classifying vision systems and the way in which they are designed leads to a broader issue; that of the state of play of vision in general. There are three different manners in which vision (or in fact most problems) can be viewed; as a *craft*, as an *engineering* discipline, or as *science*.

The craft approach encourages the solution of particular problems by any means at our disposal through the judicious use of previous experience and intuitive understanding of the problem, even though those means do not necessarily have any formal basis. Craft vision takes a problem and solves it by hammering what is known about vision and the particular application into (the required) shape. This is exactly how the visual inspection system described in chapter one was designed: by looking to see what reliable information could be easily extracted from the images of the scene, and by identifying the information which best characterized the the information. Having decided on the information, we then come to decide on the most effective way to extract, process, and analyse the information (in this instance, by thresholding, signature formation, and correlation).

Science may be regarded as a means of formally addressing problems in strict, rigorous, well-founded, and mathematical way. But it may also be viewed as a means of formalising the methods which are found, by experience, to work. For example, consider both the theory of invariance in differential geometry for extracting curves in images and the simple *ad hoc* techniques for performing the same operation.

On the other hand, engineering is the application of science to particular problems using some reasonably well understood and systematic methodology. Engineering constitutes the middle ground, the methodological and well-grounded application of theories in the context of specific application constraints.

The relationship between craft, engineering, and science is not, however, a simple linear spectrum with craft at one end, engineering in the middle, and science at the other end. For, while science and craft may appear to be poles apart, there is nonetheless a subtle relationship. Craft supplies the intuition (and experience) which is essential to the formulation of scientific principles and science rationalizes, *a posteriori*, the intuitive and heuristic by elucidating the (hidden) theory on which it rests. The relationship then is a circular one in which each of the three terms feed off one another, contributing, in total, to the advancement of the area (see figure 6.2).

It would appear that the majority of application-oriented vision systems have been developed by the craft of the vision engineer. Our goal is, or should be, the development of the scientific principles and engineering methodology underlying vision so that vision problems may all be addressed in a systematic and formal fashion (i.e. moving from craft to engineering under the guidance of science).

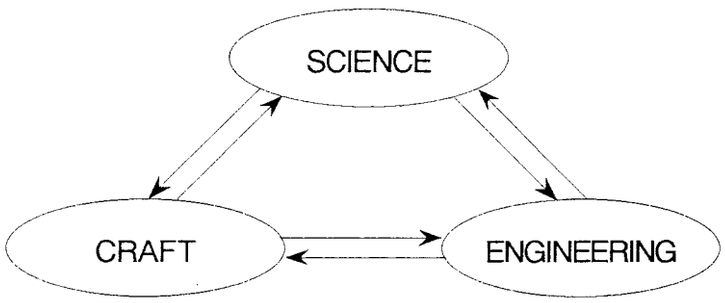


Figure 6.2: The evolution of the discipline of computer vision

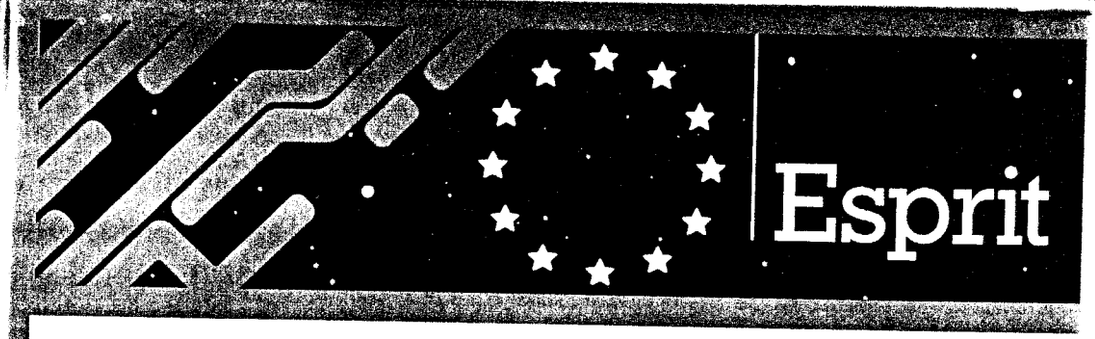
6.5 Conclusions

It is fitting that we should conclude a book such as this, which has attempted to look under the surface of the popularly-accepted mantle of computer vision, with two bold, if not brash, questions: What is the role of a vision system and what is the right way to build one?

As we have noted above, we must stipulate the type of system which is under consideration before answering such questions. For application-specific vision systems, the role must be to meet the demands required for that specific application. For more general ones, e.g. autonomous vision systems, the role is seen to be one

of survival of the perceiver. Between these two extremes, many other sub-roles can be identified, e.g., to obtain useful information from the environment or to make decisions about the state of that environment. Given the evolutionary model of Craft, Engineering, and Science of computer vision as a discipline, there was an acceptance that our current stage of evolution corresponds mostly to the Craft stage, with notable developments at both the Engineering and Science stages. Consequently, it was agreed that, while it would be ideal to be able to build general vision systems, for now the vision community should adopt, in part at least, a bottom-up approach to building application- and domain-specific vision systems in an effort to maximise the benefit of Craft approaches while developing the engineering and the scientific: The vision community cannot and should not wait until there is a perfectly complete and self-consistent theory of vision before they try to progress. The formation of taxonomies of current techniques and their extension was considered to be one way of making the transition from the craft to the engineering phase.

This is sound advice indeed. But it would be wrong to conclude on such a pragmatic and, from the research point of view, such a limiting note, for in no sense was the tenor of the discussion at workshop one which reflected a lack of ambition. On the contrary, the strong feeling was one of challenge: challenge in promoting the integrity of the vision community and challenge in advancing the state-of-the-art of computer vision. But it is a well-understood sense of challenge: knowing what is to be addressed — the autonomous, the constructionist, the representational — and why and how it is to be addressed — for the advancement of the science of vision, for the promotion of principled engineering of vision sytems, and for the exploitation of the craft-oriented experience we have in applying our current understanding of computer vision.



Basic Research Series

D. Vernon (Ed.)

Computer Vision: Craft, Engineering, and Science

Workshop Proceedings