From: COMPUTER VISION AND SENSOR-BASED ROBOTS
Edited by George G. Dodd and Lothar Rossol
(Plenum Publishing Corporation, 1979)

ARTIFICIAL INTELLIGENCE AND THE SCIENCE OF IMAGE UNDERSTANDING

B. K. P. HORN

Massachusetts Institute of Technology, Cambridge, Massachusetts

ABSTRACT

Advanced automation promises to increase productivity, improve working conditions and assure product quality. Some computer-based systems perform tasks blindly, without elaborate sensor-based feedback. In many applications however visual input can speed up an automated system by eliminating search or the need for costly fixtures that maintain exact alignment of parts. In still other situations, many inspection jobs for example, there may be no alternative to machine vision.

Unfortunately, image analysis turned out not to involve just a simple extension of some well-known subfield of computer science or optics. A long history of frustrations with techniques borrowed from other domains mixed with clever special case solutions based on *ad hoc* techniques has brought the field to a point where it is finally considered worthy of serious attention. The foundations of a science of image understanding are beginning to be built on the ancestral paradigms of image processing, pattern recognition and scene analysis.

One component of this new thrust is an improved understanding of the physics of image formation. Understanding how the measurements obtained from the vision input device are determined by the lighting, shape and surface material of the objects being imaged helps one develop methods for "inverting" the imaging process, that is, exploit physical contraints to allow one to built internal symbolic descriptions of the scene being viewed. Another ingredient of the renewed optimism is a better understanding of the computations underlying early stages of the processing of visual information in biological systems. Aside from providing existence proofs that certain aspects of a scene can be understood from an image, this also suggests computational architectures for performing such tasks in real time.

A focal point of recent work is the choice of a representation for the objects being

References pp. 75-77

viewed and their internal prototypes. The internal descriptions must be tailored to expedite the computations involved in spatial reasoning. This has turned out to be a challenging new area, with problems and methods quite different from those found in the more serial, linguistic kind of reasoning we can introspect about and that artificial intelligence research has concerned itself with in the past.

Industrial problems, such as the visual determination of the position and orientation of a part in a bin provide new challenges for machine vision researchers. At the same time reductions in computational cost make some of the more complex techniques developed now feasible for industrial exploitation.

WHY USE MACHINE VISION TECHNIQUES?

Computer-based automation appears about to induce the next quantum jump in productivity, while also contributing to improved working conditions and higher product uniformity. The flexibility of the computer leads to greater complexity in systems and increased degrees of freedom in design. One important and difficult choice the designer has to face is whether or not to employ machine vision.

Many computer-based systems perform their tasks blindly, relying on accurate positioning and other methods of traditional "fixed" automation. The added complexity of machine vision will often pay off in improvements in speed by eliminating search or costly pallets used to hold parts in precise relationships to each other and the tools. In some existing systems people load parts into these fixtures, employing *their* vision system to determine the position and orientation of the incoming parts. A boring job indeed!

Besides, there are many cases where there is no alternative to some form of noncontact sensing. Many inspection tasks seem to require visual analysis of one kind or another for example. Sometimes a small number of sensors in strategically located areas will do; but more often than not an imaging system is called for together with the necessary electronic hardware and software for interpreting the image. The applications open to less sophisticated approaches appear to be limited.

WHAT IS MACHINE VISION?

An optical system forms an image of some three-dimensional arrangement of parts. The two-dimensional image is sensed and converted into machine readable format. It is the purpose of the machine vision system to derive information from this image useful in the execution of the given task. In the simplest case the information sought will concern only the location and orientation of an isolated object—more commonly, objects have to be recognized and their spatial relationships determined. This can be viewed as a process in which a description of the scene being viewed is developed from the raw image. The description has to be appropriate to the particular application. That is, irrelevant visual features should be discarded, while needed relationships between parts of objects must be deduced from their optical projection.

One of several reasons why this is non-trivial is that one dimension is lost in the imaging system. Depth has to be inferred from a variety of cues or determined by special techniques employing optical triangulation or time-of-flight measurements [1 - 3]. Fortunately our visual world is very special and it is possible to infer a great deal of information about the scene being imaged from one, or a small number of images. This is because in most cases one is observing opaque, cohesive objects immersed in a transparent medium. In this situation the entities of interest are the *surfaces* of the objects and surfaces are essentially two dimensional. (If *volumes* were imaged a form of tomography based on large numbers of images would be needed.)

Much still has to be learned about what one can reasonably expect a vision system to derive about a three-dimensional reality from an image, and what permanent properties of the surfaces can be calculated from the raw image intensity readings. It is clear that any attempt at recognition, matching or classification ought to be based on these estimates of the permanent properties rather than directly on the image intensities, which reflect these only indirectly.

ROOTS OF MACHINE VISION

Ideas from many related fields have contributed to early progress in machine vision. We are all familiar for example with the work on recognition of printed text. Unfortunately many methods developed in such specialized areas do not generalize smoothly to more complicated situations. Characters for example appear as fixed, two dimensional patterns in essentially only two values of gray. Images of machined parts on the other hand contain many levels of gray, call for high resolution and produce image intensity "patterns" that not only vary with the attitude of the part in space, but depend on the distribution of light-sources. Specular reflections, gloss and mutual illumination further complicate the picture. Nevertheless, machine vision has built, in part, on ideas from three related fields: image processing, pattern recognition, and scene analysis.

Image processing [4 - 10] concerns itself with the production of new images from existing images. Usually these new images are obtained by application of a technique from linear systems theory and are enhanced in some fashion so as to improve their appearance to a human viewer. In machine vision we are more interested in the generation of symbolic descriptions and the use of these descriptions to permit a computer controlled system take appropriate action. Nevertheless, some of the techniques developed for processing raw image intensity values are of importance.

Pattern classification [11 - 17] deals with the mapping of feature vectors, containing measurements of objects, into class numbers. Any connection with our discussion here seems remote until it is remembered that frequently the components of the feature vector represent measurements obtained from parts of an image. Useful segmentation methods and simple measures of shape have been developed for this purpose, although many apply only to patterns of an essentially two-dimensional nature.

Scene analysis [18 - 20], finally, is largely pre-occupied with the transformation of descriptions into more abstract descriptions. When faced with the problem of vision in a world of toy blocks for example, the transformation of a line-drawing into a description in terms of three dimensional solids and how they relate spatially would be a typical task for a scene analysis system [21 - 28]. Developing the line drawing from the raw image in the first place would be a job left to some other subsystem [29 - 31]. Similar comments can be made regarding the complementary approach of region growing [32 - 34]. The methods that were developed for manipulating descriptions and relating them to known information about the world being imaged have proved invaluable to workers in the fields of machine vision.

DIFFICULTIES AND SUCCESSES

Something important has been learned: vision is difficult. This is surprising because we find it hard to believe that something that is so immediate to us could be hard. If computers can interpret mass spectrograms, perform complicated manipulations on symbolic mathematical expressions and even perform medical diagnosis in limited disease domains [20], why can't they see? Many challenging problems are succumbing to the methods being developed in artificial intelligence, yet some have turned out to be surprisingly resistant to concerted attacks. While computers become expert problem solvers in areas such as electronic circuit analysis, less impressive progress has been made on the problem of common sense reasoning, for example.

All this means is that we don't have a good idea of how difficult tasks are that we cannot introspect about. Vision is one such task. As we learn more about biological vision systems we also become more aware of the prodigous amount of computing power necessary. At this stage then we have to be very clever in our use of existing computer hardware or depend on *ad hoc* solutions applicable only in limited domains. Underestimation of the difficulties of vision led to many attempts, foolish in retrospect, to apply simple methods from other fields such as linear systems analysis, information theory or statistics.

We now know that the field demands its own methodology. In view of what has been said, it is impressive to note that machine vision has already had a significant impact. Several systems incorporating machine vision techniques have progressed beyond the laboratory demonstration stage. Systems for positioning and orienting integrated circuit chips are an example. The major cost component of a *simple* integrated circuit (I.C.) is the packaging, and the major component of that used to be the manual alignment of each I.C. to prepare it for lead bonding. The visual alignment of these chips is a fatiguing, uninteresting task possibly even harmful to the operator's vision.

As a result of pioneering work at the M.I.T. Artificial Intelligence Laboratory [35], General Motors Research Laboratory [36], Hitachi Central Research Laboratory [37] and Texas Instruments this task is now in many cases performed by a computer vision system. Curiously, these systems, developed independently, use quite different techniques for processing the raw image intensities, reflecting to some extent different target components, but also the richness of the storehouse of methods that is beginning to develop in this field. One of these systems for example depends on a careful analysis of one-dimensional profiles of intensity along strategically chosen lines in the image plane. The requisite speed and robustness is obtained in this case without recourse to special-purpose, parallel hardware. Other systems inspect printed circuit cards [38], visually guide a machine that tightens bolts on telephone-pole molds [39] and direct manipulators to pick isolated objects off a conveyor belt [40]. Recently, manufacturers of ''industrial robots'' have introduced computer-control for their products. This permits them to be programmed in more flexible ways. The use of global, rectangular coordinate systems made possible by this advance prepares these devices for input from machine vision systems such as the ones described here.

CURRENT TRENDS AND BASIC ISSUES

Many aspects of machine vision are currently being explored. I can only focus on a few issues that seem important to me. One new approach stems from a view of the machine vision system as a device for ''inverting'' the imaging process—that is, developing a description of the scene being viewed from an image. It seems reasonable to suppose that understanding the physics of the imaging process would be helpful in this endeavor, since this determines the ''forward'' transformation [41 - 46]. What has to be understood is how the measurements obtained from the vision input device are determined by the lighting, shape and surface material of the objects being imaged. Such understanding will permit the exploitation of physical constraints to allow one to build internal symbolic descriptions of the scene based on information about the physical properties of the surfaces rather than raw image intensities. It is possible for example to calculate the shape of a smooth object from a single image if enough is known about the surface reflectance properties and the distribution of light-sources [41, 42].

An extension, of interest in the case of industrial application of machine vision, depends on multiple images obtained from the same viewpoint under different lighting conditions [47-49]. In certain cases this can lead to a simple look-up table computation of local surface orientation and surface ''albedo'', thus producing just the information needed to determine the attitude of a body in space. Thus these ''photometric stereo'' methods may play a role in solving the ''bin-of-parts'' problem [50]. Fortunately the task of understanding the physics of light reflection and imaging has been made simpler by the recent introduction of a systematic formalism for the description of the reflectance properties of surfaces by the National Bureau of Standards [51].

Progress is also being made in developing better models for biological vision systems [51-54]. The central thrust of the new approach is the use of computers in evaluating competing theories with less reliance on raw neurophysiological and experimental psychology data. Instead the emphasis is on what information can be extracted from the image and what representations are appropriate for various intermediate forms of the symbolic descriptions. Suggestions for computer architectures to support the huge amounts of computation needed also may result from this work.

References pp. 75-77

An important issue concerns the appropriate representation for objects and space. This is particularly important if a computer controlled manipulator is to interact with the objects being viewed. Representations of objects in terms of mathematical equations defining surface patches, a common method in computer graphics, has not found much of a following here, where it may be necessary to reason about objects in terms of the space they occupy. Considerable progress has been made in developing representations in terms of volumes [55 - 62]. In fact, the whole question of spatial reasoning is only now drawing attention. We are once more surprised at the computational complexities involved, since we cannot easily introspect into our own ability to think about objects and space.

A representation for surfaces of objects, that arises naturally from the form of the output of the photometric methods described above, is one in terms of local surface normals. This method of capturing the information regarding the shape of the objects leads to a representation that transforms more easily with rotation than more obvious approaches. One can think of the object as if covered with ''spines''. A mapping of these spines onto the sphere of possible directions often uniquely specifies the particular object and may lead to methods for determining its attitude in space. A similar representation has been recently proposed as an intermediate form between the very crude symbolic description computed directly from the image and the more elaborate ones in terms of volumes described above [54 - 59].

The use of prior knowledge about the objects being viewed is a thorny issue. Where should this knowledge be used and how should it be represented in the computer? On the one side one finds advocates of a ''non-purposive'' approach, where one computes as detailed a description as possible based on the image, independent of the overall task at hand. Others would prefer to use the image only to verify hypotheses generated by an analysis of prior knowledge about the likely arrangement of objects in the scene. This leads to a paradigm of ''controlled hallucination''. Probably, future systems will embody a compromise, where simple operators are applied over the complete image using massive parallelism. These operations would not be influenced much by the current purpose of the overall system. The crude symbolic description computed this way would then be used by processes more goal-directed and more appropriate to the task at hand.

Other recent advances have been made in the processing of multiple images of the same scene taken from different viewpoints [63 - 67], and the use of parallel ''relaxation'' computations that process images by iteration much as numerical methods for solving elliptic partial differential equations [68, 69].

A SCIENCE OF IMAGE UNDERSTANDING

A better understanding of image formation is leading to a better exploitation of physical constraints that allow interpretation of the two-dimensional image in terms of a three-dimensional reality. Simultaneously the success of special purpose systems is supporting a more serious interest in vision problems. The simple-minded application of assorted methods borrowed from other fields has been discredited. We no longer underestimate the problems and are resigned to work hard to get solutions implemented that will be efficient enough to operate successfully on present-day hardware. A science of image understanding is beginning to emerge.

REFERENCES

1. Nitzan, D., Brain, A. E., and Duda, R. O. The measurement and use of registered reflectance and range data in scene analysis. *Proc. IEEE* 65, 2 (1977), 206-220.
2. Nevatia, R. Depth measurement by motion stereo. *Computer Graphics and Image Processing* 5, 2 (1976), 203-214.
3. Shirai, Y. Recognition of polyhedrons with a range finder. *Bul. Electro-technical Lab.* 35, 3 (1971), 209-296.
4. Andrews, H. C. *Computer Techniques in Image Processing.* Academic Press, New York, 1970.
5. Gonzalez, R. C., and Wintz, P. *Digital Image Processing.* Addison-Wesley, Reading, MA, 1977.
6. Huang, T. S. (Ed.) *Picture Processing and Digital Filtering.* Springer, NY, 1975.
7. Lipkin, B. S., and Rosenfeld, A. (Eds.) *Picture Processing and Psychopictorics.* Academic Press, New York, 1970.
8. Rosenfeld, A. *Picture Processing by Computer.* Academic Press, New York, 1969.
9. Rosenfeld, A. (Ed.) *Digital Picture Analysis.* Springer, NY, 1976.
10. Rosenfeld, A., and Kak, A. C. *Digital Picture Processing.* Academic Press, New York, 1976.
11. Cheng, G. C., et al. (Eds.) *Pictorial Pattern Recognition.* Thompson Book Co., Washington, DC, 1968.
12. Fu, K. S. *Syntactic Methods in Pattern Recognition.* Academic Press, New York, 1974.
13. Fu, K. S. (Ed.) *Digital Pattern Recognition.* Springer, NY, 1976.
14. Grasselli, A. (Ed.) *Automatic Interpretation and Classification of Images.* Academic Press, New York, 1969.
15. Tou, J. T., and Gonzalez, R. C. *Pattern Recognition Principles.* Addison-Wesley, Reading, PA, 1974.
16. Watanabe, S. *Methodologies of Pattern Recognition.* Academic Press, New York, 1969.
17. Duda, R. O., and Hart, P. E. *Pattern Classification and Scene Analysis.* John Wiley and Sons, New York, 1973.
18. Hanson, A., and Riseman, E. (Eds.) *Computer Vision Systems.* Academic Press, New York, 1979.
19. Winston, P. H. (Ed.) *The Psychology of Computer Vision.* McGraw-Hill, New York, 1975.
20. Winston, P. H. *Artificial Intelligence.* Addison-Wesley, Reading, PA, 1977.
21. Clowes, M. B. On seeing things. *Artificial Intelligence* 2, 1 (1971), 79-112.
22. Falk, G. Interpretation of imperfect line data as a three-dimensional scene. *Artificial Intelligence* 3, 2 (1972), 101-144.
23. Grape, G. R. Model based (intermediate level) computer vision. AI Memo 201, Stanford U., 1973.
24. Huffman, D. A. Impossible objects as nonsense sentences. in *Machine Intelligence* 6, B. Meltzer and D. Mitchie (Eds.) Edinburgh University Press, Edinburgh, 1971.
25. Mackworth, A. K. Interpreting pictures of polyhedral scenes. *Artificial Intelligence* 4, 2 (1973), 121-138.
26. Roberts, L. G. Machine perception of three-dimensional solids. in *Optical and Electro-optical Information Processing,* J. T. Tippet, et al (Eds.) MIT Press, Cambridge, MA, 1965.
27. Waltz, D. L. Understanding line drawings of scenes with shadows. In *The Psychology of Computer Vision,* P. H. Winston (Ed.), McGraw Hill, New York, 1975.

28. Winston, P.H. The M.I.T. robot. In *Machine Intelligence 7.* Edinburgh University Press, Edinburgh, 1972.

29. Griffith, A.K. Computer recognition of prismatic solids. MIT Project MAC, TR-73, MIT, Cambridge, MA, 1970.

30. Horn, B.K.P. The Binford-Horn line-folder. MIT AI Memo 285, MIT, Cambridge, MA, 1971.

31. Shirai, Y. Analyzing intensity arrays using knowledge about scenes. In *The Psychology of Computer Vision,* P.H. Winston, (Ed.), McGraw Hill, New York, 1975.

32. Brice, C., and Fenema, C. Scene analysis using regions. *Artificial Intelligence* 1, 3 (1970), 205-226.

33. Ohlander, R.B. Analysis of natural scenes. Ph.D. Th., Dept. of Computer Science, Carnegie-Mellon U., 1975.

34. Tenebaum, J.M., and Barrow, H.G. Experiments in interpretation-guided segmentation. *Artificial Intelligence* 8, 3 (1977), 241-274.

35. Horn, B.K.P. A problem in computer vision: orienting silicon integrated circuit chips for lead bonding. *Computer Graphics and Image Processing* 4, 3 (1975), 294-303.

36. Baird, M.L. An application of computer vision to automated IC chip manufacture. GMR-2124, General Motors Res. Labs., Warren, MI, 1976.

37. Kashioka, S., Ejiri, M., and Sakamoto, Y. A transistor wire-bonding system utilizing multiple local pattern matching techniques. *IEEE Trans. on Systems, Man, and Cybernetics* SMC-6, 8 (1976), 562-570.

38. Ejiri, M., Uno, T., Mese, M., and Ikeda, S. A process for detecting defects in complicated patterns. *Computer Graphics and Image Processing* 2 (1973), 326-339.

39. Uno, T., Ejiri, M., and Tokunaga, T. A method of real-time recognition of moving objects and its application. *Pattern Recognition* 8 (1976), 201-208.

40. Holland, S.W., Rossol, L., and Ward, M.R. CONSIGHT-I: a vision-controlled robot system for transferring parts from belt conveyors. GMR-2790, General Motors Res. Labs., Warren, MI, 1978.

41. Horn, B.K.P. Obtaining shape from shading information. In *The Psychology of Computer Vision,* P.H. Winston (Ed.), McGraw-Hill, New York, 1975.

42. Horn, B.K.P. Understanding image intensities. *Artificial Intelligence* 21, 11 (1977), 201-231.

43. Barrow, H.G., and Tenenbaum, J.M. Recovering intrinsic scene characteristics from images. In *Computer Vision Systems,* A. Hanson and E. Riseman (Eds.), Academic Press, New York, 1979.

44. Woodham, R.J. A cooperative algorithm for determining surface orientation from a single view. Proc. 5th Int. Jt. Conf. on Artificial Intelligence (1977), 635-641.

45. Woodham, R.J. Reflectance map techniques for analyzing surface defects in metal castings. TR-457, AI Lab., MIT, Cambridge, MA, 1978.

46. Horn, B.K.P., and Bachman, B.L. Using synthetic images to register real images with surface models. *Comm. ACM* 21, 11 (1978), 914-924.

47. Woodham, R.J. Photometric stereo: a reflectance map technique for determining surface orientation from image intensity. Proc. SPIE 22nd Ann. Tech. Symp. 155 (1978).

48. Woodham, R.J. Photometric stereo. AI Memo 479, MIT, Cambridge, MA, 1978.

49. Horn, B.K.P., and Woodham, R.J. Determining shape and reflectance using multiple images. AI Memo 485, MIT, Cambridge, MA, 1978.

50. Birk, J., Kelley, R., Wilson, L., Badami, V., Brownell, T., Chen, N., Duncan, D., Hall, J., Martins, H., Silva, R., and Tella, R. General methods to enable robots with vision to acquire, orient and transport workpieces. Fourth Rep., GRANT APR74-13935, U. of Rhode Island, Kingston, RI, 1978.

51. Nicodemus, F.E., Richmond, J.C., and Hsia, J.J. Geometrical considerations and nomenclature for reflectance. NBS Monograph 160, Nat. Bur. Standards, Washington, DC, 1977.

52. Marr, D. Early processing of visual information. *Phil. Trans. Roy. Soc.* B 275 (1976), 483-524.

53. Marr, D., and Poggio, T. A theory of human stereo vision. AI Memo 451, MIT, Cambridge, MA, 1977.

54. Marr, D. Representing visual information. In *Computer Vision Systems,* A. Hanson and E. Riseman (Eds.), Academic Press, New York, 1979.

55. Agin, C. J., and Binford, T.O. Computer description of curved objects. Proc. 3rd Int. Jt. Conf. on Artificial Intelligence (1973), 629-640.

56. Binford, T.O. Visual perception by computer. Proc. IFIP Conf. (1971).

57. Binford, T.O. Visual perception by computer. IEEE Conf. on Systems and Control (1971).

58. Hollerbach, J. Hierarchical shape description of objects by selection and modification prototypes. TR-346, AI Lab., MIT, Cambridge, MA, 1976.

59. Marr, D., and Nishihara, H.K. Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. Roy. Soc. London* B 200 (1977), 269-294.

60. Nevatia, R. Structured descriptions of complex curved surfaces and visual memory. AI Memo 250, Stanford U., 1974.

61. Nevatia, R., and Binford, T.O. Description and recognition of curved objects. *Artificial Intelligence* 8, 1 (1977), 77-98.

62. Soroka, B.I., and Bajcsy, R.K. Generalized cylinders from serial sections. Proc. Int. Jt. Conf. on Pattern Recognition (1976), 734-735.

63. Arnold, R.D. Local context in matching edges for stereo vision. Proc. DARPA Image Understanding Workshop, L. Baumann (Ed.), Science Applications, Inc., 1978.

64. Gennery, D.B. A stereo vision system for an autonomous vehicle. Proc. 5th Int. Jt. Conf. on Artificial Intelligence (1977), 576-582.

65. Marr, D. A note on the computation of binocular disparity in a symbolic low-level visual processor. AI Memo 327, MIT, Cambridge, MA, 1974.

66. Marr, D., and Poggio, T. Cooperative computation of stereo disparity. *Science* 194 (1976), 283-287.

67. Quam, L.H. Computer comparison of pictures. AI Memo 144, Stanford U., 1971.

68. Rosenfeld, A. Iterative methods in image analysis. Proc. IEEE Conf. on Pattern Recognition and Image Processing (1977), 14-18.

69. Rosenfeld, A., Hummel, R.A., and Zucker, S.W. Scene labelling by relaxation operations. *IEEE Trans. on Systems, Man and Cybernetics* SMC-6 (1976), 420-433.