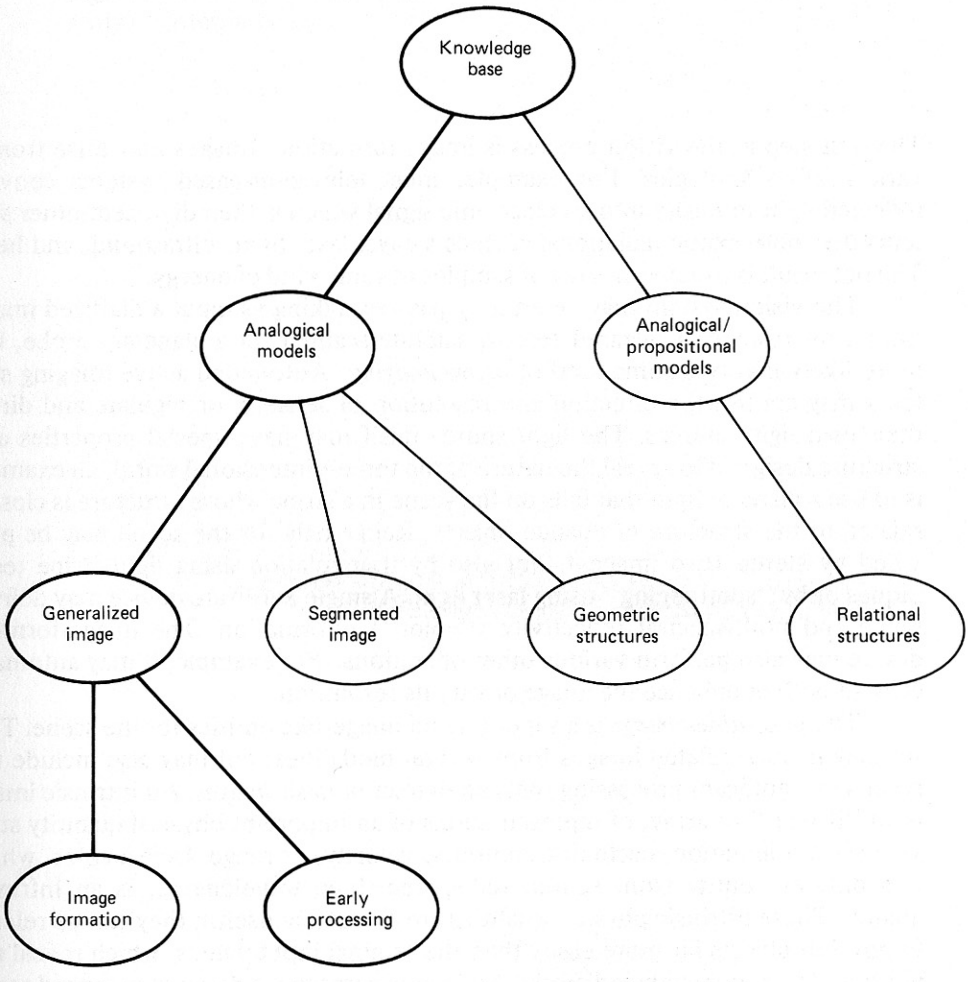
GENERALIZED IMAGES

I



The first step in the vision process is image formation. Images may arise from a variety of technologies. For example, most television-based systems convert reflected light intensity into an electronic signal which is then digitized; other systems use more exotic radiations, such as x-rays, laser light, ultrasound, and heat. The net result is usually an array of samples of some kind of energy.

The vision system may be entirely passive, taking as input a digitized image from a microwave or infrared sensor, satellite scanner, or a planetary probe, but more likely involves some kind of *active imaging.* Automated active imaging systems may control the direction and resolution of sensors, or regulate and direct their own light sources. The light source itself may have special properties and structure designed to reveal the nature of the three-dimensional world; an example is to use a plane of light that falls on the scene in a stripe whose structure is closely related to the structure of opaque objects. Range data for the scene may be provided by stereo (two images), but also by triangulation using light-stripe techniques or by "spotranging" using laser light. A single hardware device may deliver range and multispectral reflectivity ("color") information. The image-forming device may also perform various other operations. For example, it may automatically smooth or enhance the image or vary its resolution.

The *generalized image* is a set of related image-like entities for the scene. This set may include related images from several modalities, but may also include the results of significant processing that can extract *intrinsic images.* An intrinsic image is an "image," or array, of representations of an important physical quantity such as surface orientation, occluding contours, velocity, or range. Object color, which is a different entity from sensed red—green—blue wavelengths, is an intrinsic quality. These intrinsic physical qualities are extremely useful; they can be related to physical objects far more easily than the original input values, which reveal the physical parameters only indirectly. An intrinsic image is a major step toward scene understanding and usually represents significant and interesting computations.

The information necessary to compute an intrinsic image is contained in the input image itself, and is extracted by "inverting" the transformation wrought by the imaging process, the reflection of radiation from the scene, and other physical processes. An example is the fusion of two stereo images to yield an intrinsic range image. Many algorithms to recover intrinsic images can be realized with *parallel* implementations, mirroring computations that may take place in the lower neurological levels of biological image processing.

All of the computations listed above benefit from the idea of *resolution pyramids.* A pyramid is a generalized image data structure consisting of the same image at several successively increasing levels of resolution. As the resolution increases, more samples are required to represent the increased information and hence the successive levels are larger, making the entire structure look like a pyramid. Pyramids allow the introduction of many different coarse-to-fine image-resolution algorithms which are vastly more efficient than their single-level, high-resolution-only counterparts.

Image Formation 2

2.1 IMAGES

Image formation occurs when a *sensor* registers *radiation* that has interacted with *physical objects.* Section 2.2 deals with mathematical models of images and image formation. Section 2.3 describes several specific image formation technologies.

The mathematical model of imaging has several different components.

1. An *image function* is the fundamental abstraction of an image.
2. A *geometrical model* describes how three dimensions are projected into two.
3. A *radiometrical model* shows how the imaging geometry, light sources, and reflectance properties of objects affect the light measurement at the sensor.
4. A *spatial* frequency model describes how spatial variations of the image may be characterized in a transform domain.
5. A *color model* describes how different spectral measurements are related to image colors.
6. A *digitizing model* describes the process of obtaining discrete samples.

This material forms the basis of much image-processing work and is developed in much more detail elsewhere, e.g., [Rosenfeld and Kak 1976; Pratt 1978]. Our goals are not those of image processing, so we limit our discussion to a summary of the essentials.

The wide range of possible sources of samples and the resulting different implications for later processing motivate our overview of specific imaging techniques. Our goal is not to provide an exhaustive catalog, but rather to give an idea of the range of techniques available. Very different analysis techniques may be needed depending on how the image was formed. Two examples illustrate this point. If the image is formed by reflected light intensity, as in a photograph, the image records both light from primary light sources and (more usually) the light reflected off physical surfaces. We show in Chapter 3 that in certain cases we can use these kinds of images together with knowledge about physics to derive the orientation of the surfaces. If, on the other hand, the image is a computed tomogram of the human body (discussed in Section 2.3.4), the image represents tissue density of internal organs. Here orientation calculations are irrelevant, but general segmentation techniques of Chapters 4 and 5 (the agglomeration of neighboring samples of similar density into units representing organs) are appropriate.

2.2 IMAGE MODEL

Sophisticated image models of a statistical flavor are useful in image processing [Jan 1981]. Here we are concerned with more geometrical considerations.

2.2.1 Image Functions

An *image function* is a mathematical representation of an image. Generally, an image function is a vector-valued function of a small number of arguments. A special case of the image function is the *digital (discrete) image function*, where the arguments to and value of the function are all integers. Different image functions may be used to represent the same image, depending on which of its characteristics are important. For instance, a camera produces an image on black-and-white film which is usually thought of as a real-valued function (whose value could be the density of the photographic negative) of two real-valued arguments, one for each of two spatial dimensions. However, at a very small scale (the order of the film grain) the negative basically has only two densities, "opaque" and "transparent."

Most images are presented by functions of two *spatial* variables $f(\mathbf{x}) = f(x, y)$, where $f(x, y)$ is the brightness of the gray level of the image at a spatial coordinate (x, y). A multispectral image $\mathbf{f}$ is a vector-valued function with components $(f_1 \ldots f_n)$. One special multispectral image is a color image in which, for example, the components measure the brightness values of each of three wavelengths, that is,

$$\mathbf{f}(\mathbf{x}) = \left\{ f_{\text{red}}(\mathbf{x}), f_{\text{blue}}(\mathbf{x}), f_{\text{green}}(\mathbf{x}) \right\}$$

Time-varying images $\mathbf{f}(\mathbf{x}, t)$ have an added temporal argument. For special three-dimensional images, $\mathbf{x} = (x, y, z)$. Usually, both the domain and range of $\mathbf{f}$ are bounded.

An important part of the formation process is the conversion of the image representation from a continuous function to a discrete function; we need some way of describing the images as samples at discrete points. The mathematical tool we shall use is the *delta function*.

Formally, the delta function may be defined by