COMPUTER VISION

J. M. BRADY, editor

north-holland

Computer Vision

Edited by

Michael Brady

Artificial Intelligence Laboratory, M.I.T., Cambridge, MA, U.S.A.

Contributors

H.G. Barrow
T.O. Binford
J.M. Brady
R.A. Brooks
B. Chandrasekaran
L.S. Davis
S.W. Draper
B.E. Flinchbaugh
J.P. Frisby
B.K.P. Horn
K. Ikeuchi
T. Kanade
J.E.W. Mayhew
H.K. Nishihara
A. Rosenfeld
B.G. Schunck
K.A. Stevens
J.M. Tenenbaum
A.P. Witkin
R.J. Woodham

50
1931 N·H 1981
P C

North-Holland Publishing Company
Amsterdam

Reprinted from the Journal
Artificial Intelligence, Vol. 17, August 1981

PRINTED IN THE NETHERLANDS

COMPUTER VISION

CONTENTS

Preface — The Changing Shape of Computer Vision

Michael Brady
Artificial Intelligence Laboratory, MIT, Cambridge, MA 02139, U.S.A.

1. Introduction

This special volume of the *Artificial Intelligence* (*AI*) Journal recognises the considerable advances that have taken place in Computer Vision over the past decade. It contains fourteen papers that are representative of the best work currently in the field. Apart from being a state-of-the-art account, the issue has been designed, as far as possible, to serve two rather different aims.

First, it is intended to give AI researchers in fields other than Vision an opportunity to become familiar with recent developments in the field. The continuing growth of AI inevitably makes it difficult to keep abreast of progress in any but a narrow area. As we shall see, the increasingly technical content of Vision, and its growing concentration on visual perception, rather than on general AI ideas, make it doubly forbidding to the casual AI reader.

Second, it is intended to enable vision researchers from fields other than AI to get a clearer picture of what an AI approach to their problem might be, or might contribute. Certainly, there is increasing interest in Computer Vision among researchers in fields as disparate as psychophysics, neurophysiology, signal and image processing, optical engineering, and photogrammetry. In addition, remote sensing, visual inspection of industrial products, and other applications in the growing field of Robotics, ensure that Vision will continue to be a topic of considerable importance for many years to come.

2. As it Was in the Beginning

Even as late as 1975, Computer Vision looked rather different than it does today. A great deal of effort had been expended on the 'blocks' microworld of scenes of polyhedra. Huffman [26] and Clowes [11] had noted the advantage of making the image forming process explicit. They observed that *picture lines* and *junctions* were the images of *scene edges* and *vertices*, and they catalogued all

Artificial Intelligence **17** (1981) 1–15
North-Holland Publishing Company

those interpretations of lines and junctions that were possible, given the prior assumption of planarity and the restriction that at most three surfaces were allowed to meet at a vertex. These interpretations, taking the form of 'labellings' of lines, amounted to local constraints on the volume occupied by a vertex. The local constraints propagated along picture lines since planar edges can not change their nature between two vertices. (This is not so for curved lines, as Huffman noted. Turner [59] described one possible extension to curves of the Huffman–Clowes approach. Binford, Barrow and Tenenbaum describe rather different approaches in this volume.)

Huffman [26] further pointed out that the local vertex constraints were not sufficient to capture the important restriction that picture regions were the images of planar surfaces. Mackworth's [34] development of *Gradient Space* was expressly intended to repair this deficit. Draper's article in this volume describes the greater competence of Mackworth's program, as well as its shortcomings. Despite this, most line drawings had a remarkable number of possible interpretations. Waltz's [60] work introduced the inherently global constraint afforded by shadows cast by a single distant light source, and showed that the multiple ambiguities possible without lighting were often completely resolved to a unique interpretation with lighting. Furthermore, the process by which the unique interpretation was discovered, naturally lent itself to parallel processing of a particular sort. Each vertex had an associated processor, and they all operated in strict synchrony. At each stage, the processors changed their states depending on the state of those directly connected to them. Rosenfeld, Hummel, and Zucker [54] noted the connection between this scheme and relaxation processes in numerical analysis. Actually, several authors had suggested the use of local parallel processing for Vision rather earlier, see for example the historical remarks in Subsection 5.2 of the paper by Ikeuchi and Horn in this volume.

Waltz's scheme had a number of drawbacks. For example, as Winston [63] pointed out, the program could make no use of the direction of lines in the image. On the other hand, Mackworth's program could, since gradient lines are perpendicular to image lines (see the discussions of Gradient Space in the papers of Draper, Kanade, Woodham, and Ikeuchi and Horn in this volume.) Huffman [27] later defined *Dual Space* in which this information could be made explicit (see the articles of Draper (this volume) and Spacek [58]). Again, the Waltz labellings were extremely complex and unstructured. In essence this was because they constituted interpretations that confounded many different sources of information about lighting, surface cracks, occlusion, and edge type into a single label. Clearly these different contributions to the entire percept should be made explicit and exploited separately, as they are in the human visual system. Binford's paper in this volume reconsiders the possible interpretations of edges.

A second strand in the development of Computer Vision concerns what was referred to as 'low level' processing. It was more art than science, and largely consisted of methods for the extraction of the 'important' intensity changes in an image. In the blocks world these correspond to shadow boundaries and the edges of visible surfaces, including depth discontinuities. The approach mostly consisted of convolving images with local operators (typically 3 by 3 on a 256 by 256 image) to estimate the position, contrast, and orientation of the important intensity changes. Operators were tuned to particular applications, and fared badly outside their limited domain and in the presence of noise. Little serious analysis of actual intensity changes, including the signal to noise characteristics of real images, had been carried out. A singular exception was the work of Herskovits and Binford [16]. They suggested that there were basically three qualitatively different types of intensity changes; and that particular changes often combine features of two types. This analysis formed the basis of a line finder whose performance considerably advanced the state of the art.

Other work in 'low level' vision largely consisted of the design and construction of region finders. Region finding, essentially the dual of edge finding, aimed to isolate those regions of an image that were the images of perceptual surface patches. It was thought that such regions might be isolated by defining some descriptor with respect to which they were uniform, and distinguishable from surrounding regions. It was soon clear [2, 9] that even if such descriptors existed, they were not defined simply in terms of grey level intensity values. Some researchers proposed multi-spectral descriptors [48], while others later flatly denied that it is possible to define adequate descriptors at all [36, p. 64]. Binford's paper in this volume discusses region finding in some detail.

By the early 1970's, the consensus was that 'low level' vision was inherently incapable of producing rich useful descriptions. It was observed, by analogy to the apparent need for semantics in parsing English sentences, that downward flowing knowledge of the scene could provide additional constraint. This in turn could inform local decision making. A number of program structures were proposed to effect this interaction between top down and bottom up processing of information [4, 7, 12, 43, 55, 64]. Similar ideas were advanced about natural language understanding, and speech perception. This influenced the design of, for example, Hearsay 2 [31]. To experiment with these ideas, entire systems were constructed which mobilised knowledge at all levels of the visual system as well as information specific to some domain of application. In order to complete the construction of these systems, it was inevitable that corners were cut and many over simplified assumptions were made. By and large, the performance of these systems did not give grounds for unbridled celebration. The authors of the KRL proposal (Bobrow and Winograd [6]), for example, listed several common failings (see also [7]).

3. Is Now

Perhaps the most fundamental differences between Computer Vision as it is now and as it was a decade ago, stem from the current concentration on topics corresponding to identifiable modules in the human visual system. This volume contains papers, for example, on stereopsis, the interpretation of surface contours, the determination of surface orientation from texture, and the grouping of motion primitives. To be sure, there is still a considerable amount of work oriented toward applications, but it is also increasingly based on detailed and precise analyses of specific visual abilities. The focus of research is more narrowly defined in terms of visual abilities than in terms of a domain, and the depth of analysis is correspondingly greater. This change has produced a number of far-reaching effects in the way vision is researched. This section attempts to make them explicit.

One obvious effect has been a sharp decline in the construction of entire vision systems. Most AI vision workers have thankfully abandoned the idea that visual perception can profitably be studied in the context of a priori commitment to a particular program or machine architecture. There is, for example, no more reason to believe that 'relaxation' style processing will of itself tell us more about vision than did the excursions into heterarchy. There is no obvious reason to be encouraged by Reddy's [51] claim that the Hearsay 2 model can be adapted mutatis mutandis to vision.

What identifies a particular operation as a distinguishable module in the visual system? Normal vision confronts and exploits massive redundancy. Some of the most solid evidence for the claims of individual modules is offered by psychophysical demonstrations. Care is taken, as far as possible, to isolate a particular source of information and show that the operation in question survives. One particular instance of this is the study of patients with certain disabilities resulting from brain lesions (for example [42, 57, 61]). Many psychophysical experiments, seemingly isolating particular modules of the (human) visual system, have been reported in the literature. Notable examples include Land's demonstration of the computation of lightness [19, 30] and Julesz's [28] demonstration of stereoscopic fusion without monocular cues. In some cases there is clear evidence of a human perceptual ability, although such evidence would hardly be referred to as psychophysical. Horn's work (see the papers by Woodham, and Ikeuchi and Horn in this volume) concerns the highly developed human ability to infer shape from shading. Steven's paper concerns the human three-dimensional interpretation of surface contours. On the other hand, it is equally clear that we do not have a specific module in our visual system to recognise 'yellow Volkswagens' (see for example [62]). It is less clear whether we compute depth directly, as opposed to indirectly through integrating over surface orientations, or what use we make of directional selectivity, optical flow, or texture gradients.

Not all modules work directly on the image. Indeed, it seems that few do. Instead they operate on *representations* of the information computed, or made explicit by other processes. In the case of stereopsis, Marr and Poggio [40] argue against correlating the intensity information in the left and right views.

Instead they suggest that so called zero-crossings are matched (see [15] or Nishihara's paper in this volume). The paper by Mayhew and Frisby argues that the matching actually takes place on a different representation, called the primal sketch [35]. In any case, a great deal of attention has centered on the isolation and study of individual modules, and in each case on the development of the representations on which they operate, and on those that they produce. The first of these representations, and the one whose structure is least subject to dispute, is the image itself. Not surprisingly then, most attention has centered on those modules that operate upon the image. As we shall see, the further we progress up the processing hierarchy, the less secure the story becomes, as the exact structure of the representations becomes more subject to dispute. Again, this is not surprising. The image aside any representation is one module's co-domain and another's domain. All of them shape its eventual structure.

3.1. Modules that operate on the image

A great deal of effort has been devoted to understanding how the important intensity changes in an image can be extracted, and how the information can be best represented. Marr [35] coined the term primal sketch to describe such a representation, and he described a particular algorithm by which it might be computed. A novel feature of the work was its direct reference to neurophysiological and psychophysical findings, a commitment Marr was to continue to stress in later work. His work with Poggio led to a revision of the process of construction of the Primal Sketch. Instead they advocated the use of zero-crossings of the second derivative of the filtered image. This idea was developed in turn by Marr and Hildreth [38], who propose that an image is first filtered by four Gaussians having different bandpass characteristics. Then each filtered image is convolved with a Laplacian operator (see Nishihara's paper in this volume for more detail). One of the novel features of the Marr–Hildreth account is the size of the operators involved, the smallest being roughly 35 picture elements square. This is in stark contrast to conventional operators, which are still typically on the order of 5 by 5. Such a large operator can be in much closer agreement with a Gaussian (or any other filter for that matter) than a small operator, and its effects are therefore more predictable. Unfortunately it is no longer obvious how to compute the assertions that Marr had previously advocated for inclusion in the primal sketch (see [17, p. 75]). The whole issue of constructing the primal sketch from zero-crossings is far from being resolved. Binford's paper in this volume considers this issue, as well as the choice of an optimal filter and the use of non-oriented masks, in fair detail.

Intensity changes aside, Horn and his colleagues have studied the perception of surface shape from shading. Their work is represented in the current volume by the papers of Ikeuchi and Horn, and Woodham. In brief outline, Horn has formulated a second order differential equation which he calls the image irradiance equation which relates the orientation of the local surface normal of a visible surface, the surface reflectance characteristics, and the lighting, to the intensity value recorded at the corresponding point in the image. Horn quickly realised the need for a representation which makes such surface orientations explicit. Two parameters are needed. Horn [22] observed that gradient space provides such a parameterisation, and showed how the relationship between intensity values and surface orientations could be added to gradient space to form what he called the *reflectance map*. The papers by Ikeuchi and Horn, and Woodham give details. Gradient space is by no means the only two parameter representation of surface orientations. Ikeuchi and Horn investigate *stereographic space*, which has the additional desirable property that the constraints offered by occluding boundaries can be represented and exploited. The output of shape-from-shading is a representation that makes explicit the orientation of visible surfaces, and may make other information such as depth and surface orientation discontinuities explicit also. Horn [23] suggests the name *needle map* for the representation. Other representations have been proposed which make substantially the same information explicit. Marr [36] labels this representation the $2\frac{1}{2}D$ *Sketch*, and Barrow and Tenenbaum [4] discuss *intrinsic images*. Again the exact nature of the representation (or representations) is currently far from clear. In part this is because very little work has been devoted to modules which operate upon it.

Finally, Horn and Schunck (this volume) propose a method by which the so-called *optical flow* can be determined from a sequence of images. Several authors have investigated the information that can in principle be computed from ideal optical flow fields (see the references in Horn and Schunck's paper), but no proposals have previously been made for its computation.

3.2. Modules which operate on zero-crossings and the primal sketch

We pointed out in the previous section that there remain a vast number of unresolved issues concerning the nature of the primal sketch and its computation from zero-crossings of whatever kind of filtered image. Nevertheless, the broad outlines are clear enough for work to proceed to investigate modules which are assumed to operate upon those representations. Indeed it is necessary that it does, as it will also contribute to our understanding of the information that needs to be made explicit in the primal sketch, and hence its eventual form. One area that is not represented in this volume, but that is of considerable importance, is the investigation of the processes which impose hierarchical structure on the primal sketch (what Marr [35] called the *full primal sketch*). Riley [53] has made an initial study of such processes for static scenes. Motion is an important source of information of determining structure.

The paper by Flinchbaugh and Chandrasekaran in this volume addresses grouping on the basis of motion cues. Such grouping operations play an important role on all the representations used by the visual system, and for the most part they are poorly understood. Little if any work has been done on grouping operations on what we call the *surface orientation map*.[1]

Considerable attention has been paid to *stereopsis*. Marr and Poggio's [40] theory of human stereopsis, and its implementation and refinement by Grimson [15] is discussed at length by Mayhew and Frisby, who propose a number of further refinements.

Ever since Gibson [14] stressed the importance of texture gradients for the perception of depth and surface shape, they have been the subject of detailed psychophysical and computational investigation. Pattern recognition approaches typically consist of computing crude statistics on the image intensities. This does not work at all well since, as Horn in particular has shown, an individual intensity value is a complex encoding of the lighting, the surface reflectance characteristics, and the local surface orientation. Witkin's paper in this volume once more underscores the importance of making the image forming process explicit. His approach relies upon statistical arguments but, crucially, does not require that natural textures are uniformly distributed. Rather, it requires that their non-uniformity does not mimic projection. It relies upon deriving a probability density function which relates the orientation of a scene element via projection into an image element.

The papers of Draper, Kanade, Stevens, and Tenenbaum and Barrow address various aspects of the human ability to perceive surface shape from line drawings. The first two of these assume that the scene is composed of plane-faced objects. As such, they continue the tradition of work discussed in Section 1. Kanade's paper combines the ideas of gradient space and edge labellings. It proposes the two additional assumptions of parallelism and skewed symmetry to further constrain the orientation of a planar surface. Matching the intensity profiles across two edges provides further constraint. Crucially, the program is able to make the conservative inference that two edges have the same interpretation without knowing exactly what it is. Draper's paper discusses the limitations of gradient and dual space in supporting possible processes that interpret line drawings of polyhedra. He proposes instead symbolic reasoning about 'sidedness'. Unfortunately, such inferences are inherently long range, since they rely upon the observation that the relationship between planar regions is fixed, and therefore common to all points at which they intersect. Such reasoning is likely to be of limited usefulness when applied to images of natural or curved scenes. Tenenbaum and

[1] We sincerely hope that this name does not become established in the literature, as it only serves as a name for the intuitive notion which is rendered more or less precise in the three published versions referenced (namely the $2\frac{1}{2}$D sketch, needle map, or intrinsic image).

Barrow address the subject of interpreting line drawings of curved surfaces. They use junction labellings to determine whether a bounding curve depicts an extremal boundary or a depth discontinuity. Then they propose two mechanisms: one for computing the spatial layout of the bounding curves and one for interpolating local surface orientation from the boundary values. In the remaining paper on this topic, Stevens proposes a taxonomy of interpretations of surface contours. By investigating intersecting contours in an image, a local decision can be reached about the nature of the underlying three dimensional surface.

3.3. Object representations

Considerably less is known about the modules which operate upon the surface orientation map to produce object representations, and the nature of those representations is very far from clear. Some work has been done, and it is well represented in this volume. Binford [5] proposed a volumetric primitive known as *generalized cylinders*. Nevatia and Binford [46], Hollerbach [18], and later Marr and Nishihara [39] developed representational schemes based upon such volumetric primitives. Brooks (in this volume) describes the representation of complex objects such as motors and airplanes, the incorporation of constraints such as symmetry, and the specification of affixment relations by which the local coordinate frames of two objects can be inter-related. Marr and Nishihara [39] discuss the role which such representations might play in human vision (see Nishihara's paper in this volume).

3.4. Methodological comments

The previous sections have discussed some of the modules and the important representations which have begun to emerge in Computer Vision. The broad outlines are clear, even if there are many major unresolved questions in nearly every facet of the subject. We may also note some further common themes which have crystallized over the past decade.

Most of the analyses sketched above start out with a precise description of the domain and co-domain of the visual process under scrutiny. Increasingly, 'precise' means 'mathematically precise', and so Computer Vision has become steadily more technical. This is not to say that Vision was not technical before, rather it alludes to the increasing occurrence and sophistication of mathematical analyses in Vision. Many observations about the world, as well as our assumptions about it, are naturally articulated in terms of 'smoothness' of some appropriate quantity. This intuitive idea is made mathematically precise in a number of ways in real analysis, for example in conditions for differentiability. Relationships between smoothly varying quantities give rise to differential equations, such as Horn's *image irradiance equation*. We commented several times above on the value of making the image forming process explicit. This in turn leads to a concern with geometry, such as the properties of the gradient, stereographic, and dual spaces. Combining the considerations of geometry and smoothness leads naturally to multi-variate vector analysis and to differential geometry [13]. Mostly, a representation does not of itself contain sufficient information to guarantee that a module can uniquely arrive at the result computed so effortlessly by the human visual system. Additional assumptions, in the form of constraints, are required. This observation has led to a concern with constraint satisfaction and equation solving, using the techniques of numerical analysis such as Gauss–Seidel iteration and Lagrange multipliers (especially in the form of the calculus of variations). Examples of all of these approaches can be found in the papers in this volume.

For many authors, the changing style of research in Computer Vision has not been simply a matter of a narrowing of attention and a more highly developed technical content. Instead, greater significance is attached to the desire to make explicit the links between their work and corresponding theories in psychophysics and neurophysiology. From this perspective Computer Vision has as its goal the construction of computational theories of human visual perception. In large part, this approach stems from a series of papers written by David Marr and his colleagues at MIT. Marr's work stems from a background in neurophysiology, and is expressly addressed to psychophysicists and neurophysiologists. In particular, it is couched in terms they are accustomed to, and makes extensive reference to their literature, rather than that of Computer Vision. The work of the MIT group has excited considerable interest among psychologists and neurophysiologists, and is extensively referenced in the papers in this volume. A book summarising Marr's thoughts about human visual perception [37] and incorporating summaries of the contributions he and his colleagues have made across the entire range of the subject is currently in press.

There is considerably less diversity in emphasis, subject matter, and technical content than might be imagined between those researchers who see themselves constructing a computational theory of human visual perception and those for whom human visual perception is at most a matter of secondary concern. Compare, for example, the ACRONYM representation of objects based upon generalized cylinders (see the paper by Brooks in this volume) with that proposed by Marr and Nishihara [39], or the work on early processing of motion by Horn and Schunck (this volume) with Marr and Ullman [41]. Another common research theme is the need for local parallel processing which can discover global information through propagation. The paper by Davis and Rosenfeld (this volume) considers one such class of program structures, while others can be found in the papers by Horn and Schunck, Ikeuchi and Horn, Tenenbaum and Barrow, and Woodham. Such architectures naturally lend themselves to realization in hardware. Nishihara describes one such realization.

4. And Ever Shall Be?

As this introductory survey suggests, Computer Vision has progressed considerably on many fronts over the past decade. There has been a change in the style of research as well as in its substance. However, most issues are still poorly understood, from the exact form of representations, through the detailed understanding of the individual modules, to topics that have so far received little or no attention. A sampling of unresolved problems follows in the next few paragraphs. It is by no means exhaustive.

First, the details of what we have called the surface orientation map need to be made precise. Marr [36], Horn [23], and Barrow and Tenenbaum [4] have suggested that it records local surface orientation, as well as depth discontinuities; but it is unclear how they are recorded. Suggestions include Cartesian and polar formulations of the gradient, 'sequins' versus 'quills' [23], and the separation of various kinds of information into separate 'intrinsic' images. Nor is it obvious how accurately values are recorded. It is clear that surface information needs to be represented at different levels of resolution: a pebbled path may be considered approximately planar by a human who is walking along it. Yet an ant or person on roller skates may find the same path extremely difficult to navigate; in such cases the path is unlikely to be considered planar. As this example indicates, the level of resolution of a representation is determined largely by the process operating upon the representation, and there has been little investigation of such processes to date.

It is equally clear that grouping operations need to be defined at each level of resolution of each representation in the visual system, in order to impose hierarchical structure upon the representation. The advantages that should accrue from imposing such structure are likely to be precisely those which have inspired the development of data structures generally in computer science. Consider as an example a simple egg tray. The pattern of identical depressions to hold the eggs is immediately obvious, even though the detailed description of the individual egg cells is not.

A related set of problems concerns the determination of surface properties such as its color, manufacture, and whether or not it is wet, slippery, or prickly. Granted that we make such properties explicit, we need to determine whether they are attached as local descriptors to representations such as the surface orientation (say), or whether they are the content of separate representations. It may be that there is a separate albedo map [24, 25] or it may be that albedo information is embedded in the surface orientation map. Actually, the entire question of the computer perception of color is still very much in its infancy, despite its enormous literature.

Our current understanding of motion perception is crude. Horn and Schunck's paper is a preliminary account of the computation of optical flow from grey levels. It is less clear what information can be recovered from optical flow. Some authors are enthusiastic about the richness of the information it can provide (Clocksin [10]), while others are more sanguine (Prazdny [50]). Marr and Ullman [41], and Richter and Ullman [52] have made a start towards determining motion from the displacement of intensity changes. Ullman [65], and Flinchbaugh and Chandrasekaran (this volume) consider the grouping of primal sketch tokens in motion. Even less is known about motion computed on the surface orientation map or on object representations. It is reasonable to suppose that the description of such object motions will need to incorporate a formulation of the object's kinematics. This has proved to be quite difficult even for simple robot arms (see for example [49]).

Perhaps the most difficult problem of all concerns the perception or planning of movements through cluttered space. Space, considered as an object, typically occupies a volume and surface whose descriptions push current representational frameworks to their limits, if not far beyond them. Some progress has been made in Robotics [32]. A further important application lies in making precise the rather vague motion of cognitive map. It is usually supposed [33] that this only refers to object representations. Actually it seems that we have quite considerable navigational processes which operate on the surface orientation map.

The current rapid pace of developments in VLSI technology has further motivated research into what were referred to above as local parallel programming architectures. It is likely that our conception of computation will change as a result of such developments. Vision will be one of the first areas to benefit from such advances. It seems that it will also be a continuing source of inspiration to VLSI designers [1, 47].

Finally, we certainly need a better understanding of the extent and use of domain specific information in visual perception. Yesterday's heterarchy and today's multi-layered relaxation systems both derive from a priori commitment to a particular mechanism. The experience of the past decade should certainly have made us wary about jumping to premature conclusions regarding which phenomena appear to inevitably implicate such downward flow. This has certainly been true of our ability to compute rich useful descriptions of the information provided in an image. It also seems reasonable to suppose that the three dimensional structure of jointed rigid objects can be recovered from a time succession of images without knowing a great deal about human physiology. This would provide an explanation for, amongst other things, the demonstrations of Johannson [29] and Muybridge [44], knowing only the basic facts of dynamics.

There is every reason to believe that there will be considerable advance on these and other issues over the next few decades, probably resulting in changes in our conception of Computing and Vision at least as large as those which have occurred over the past decade. It would be a very brave person indeed who claimed to understand other than the broadest outlines of the subject now.

5. Professor David Marr

One paper which was to be written especially for this special issue of *Artificial Intelligence* will unfortunately never appear. It would have been authored by Professor David Marr, who died toward the end of 1980 after a protracted illness. The influence of the group which he founded at MIT is evident from the preceding pages.

David's background was in neurophysiology, after completing a mathematics degree at Cambridge University. His early work proposed mathematical theories of the neocortex, archicortex, and, perhaps best known of all, the cerebellum. He was to remain deeply commited to the study of human perception and memory for the rest of his life. In 1974 he was invited to spend a little time at the Artificial Intelligence Laboratory at MIT, and stayed for six years, eventually accepting a Professorship in the Department of Psychology. He quickly appreciated that computational concepts provided a further dimension for the expression of theories of human perception, and, together with a growing group of Ph.D. students, he set out to construct what he called a computational theory of human vision. The group has been enormously creative, publishing studies across the entire breadth of human vision.

David's work was notable in many ways, but in particular notable for its style. He made extensive reference to the literatures of neurophysiology and psychology, which were his background and to which he directed his contributions. In particular, he published in the journals which would be read by his intended audience, and encouraged his students to do so too. He argued for a mathematical analysis of a perceptual problem independent of, and prior to, consideration of issues concerning the choice of an algorithm. Under the heading of 'natural computation', he championed the isolation of the constraints which the world imposes upon perception, as well as the perceiver's prior beliefs about it. Though a good deal of Marr's work was mathematical in nature, its ramifications were stated in elegant prose. A book summarising his thoughts about human visual perception [37] and incorporating summaries of the contributions he and his colleagues have made across the entire range of the subject is currently in press.

David will be missed by the wide community of scholars whose work brought them in contact with his writing. He will be missed especially by those whose lives were enriched by knowing him or working with him.

ACKNOWLEDGMENT

This paper describes research done in part at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the Laboratory's Artificial Intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-75-C-0643. I thank Ellen Hildreth, Marilyn Matz, and Demetri Terzopoulos for their comments on a draft of this paper.