1

INTRODUCTION

Machine perception is a field of endeavor attempting to enable man-made machines to perceive their environment by sensory means as humans and animals do. Biological systems sense their environments by a variety of sources such as sight, sound, touch and smell. The process of perception involves making useful models of the environment from a confusing mass of sensory input data.

Visual data are the most complex and most useful sensory input for humans. Visual machine perception is concerned with the machine interpretation of similar visual data. There are many similarities in the ways in which different sensory data are perceived; in this book we will treat visual perception only.

Machines that perceive their environments and perform required tasks have an obvious usefulness for diverse application areas such as industrial assembly and inspection, planetary space exploration, automated medical x-ray screening, monitoring of earth resources by remote sensors, and a variety of military applications. They could assist in many tasks that are routine, tedious and even dangerous for humans to perform but are difficult or impossible to automate without some perceptual ability.

Most industrial assembly tasks, including seemingly simple tasks such as mounting of wheels on an automobile, normally require use of vision (however, blind people do learn to perform some of these tasks effectively by using touch). In hostile environments, such as outer space and undersea, or in the handling of hazardous materials, the use of machines may be essential. For many of these applications, machines need to be autonomous. As an example, for a vehicle to explore the surface of the planet Mars, the long delay in signal transmissions to earth (8 to 30 minutes round trip) virtually rules out human control from earth (by using "tele-operator" systems). Advantages of earthbound self-guided vehicles are also clear for transportation as well as military purposes.

Another important application area is the interpretation of images taken from aircraft or satellites for the monitoring of earth resources, weather patterns, and military surveillance. These tasks are tedious for humans to perform, and the manual interpretation process is a major bottleneck in the total processing system. Also, owing to the high volume of data, on board processing is highly desirable.

The above examples suggest some of the ambitious goals of the field of visual machine perception. Machine perception is a part of the larger field of *artificial intelligence*, which aims at building machines that behave intelligently, in perceptual and other domains. Visual imagery plays a prominent role in human intelligence for diverse tasks, and it is believed that understanding of machine perception will aid in the development of other forms of machine intelligence as well.

1.1 SOME EXAMPLES AND PROBLEMS

Visual perception is so immediate and effortless for us that the noninitiated imagine little difficulty in automating this process. Isn't the process of recognizing an object in an image merely a matter of finding a particular pattern of light intensity seen previously? A little thought, however, reveals numerous traps and complexities. A few simple examples will expose the basic problems.

First, consider the problem of recognizing isolated characters of the English alphabet. A character may be recognized by comparing it to those in the alphabet and choosing the "closest" one. This method will work if the characters are always of the same size and occur in the same orientation and the images contain little noise. But what of changes in scale and orientation, as shown in Fig. 1-1? To meet such situations we may postulate "shape" of a character, say "A," as consisting of two equal length line segments meeting at a point and at an angle of between 20 and 60 degrees, joined by a horizontal bar close to halfway between them. Assuming that such *descriptions* can be extracted from an image, the problems of change in scale and orientation seem to be in hand.

However, the characters in Fig. 1-2 are also likely to be perceived by humans as the letter "A," even though the reader may not have seen these forms before and they do not fit the above description of "A." Clearly, the *representation* of the pattern of the letter "A" is a rather complex and important issue.

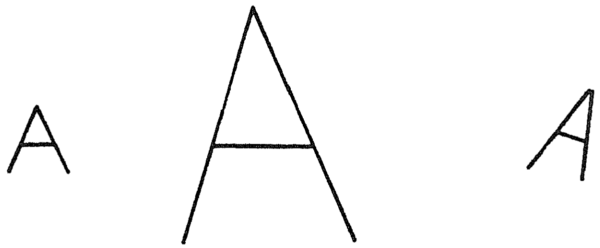


Figure 1-1: Character "A" in different sizes and orientations

Figure 1-2: Different ways of drawing the same character

The words shown in Fig. 1-3 are easily perceived to be "THE CAT." However, the letter representing "H" is printed in the same manner as that representing "A." It is clear that context and our expectations mediate such perception.

Figure 1-3: An example to illustrate context dependency in character recognition

Now, consider the more complex example shown in Fig. 1-4. There is likely to be little disagreement that the figure is a drawing representing a cube or a box. However, the figure is inherently ambiguous. In the extreme, the figure is just a pattern of lines forming connected parallelograms in a plane. Even if a three-dimensional interpretation is required, there is no direct evidence that the "hidden" part of the surface is that of a cube. The choice of a suitable representation is even more important now.

Figure 1-5 shows a line drawing of one box-shaped object behind

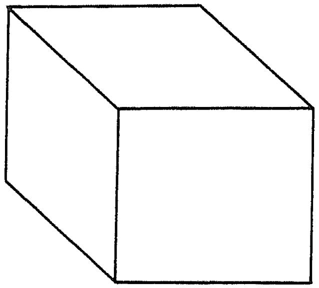


Figure 1-4: Line drawing of a cube

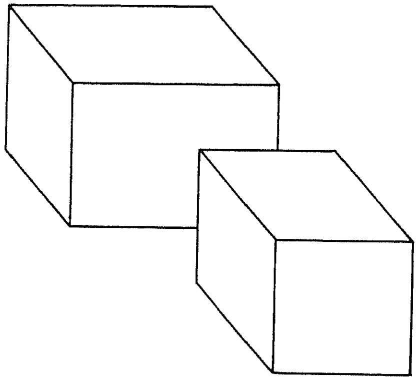
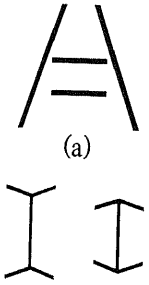


Figure 1-5: Line drawing of two objects



(a)

(b)

Figure 1-6: Two illusions

another. We conclude that the figure contains two objects, even though the lines defining them are connected. The distant object does not seem to be ambiguous even though not all of the "front" surface is visible.

These examples point out the difficulties caused by perception of a three-dimensional world through its two-dimensional projections on an imaging device. Most acute is the occlusion of objects by one another or self-occlusion of parts of their own surfaces. Such occlusion is common among scenes viewed at close range.

Human perception of figures, though remarkably accurate, is hardly perfect. Figures 1-6 (a) and (b) show examples of lines of equal length which are perceived unevenly. These "illusions" can be explained as being due to three-dimensional interpretations of the two-dimensional patterns.

Images of practical interest and of our everyday experience are rather complex. Figure 1-7 shows an image of some industrial parts; Fig. 1-8 shows a typical chest x-ray; Fig. 1-9 shows an aerial outdoor image (of the San Francisco area). In all three examples the major features, such as the parts in the industrial scene, the ribs in the x-ray, and the bay, roads, urban and suburban areas, and hills in the aerial image, are likely to be perceived immediately and with high confidence, even if the reader is unfamiliar with some of the objects in the scene or with the geography of the location of the aerial image. To analyze pictures of such complexity, we need to extract subtle and complex pieces of information from the pictures and efficiently utilize our store of knowledge about the structures and constraints on the objects in our perceptual world.

1.2 AN OVERVIEW AND HISTORY

Two major approaches to pattern perception have evolved: (1) those of mathematical pattern recognition and (2) the descriptive techniques of scene analysis. The first concentrates on the problem of assigning a pattern, not necessarily visual, to one of the known classes of patterns, based on some measurements, known as features, computed from the pattern. Studies in this approach have concentrated on classification methods and their generality rather than on defining useful features.

Scene analysis methods attempt to describe a pattern in terms of simpler primitives extracted from the input, and recognition is by matching of their descriptions. The emphasis is on the choice of the primitives and their descriptions. Utilization of known constraints in the image formation process is encouraged. The techniques developed

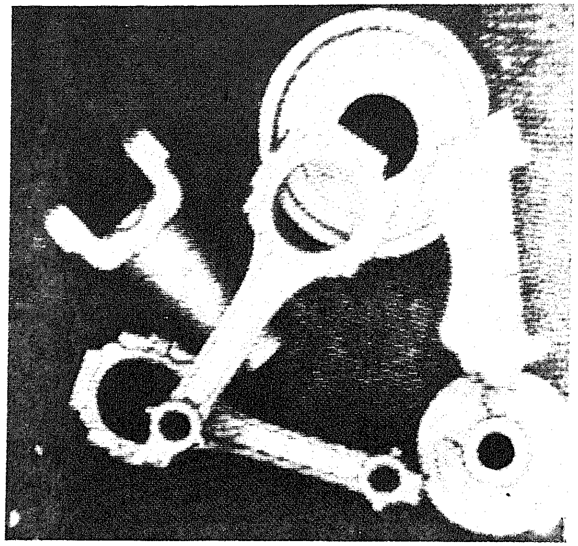


Figure 1-7: Picture of some industrial parts (from Perkins [1])

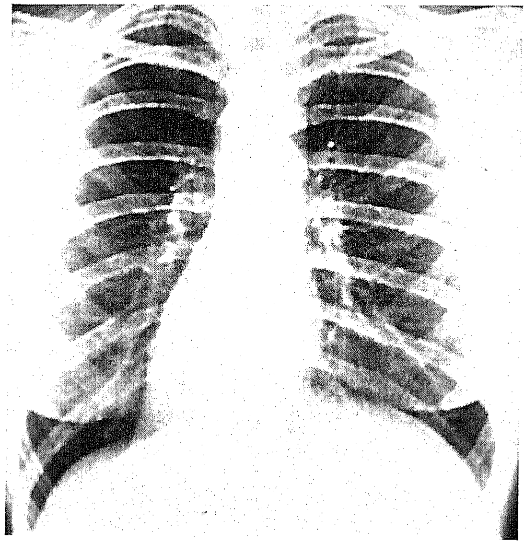


Figure 1-8: A chest x-ray

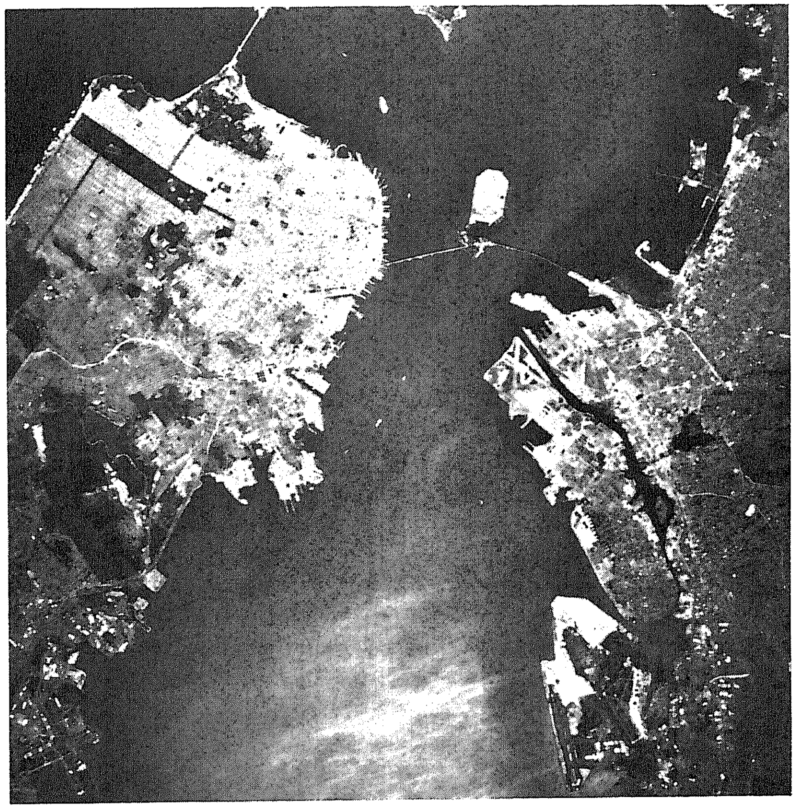


Figure 1-9: An image of the San Francisco area

often tend to follow those thought to be used by humans, though this is not an essential goal. This book is about these descriptive techniques; pattern recognition techniques are discussed only briefly for comparison (in Chapter 2).

The field of three-dimensional scene analysis originated, in the early 1960s with the pioneering work of L. G. Roberts [2]. He considered the problem of recognizing simple 3-D polyhedral objects from a single view. His system extracted a line drawing and its description from a digitized TV image of a scene. Objects were recognized by matching against computed projections of known objects. In computation of such projections, Roberts also originated the field of three-dimensional computer graphics. His system handled objects with a limited amount of occlusion.

For the next several years, for presumed reasons of simplicity, attention remained focused on scenes consisting of polyhedral objects. The central problem was recognized to be that of *segmentation*—that is, determining which lines and surfaces belonged to a single object. This domain is now well understood, and techniques exist to segment complex scenes of polyhedra, *if* complete and perfect line drawings are available.

In experiments with real images of polyhedra, it was realized that perfect line drawings were rarely extracted; consequently, some techniques that work with incorrect or missing data were developed. It was also realized that a partial analysis of line drawings may provide the essential context for extracting other weaker lines. Visual processing was now thought to proceed at many levels of abstraction, with results of higher level processing sometimes guiding the processing at lower levels in a feedback arrangement. The term *heterarchy*, as contrasted with *hierarchy*, has been applied to such a scheme.

Building on the experience with the so-called *blocks world*, further research has explored many divergent paths. A few attempts were made to extend the techniques of the blocks world to scenes of more general objects, including curved objects. For polyhedra, topological properties of lines and vertices often suffice for their recognition. More powerful descriptions of shape, necessary for curved objects, were developed. Experience with complex, natural outdoor and indoor scenes indicated that the contextual relationships of different objects to each other are extremely useful, if not essential, in their analysis.

Another line of research has been to build special purpose systems for particular applications, taking advantage of the a priori knowledge of the limited environment to simplify the processings. Proper organization and utilization of such knowledge are the key issues for such systems. A wide variety of application areas have been investigated such as industrial automation, biomedical image processing, and interpretation of aerial images. Some typical scenes are shown in Figs. 1-7, 1-8, and 1-9.

Concurrent with the work on *scene analysis* has been work on *image analysis.* Image analysis is concerned with techniques for extracting descriptions from images that are necessary for higher-level scene-analysis methods. The image analysis techniques include computation of perceived brightness and color, partial or complete recovery of three-dimensional data in the scene, location of discontinuities corresponding to objects in the scene, and characterization of the properties of uniform regions in the image.

The capabilities of machines to see are far less than human capabilities, the fantasies of the science fiction writers notwithstanding, and many fundamental issues of processing and representation are unresolved. However, the known techniques are adequate for building machines with limited, but useful, abilities to see for applications to real problems. Experimental and production machines exist to inspect printed circuit boards for defects and to orient integrated circuit chips for lead bonding, and many machines perform interesting assembly tasks in laboratory environments. A prototype of a Mars rover that functions in a highly simplified environment has been built. Use of image-based surveillance for earth resource monitoring and guidance for military vehicles is growing.

1.3 RELATIONSHIP TO PSYCHOLOGY

The problems of machine perception are analogous to those of human perception that have been studied for a long time by psychologists. A student of either field will benefit from studying the literature of the other. The known results of psychologica l research as well as insights gained by introspection have guided much of the research in machine perception, even though the motivation is not necessarily to simulate human perception. Rather, human (animal) perception is viewed as the only known example of a perceiving system.

The psychological research is not advanced to the extent of providing many concrete models that can be directly programmed for a machine. More is known about the early processing of signals at or near the eye than in in the cortex of the brain. Digital computers provide a rigorous way to test the predictions of psychological theories, and research in machine perception should suggest further psychology experiments. In the near term, the major benefit to machine perception may come from the wealth of experimental data about human visual perception, and its limitations, that is available in the psychology literature.

No attempt is made in this book to relate machine perception techniques in a formal way with psychological models, but analogies with human perception are cited informally. Introductory material about human visual perception may be found in [3-6].

1.4 PLAN OF THE BOOK

The organization of the book follows the overview in Section 1.2. The ordering of material is neither chronological nor hierarchical in levels of processing. Pattern classification techniques are discussed only briefly in Chapter 2, mostly for purposes of comparison; several textbooks are available on this subject [7-9].

The higher level scene analysis techniques for scenes of polyhedral objects are presented in Chapters 3 and 4, followed by a discussion of techniques for nonpolyhedral objects in Chapter 5. Lower-level image analysis techniques are presented in Chapters 6 through 9, which discuss the perception of brightness, color, and depth as well as low-level segmentation techniques. The high-level scene analysis techniques are presented first to give the reader a view of the goals and the desired results of the lower level image analysis. Complete systems and application areas are described in Chapter 10.

Each chapter of the book contains a bibliography of historically important papers and representative papers of the major approaches. More complete and current references may be found in [10].

This book excludes discussion of a field commonly known as *image processing,* which is concerned with the processing of images to produce new images *for human viewing.* Typical applications are enhancement or restoration of degraded images and bandwidth reduction for image transmission. These subjects are covered in several textbooks [11-15]. (Some of these books also contain material overlapping the subject of this book.) To draw a distinction from image processing, the term *image understanding* is also used synomously with visual machine perception.

REFERENCES

[1] W. A. Perkins, "A Model-Based Vision System for Industrial Parts," General Motors Research Laboratories Report GMR-2410, 1977.

[2] L. G. Roberts, "Machine Perception of Three-Dimensional Solids," in *Optical and Electro-Optical Information Processing*, J. T. Tippett, et al. (eds.), MIT Press, Cambridge, Mass., 1963, pp. 159-197.

[3] R. L. Gregory, *The Intelligent Eye*, McGraw-Hill Paperbacks, New York, 1970.

[4] P. H. Lindsey and D. A. Norman, *Human Information Processing*, Academic Press, New York, 1972.

[5] T. N. Cornsweet, *Visual Perception*, Academic Press, New York, 1970.

[6] I. Rock, *Introduction to Perception*, Macmillan, New York, 1975.

[7] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, New York, 1973.

[8] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, New York, 1972.

[9] H. C. Andrews, *Introduction to Mathematical Techniques in Pattern Recognition*, John Wiley & Sons, New York, 1972.

[10] A. Rosenfeld, "Picture Processing," published annually in *Computer Graphics and Image Processing*, since 1975.

[11] W. K. Pratt, *Digital Image Processing*, John Wiley & Sons, New York, 1978.

[12] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, Academic Press, New York, 1976.

[13] K. R. Castleman, *Image Processing*, Prentice-Hall, Englewood Cliffs, N.J., 1979.

[14] E. L. Hall, *Computer Image Processing and Recognition*, Academic Press, New York, 1979.

[15] R. C. Gonzales and P. A. Wintz, *Digital Image Processing*, Addison-Wesley, Reading, Mass., 1977.