Robot Vision

Book · January 1986
Source: DBLP

1 author:

Berthold K. P. Horn
Massachusetts Institute of Technology

1

Introduction

In this chapter we discuss what a machine vision system is, and what tasks it is suited for. We also explore the relationship of machine vision to other fields that provide techniques for processing images or symbolic descriptions of images. Finally, we introduce the particular view of machine vision exploited in this text and outline the contents of subsequent chapters.

1.1 Machine Vision

Vision is our most powerful sense. It provides us with a remarkable amount of information about our surroundings and enables us to interact intelligently with the environment, all without direct physical contact. Through it we learn the positions and identities of objects and the relationships between them, and we are at a considerable disadvantage if we are deprived of this sense. It is no wonder that attempts have been made to give machines a sense of vision almost since the time that digital computers first became generally available.

Vision is also our most complicated sense. The knowledge we have accumulated about how biological vision systems operate is still fragmentary and confined mostly to the processing stages directly concerned with signals from the sensors. What we do know is that biological vision systems are complex. It is not surprising, then, that many attempts to provide machines with a sense of vision have ended in failure. Significant progress has been made nevertheless, and today one can find vision systems that successfully deal with a variable environment as parts of machines.

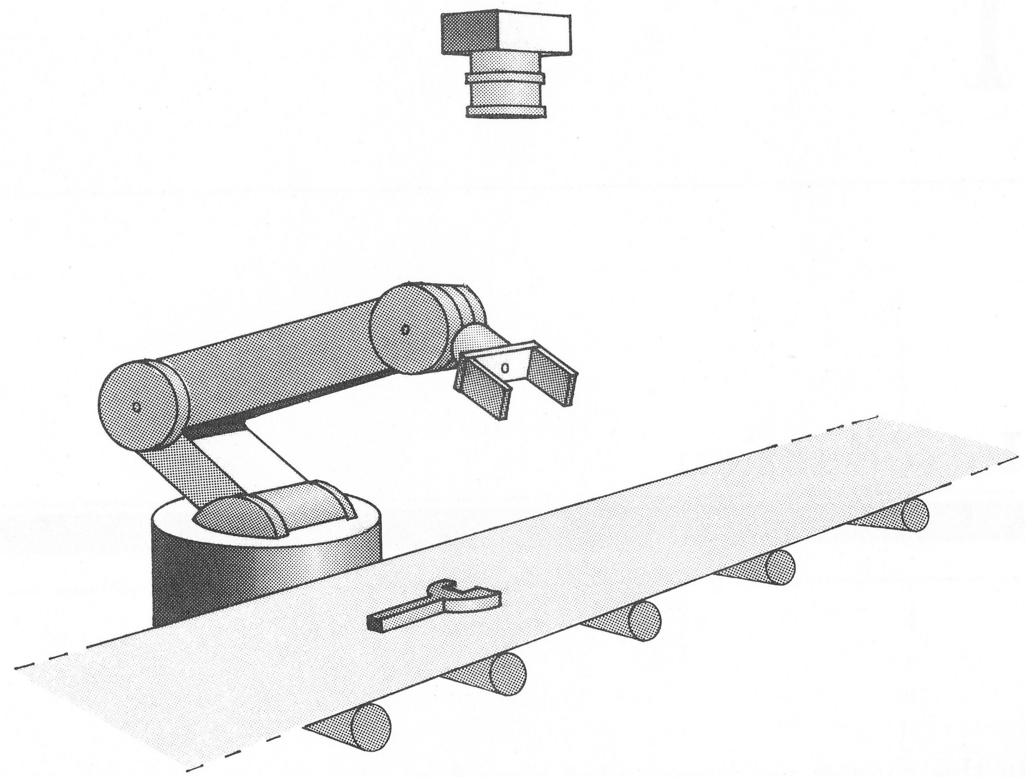

Figure 1-1. A machine vision system can make a robot manipulator much more versatile by allowing it to deal with variations in part position and orientation. In some cases simple binary image-processing systems are adequate for this purpose.

Most progress has been made in industrial applications, where the visual environment can be controlled and the task faced by the machine vision system is clear-cut. A typical example would be a vision system used to direct a robot arm to pick parts off a conveyor belt (figure 1-1).

Less progress has been made in those areas where computers have been called upon to extract ill-defined information from images that even people find hard to interpret. This applies particularly to images derived by other than the usual optical means in the visual spectrum. A typical example of such a task is the interpretation of X-rays of the human lung.

It is of the nature of research in a difficult area that some early ideas have to be abandoned and new concepts introduced as time passes. While frustrating at times, it is part of the excitement of the search for solutions. Some believed, for example, that understanding the image-formation process was not required. Others became too enamored of specific computing methods of rather narrow utility. No doubt some of the ideas presented here will also be revised or abandoned in due course. The field is evolving too rapidly for it to be otherwise.

We cannot at this stage build a "universal" vision system. Instead, we address ourselves either to systems that perform a particular task in a controlled environment or to modules that could eventually become part of a general-purpose system. Naturally, we must also be sensitive to practical considerations of speed and cost. Because of the enormous volume of data and the nature of the computations required, it is often difficult to reach a satisfactory compromise between these factors.

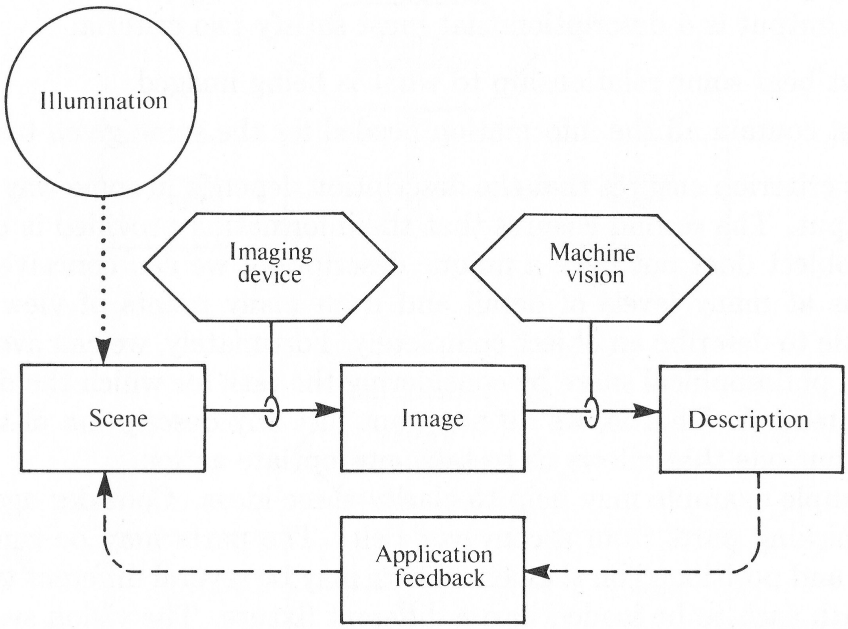

Figure 1-2. The purpose of a machine vision system is to produce a symbolic description of what is being imaged. This description may then be used to direct the interaction of a robotic system with its environment. In some sense, the vision system's task can be viewed as an inversion of the imaging process.

1.2 Tasks for a Machine Vision System

A machine vision system analyzes images and produces descriptions of what is imaged (figure 1.2). These descriptions must capture the aspects of the objects being imaged that are useful in carrying out some task. Thus we consider the machine vision system as part of a larger entity that interacts with the environment. The vision system can be considered an element of a feedback loop that is concerned with sensing, while other elements are dedicated to decision making and the implementation of these decisions.

The input to the machine vision system is an image, or several images, while its output is a description that must satisfy two criteria:

- It must bear some relationship to what is being imaged.
- It must contain all the information needed for the some given task.

The first criterion ensures that the description depends in some way on the visual input. The second ensures that the information provided is useful.

An object does not have a unique description; we can conceive of descriptions at many levels of detail and from many points of view. It is impossible to describe an object completely. Fortunately, we can avoid this potential philosophical snare by considering the task for which the description is intended. That is, we do not want just any description of what is imaged, but one that allows us to take appropriate action.

A simple example may help to clarify these ideas. Consider again the task of picking parts from a conveyor belt. The parts may be randomly oriented and positioned on the belt. There may be several different types of parts, with each to be loaded into a different fixture. The vision system is provided with images of the objects as they are transported past a camera mounted above the belt. The descriptions that the system has to produce in this case are simple. It need only give the position, orientation, and type of each object. The description could be just a few numbers. In other situations an elaborate symbolic description may be called for.

There are cases where the feedback loop is not closed through a machine, but the description is provided as output to be interpreted by a human. The two criteria introduced above must still be satisfied, but it is harder in this case to determine whether the system was successful in solving the vision problem presented.

1.3 Relation to Other Fields

Machine vision is closely allied with three fields (figure 1-3):

- Image processing.
- Pattern classification.
- Scene analysis.

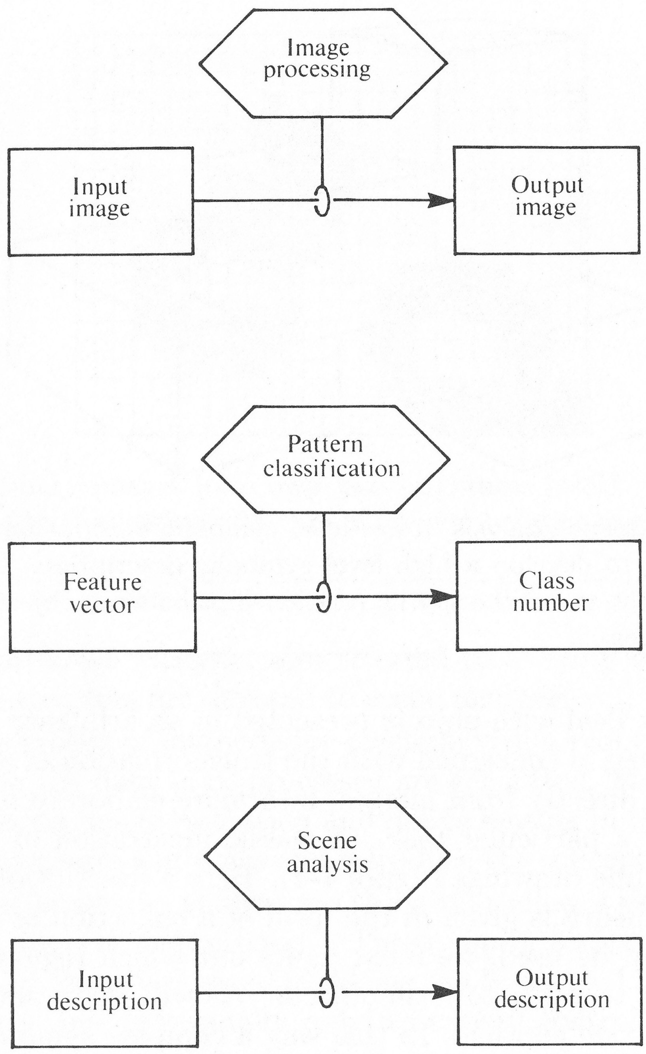

Figure 1-3. Three ancestor paradigms of machine vision are image processing, pattern classification, and scene analysis. Each contributes useful techniques, but none is central to the problem of developing symbolic descriptions from images.

Image processing is largely concerned with the generation of new images from existing images. Most of the techniques used come from linear systems theory. The new image may have noise suppressed, blurring removed, or edges accentuated. The result is, however, still an image, usually meant to be interpreted by a person. As we shall see, some of the techniques of image processing are useful for understanding the limitations of image-forming systems and for designing preprocessing modules for machine vision.

Pattern classification has as its main thrust the classification of a "pattern," usually given as a set of numbers representing measurements of an object, such as height and weight. Although the input to a classifier is not an image, the techniques of pattern classification are at times useful for analyzing the results produced by a machine vision system. To recognize an object means to assign it to one of a number of known classes. Note, however, that recognition is only one of many tasks faced by the machine vision system. Researchers concerned with classification have created simple methods for obtaining measurements from images. These techniques, however, usually treat the images as a two-dimensional pattern of brightness and cannot deal with objects presented in an arbitrary attitude.

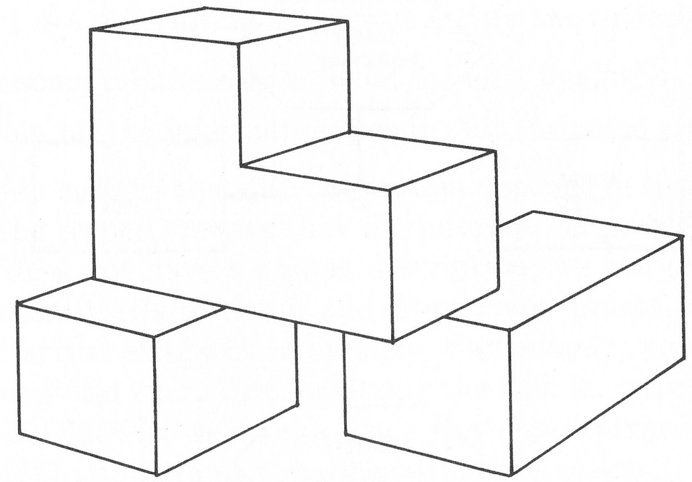

Figure 1-4. In scene analysis, a low-level symbolic description, such as a line drawing, is used to develop a high-level symbolic description. The result may contain information about the spatial relationships between objects, their shapes, and their identities.

Scene analysis is concerned with the transformation of simple descriptions, obtained directly from images, into more elaborate ones, in a form more useful for a particular task. A classic illustration of this is the interpretation of line drawings (figure 1-4). Here a description of the image of a set of polyhedra is given in the form of a collection of line segments. Before these can be used, we must figure out which regions bounded by the lines belong together to form objects. We will also want to know how objects support one another. In this way a complex symbolic description of the image can be obtained from the simple one. Note that here we do not start with an image, and thus once again do not address the central issue of machine vision:

- Generating a symbolic description from one or more images.

1.4 Outline of What Is to Come

The generation of descriptions from images can often be conveniently broken down into two stages. The first stage produces a *sketch*, a detailed but undigested description. Later stages produce more parsimonious, structured descriptions suitable for decision making. Processing in the first stage will be referred to as *image analysis*, while subsequent processing of the results will be called *scene analysis*. The division is somewhat arbitrary, except insofar as image analysis starts with an image, while scene analysis begins with a sketch. The first thirteen chapters of the book are concerned with image analysis, also referred to as *early vision*, while the remaining five chapters are devoted to scene analysis.

The development of methods for machine vision requires some understanding of how the data to be processed are generated. For this reason we start by discussing image formation and image sensing in chapter 2. There we also treat measurement noise and introduce the concept of convolution.

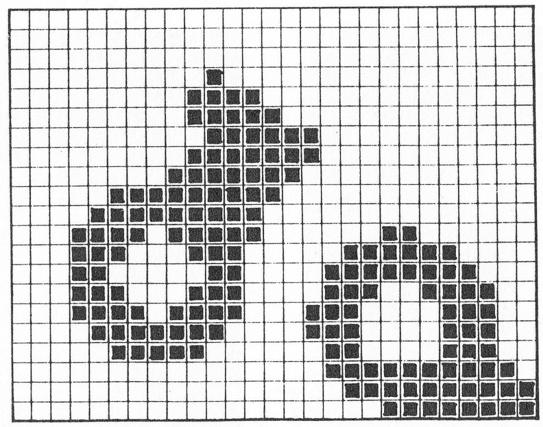

Figure 1-5. Binary images have only two brightness levels: black and white. While restricted in application, they are of interest because they are particularly easy to process.

The easiest images to analyze are those that allow a simple separation of an "object" from a "background." These *binary images* will be treated first (figure 1-5). Some industrial problems can be tackled by methods that use such images, but this usually requires careful control of the lighting. There exists a fairly complete theory of what can and cannot be accomplished with binary images. This is in contrast to the more general case of *gray-level images*. It is known, for example, that binary image techniques are useful only when possible changes in the attitude of the object are confined to rotations in a plane parallel to the image plane. Binary image processing is covered in chapters 3 and 4.

Many image-analysis techniques are meant to be applied to regions of an image corresponding to single objects, rather than to the whole image. Because typically many surfaces in the environment are imaged together, the image must be divided up into regions corresponding to separate entities in the environment before such techniques can be applied. The required segmentation of images is discussed in chapter 5.

In chapters 6 and 7 we consider the transformation of gray-level images into new gray-level images by means of linear operations. The usual intent of such manipulations is to reduce noise, accentuate some aspect of the image, or reduce its dynamic range. Subsequent stages of the machine vision system may find the processed images easier to analyze. Such filtering methods are often exploited in edge-detection systems as preprocessing steps.

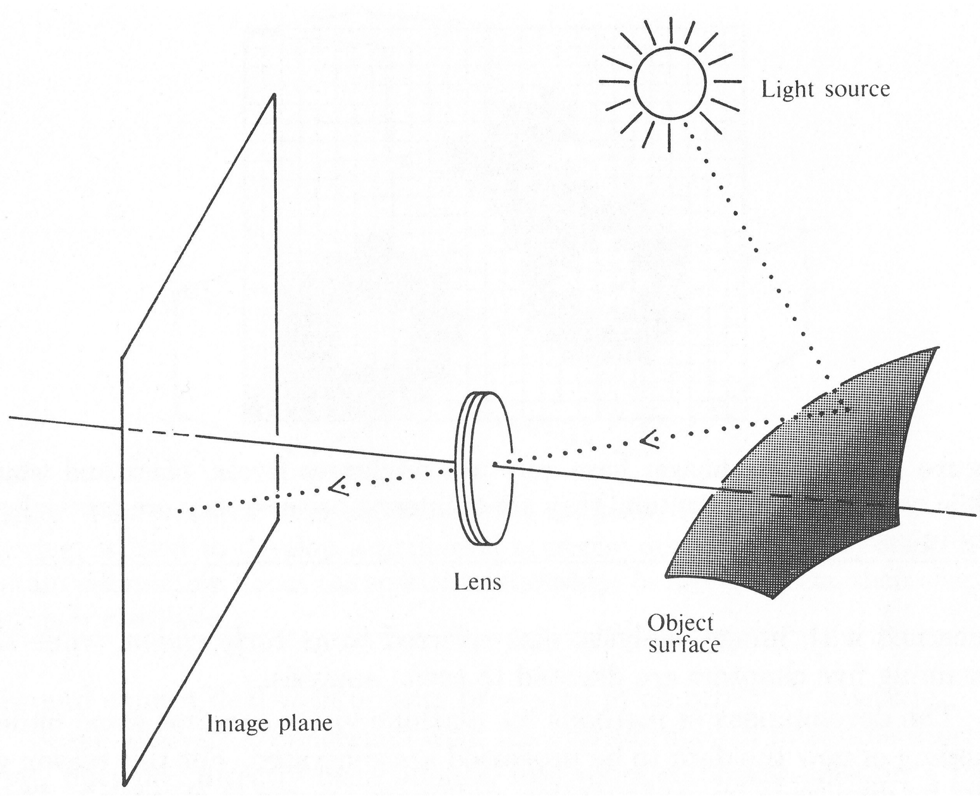

Figure 1-6. In order to use images to recover information about the world, we need to understand image formation. In some cases the image formation process can be inverted to extract estimates of the permanent properties of the surfaces of the objects being imaged.

Complementary to image segmentation is edge finding, discussed in chapter 8. Often the interesting events in a scene, such as a boundary where one object occludes another, lead to discontinuities in image brightness or in brightness gradient. Edge-finding techniques locate such features. At this point, we begin to emphasize the idea that an important aspect of machine vision is the estimation of properties of the surfaces being imaged. In chapter 9 the estimation of surface reflectance and color is addressed and found to be a surprisingly difficult task.

Finally, we confront the central issue of machine vision: the generation of a description of the world from one or more images. A point of view that one might espouse is that the purpose of the machine vision system is to invert the projection operation performed by image formation. This is not quite correct, since we want not to recover the world being imaged, but to obtain a symbolic description. Still, this notion leads us to study image formation carefully (figure 1-6). The way light is reflected from a surface becomes a central issue. The apparent brightness of a surface depends on three factors:

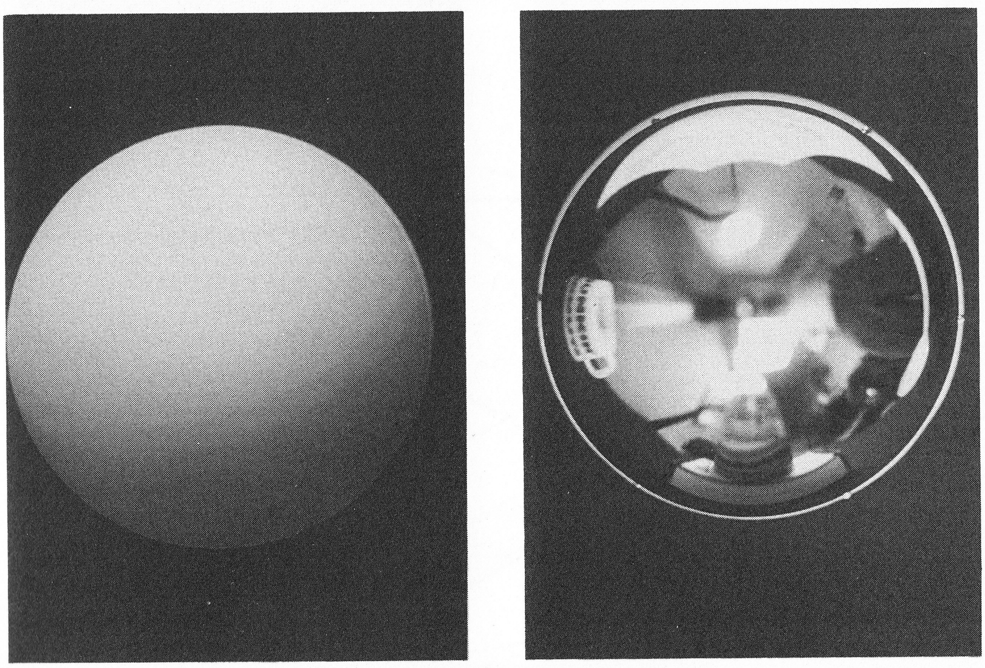

Figure 1-7. The appearance of the image of an object is greatly influenced by the reflectance properties of its surface. Perfectly matte and perfectly specular surfaces present two extreme cases.

Figure 1-8. The appearance of the image of a scene depends a lot on the lighting conditions. To recover information about the world from images we need to understand how the brightness patterns in the image are determined by the shapes of surfaces, their reflectance properties, and the distribution of light sources.

- Microstructure of the surface.
- Distribution of the incident light.
- Orientation of the surface with respect to the viewer and the light sources.

In figure 1-7 we see images of two spherical surfaces, one covered with a paint that has a matte or diffuse reflectance, the other metallic, giving rise to specular reflections. In the second case we see a virtual image of the world around the spherical object. It is clear that the microstructure of the surface is important in determining image brightness.

Figure 1-8 shows three views of Place Ville-Marie in Montreal. The three pictures were taken from the same hotel window, but under different lighting conditions. Again, we easily recognize that the same objects are depicted, but there is a tremendous difference in brightness patterns between the images taken with direct solar illumination and those obtained under a cloudy sky.

In chapters 10 and 11 we discuss these issues and apply the understanding developed to the recovery of surface shape from one or more images. Representations for the shape of a surface are also introduced there. In developing methods for recovering surface shape, we often consider the surface broken up into tiny patches, each of which can be treated as if it were planar. Light reflection from such a planar patch is governed by three angles if it is illuminated by a point source (figure 1-9).

The same systematic approach, based on an analysis of image brightness, is used in chapters 12 and 13 to recover information from time-varying images and images taken by cameras separated in space. Surface shape, object motion, and other information can be recovered from images using the methods developed in these two chapters. The relations between various coordinate systems, either viewer-centered or object-centered, are uncovered in the discussion of photogrammetry in chapter 13, along with an analysis of the binocular stereo problem. In using a machine vision system to guide a mechanical manipulator, measurements in the camera's coordinate system must be transformed into the coordinate system of the robot arm. This topic naturally fits into the discussion of this chapter also.

At this point, we turn from image analysis to scene analysis. Chapter 14 introduces methods for classifying objects based on feature measurements. Line drawings obtained from images of polyhedral objects are analyzed in chapter 15 in order to recover the spatial relationships between the objects.

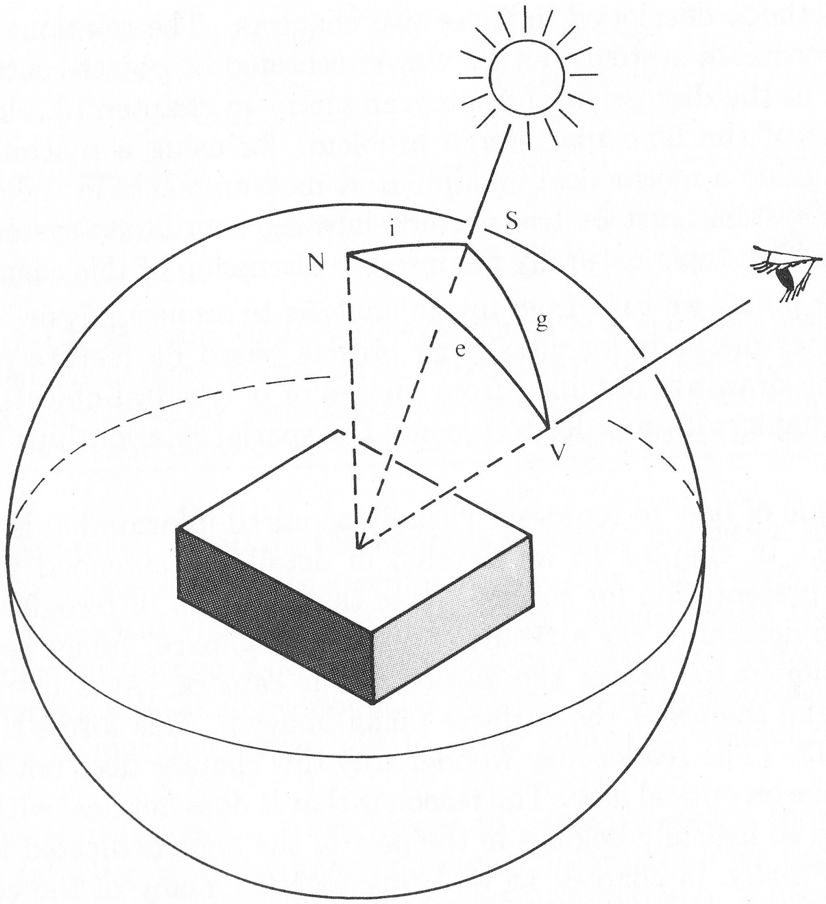

Figure 1-9. The reflection of light from a point source by a patch of an object's surface is governed by three angles: the incident angle i, the emittance angle e, and the phase angle g. Here N is the direction perpendicular, or normal, to the surface, S the direction to the light source, and V the direction toward the viewer.

The issue of how to represent visually acquired information is of great importance. In chapter 16 we develop in detail the extended Gaussian image, a representation for surface shape that is useful in recognition and allows us to determine the attitude of an object in space. Image sequences can be exploited to recover the motion of the camera. As a by-product, we obtain the shapes of the surfaces being imaged. This forms the topic of chapter 17. (The reader may wonder why this chapter does not directly follow the one on optical flow. The reason is that it does not deal with image analysis and so logically belongs in the part of the book dedicated to scene analysis.) Finally, in chapter 18 we bring together many of the concepts developed in this book to built a complete hand–eye system. A robot arm is guided to pick up one object after another out of a pile of objects. Visual input provides the system with information about the positions of the objects and their attitudes in space. In this chapter we introduce some new topics, such as methods for representing rotations in three-dimensional space, and discuss some of the difficulties encountered in building a real-world system.

Throughout the book we start by discussing elementary issues and well-established techniques, progress to more advanced topics, and close with less certain matters and subjects of current research. In the past, machine vision may have appeared to be a collection of assorted heuristics and ad hoc tricks. To give the material coherence we maintain a particular point of view here:

- Machine vision should be based on a thorough understanding of image formation.

This emphasis allows us to derive mathematical models of the image-analysis process. Algorithms for recovering a description of the imaged world can then be based on these mathematical models.

An approach based on the analysis of image formation is, of course, not the only one possible for machine vision. One might start instead from existing biological vision systems. Artificial systems would then be based on detailed knowledge of natural systems, provided these can be adequately characterized. We shall occasionally discuss alternate approaches to given problems in machine vision, but to avoid confusion we will not dwell on them.

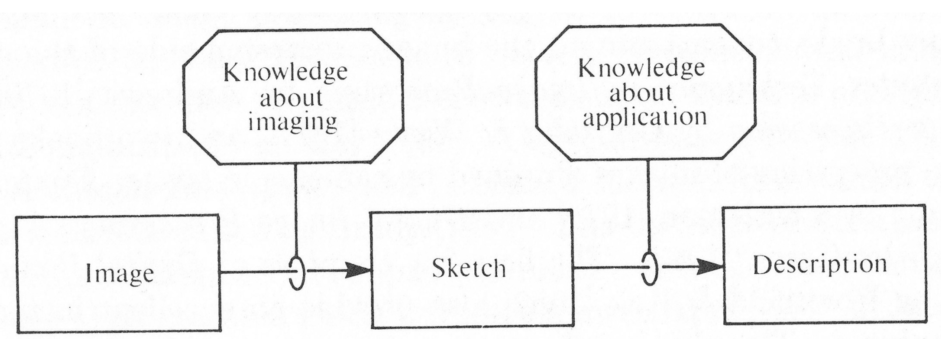

Figure 1-10. In many cases, the development of a symbolic description of a scene from one or more images can be broken down conveniently into two stages. The first stage is largely governed by our understanding of the image-formation process; the second depends more on the needs of the intended application.

The transformation from image to sketch appears to be governed mostly by what is in the image and what information we can extract directly from it (figure 1-10). The transformation from a crude sketch to a full symbolic description, on the other hand, is mostly governed by the need to generate information in a form that will be of use in the intended application.

1.5 References

Each chapter will have a section providing pointers to background reading, further explanation of the concepts introduced in that chapters, and recent results in the area. Books will be listed first, complete with authors and titles. Papers in journals, conference proceedings, and internal reports of universities and research laboratories are listed after the books, but without title. Please note that the bibliography has two sections: the first for books, the second for papers.

There are now numerous books on the subject of machine vision. Of these, *Computer Vision* by Ballard & Brown [1982] is remarkable for its broad coverage. Also notable are *Digital Picture Processing* by Rosenfeld & Kak [1982], *Computer Image Processing and Recognition* by Hall [1979], and *Machine Perception* [1982], a short book by Nevatia. A recent addition is *Vision in Man and Machine* [1985] by Levine, a book that has a biological vision point of view and emphasizes applications to biomedical problems.

Many books concentrate on the image-processing side of things, such as *Computer Techniques in Image Processing* by Andrews [1970], *Digital Image Processing* by Gonzalez & Wintz [1977], and two books dealing with the processing of images obtained by cameras in space: *Digital Image Processing* by Castleman [1979] and *Digital Image Processing: A Systems Approach* by Green [1983]. The first few chapters of *Digital Picture Processing* by Rosenfeld & Kak [1982] also provide an excellent introduction to the subject. The classic reference on image processing is still Pratt's encyclopedic *Digital Image Processing* [1978].

One of the earliest significant books in this field, *Pattern Classification and Scene Analysis* by Duda & Hart [1973], contains more on the subject of pattern classification than one typically needs to know. *Artificial Intelligence* by Winston [1984] has an easy-to-read, broad-brush chapter on machine vision that makes the connection between that subject and artificial intelligence.

A number of edited books, containing contributions from several researchers in the field, have appeared in the last ten years. Early on there was *The Psychology of Computer Vision*, edited by Winston [1975], now out of print. Then came *Digital Picture Analysis*, edited by Rosenfeld [1976], and *Computer Vision Systems*, edited by Hanson & Riseman [1978]. Several papers on machine vision can be found in volume 2 of *Artificial Intelligence: An MIT Perspective*, edited by Winston & Brown [1979]. The collection *Structured Computer Vision: Machine Perception through Hierarchical Computation Structures*, edited by Tanimoto & Klinger, was published in 1980. Finally there appeared the fine assemblage of papers *Image Understanding 1984*, edited by Ullman & Richards [1984].

The papers presented at a number of conferences have also been collected in book form. Gardner was the editor of a book published in 1979 called *Machine-aided Image Analysis, 1978*. Applications of machine vision to robotics are explored in *Computer Vision and Sensor-Based Robots*, edited by Dodd & Rossol [1979], and in *Robot Vision*, edited by Pugh [1983]. Stucki edited *Advances in Digital Image Processing: Theory, Application, Implementation* [1979], a book containing papers presented at a meeting organized by IBM. The notes for a course organized by Faugeras appeared in *Fundamentals in Computer Vision* [1983].

Because many of the key papers in the field were not easily accessible, a number of collections have appeared, including three published by IEEE Press, namely *Computer Methods in Image Analysis*, edited by Aggarwal, Duda, & Rosenfeld [1977], *Digital Image Processing*, edited by Andrews [1978], and *Digital Image Processing for Remote Sensing*, edited by Bernstein [1978].

The IEEE Computer Society's publication *Computer* brought out a special issue on image processing in August 1977, the *Proceedings of the IEEE* devoted the May 1979 issue to pattern recognition and image processing, and *Computer* produced a special issue on machine perception for industrial applications in May 1980. A special issue (Volume 17) of the journal *Artificial Intelligence* was published in book form under the title *Computer Vision*, edited by Brady [1981]. The Institute of Electronics and Communication Engineers of Japan produced a special issue (Volume J68-D, Number 4) on machine vision work in Japan in April 1985 (in Japanese).

Not much is said in this book about biological vision systems. They provide us, on the one hand, with reassuring existence proofs and, on the other, with optical illusions. These startling effects may someday prove to be keys with which we can unlock the secrets of biological vision systems. A computational theory of their function is beginning to emerge, to a great extent due to the pioneering work of a single man, David Marr.

His approach is documented in the classic book *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* [1982].

Human vision has, of course, always been a subject of intense curiosity, and there is a vast literature on the subject. Just a few books will be mentioned here. Gregory has provided popular accounts of the subject in *Eye and Brain* [1966] and *The Intelligent Eye* [1970]. Three books by Gibson—*The Perception of the Visual World* [1950], *The Senses Considered as Perceptual Systems* [1966], and *The Ecological Approach to Visual Perception* [1979]—are noteworthy for providing a fresh approach to the problem. Cornsweet's *Visual Perception* [1971] and *The Psychology of Visual Perception* by Haber & Hershenson [1973] are of interest also. The work of Julesz has been very influential, particularly in the area of binocular stereo, as documented in *Foundations of Cyclopean Perception* [1971]. More recently, in the wonderfully illustrated book *Seeing*, Frisby [1982] has been able to show the crosscurrents between work on machine vision and work on biological vision systems. For another point of view see *Perception* by Rock [1984].

Twenty years ago, papers on machine vision were few in number and scattered widely. Since then a number of journals have become preferred repositories for new research results. In fact, the journal *Computer Graphics and Image Processing*, published by Academic Press, had to change its name to *Computer Vision, Graphics and Image Processing* (CVGIP) when it became the standard place to send papers in this field for review. More recently, a new special-interest group of the Institute of Electrical and Electronic Engineers (IEEE) started publishing the *Transactions on Pattern Analysis and Machine Intelligence* (PAMI). Other journals, such as *Artificial Intelligence*, published by North-Holland, and *Robotics Research*, published by MIT Press, also contain articles on machine vision. There are several journals devoted to related topics, such as pattern classification.

Some research results first see the light of day at an "Image Understanding Workshop" sponsored by the Defense Advanced Research Projects Agency (DARPA). Proceedings of these workshops are published by Science Applications Incorporated, McLean, Virginia, and are available through the Defense Technical Information Center (DTIC) in Alexandria, Virginia. Many of these papers are later submitted, possibly after revision and extension, to be reviewed for publication in one of the journals mentioned above.

The Computer Society of the IEEE organizes annual conferences on

Computer Vision and Pattern Recognition (CVPR) and publishes their proceedings. Also of interest are the proceedings of the biannual International Joint Conference on Artificial Intelligence (IJCAI) and the national conferences organized by the American Association for Artificial Intelligence (AAAI), usually in the years in between.

The thorough annual surveys by Rosenfeld [1972, 1974, 1975, 1976, 1977, 1978, 1979, 1980, 1981, 1982, 1983, 1984a, 1985] in *Computer Vision, Graphics and Image Processing* are extremely valuable and make it possible to be less than complete in providing references here. The most recent survey contained 1,252 entries! There have been many analyses of the state of the field or of particular views of the field. An early survey of image processing is that of Huang, Schreiber, & Tretiak [1971]. While not really a survey, the influential paper of Barrow & Tenenbaum [1978] presents the now prevailing view that machine vision is concerned with the process of recovering information about the surfaces being imaged. More recent surveys of machine vision by Marr [1980], Barrow & Tenenbaum [1981a], Poggio [1984], and Rosenfeld [1984b] are recommended particularly. Another paper that has been influential is that by Binford [1981].

Once past the hurdles of early vision, the representation of information and the modeling of objects and the physical interaction between them become important. We touch upon these issues in the later chapters of this book. For more information see, for example, Brooks [1981] and Binford [1982].

There are many papers on the application of machine vision to industrial problems (although some of the work with the highest payoff is likely not to have been published in the open literature). Several papers in *Robotics Research: The First International Symposium*, edited by Brady & Paul [1984], deal with this topic. Chin [1982] and Chin & Harlow [1982] have surveyed the automation of visual inspection.

The inspection of printed circuit boards, both naked and stuffed, is a topic of great interest, since there are many boards to be inspected and since it is not a very pleasant job for people, nor one that they are particularly good at. For examples of work in this area, see Ejiri et al. [1973], Danielsson & Kruse [1979], Danielsson [1980], and Hara, Akiyama, & Karasaki [1983]. There is a similar demand for such techniques in the manufacture of integrated circuits. Masks are simple black-and-white patterns, and their inspection has not been too difficult to automate. The inspection of integrated circuit wafers is another matter; see, for example, Hsieh & Fu [1980].

Machine vision has been used in automated alignment. See Horn [1975b], Kashioka, Ejiri, & Sakamoto [1976], and Baird [1978] for examples in semiconductor manufacturing. Industrial robots are regularly guided using visually obtained information about the position and orientation of parts. Many such systems use binary image-processing techniques, although some are more sophisticated. See, for example, Yachida & Tsuji [1977], Gonzalez & Safabakhsh [1982], and Horn & Ikeuchi [1984]. These techniques will not find widespread application if the user has to program each application in a standard programming language. Some attempts have been made to provide tools specifically suited to the vision applications; see, for example, Lavin & Lieberman [1982].

Papers on the application of machine vision methods to the vectorization of line drawings are mentioned at the end of chapter 4; references on character recognition may be found at the end of chapter 14.

1.6 Exercises

1-1 Explain in what sense one can consider pattern classification, image processing, and scene analysis as "ancestor paradigms" to machine vision. In what way do the methods from each of these disciplines contribute to machine vision? In what way are the problems addressed by machine vision different from those to which these methods apply?