1 Introduction

Our immediate environment is a magnificent tapestry of information-bearing signals of many kinds: some are man-made signals and some are not, reaching us from many directions. Some signals, such as optical signals and acoustic signals, are immediately compatible with our senses. Other signals, as in the radio and radar bands, or as in the infrared, ultraviolet, and X-ray bands, are not directly compatible with human senses. To perceive one of these signals, we require a special apparatus to convert it into observable form.

A great variety of man-made sensors now exist that collect signals and process those signals to form some kind of image, normally a visual image, of an object or a scene of objects. We refer to these as sensors for remote surveillance. There are many kinds of sensors collected under this heading, differing in the size of the observed scene, as from microscopes to radio telescopes; in complexity, as from the simple lens to synthetic-aperture radars; and in the current state of development, as from photography to holography and tomography. Each of these devices collects raw sensor data and processes that data into imagery that is useful to a user. This processing might be done by a digital computer, an optical computer, or an analog computer. The development and description of the processing algorithms will often require a sophisticated mathematical formulation.

In this book, we shall bring together a number of signal-processing concepts that will form a background for the study and design of the many kinds of image formation and remote surveillance systems. The signal-processing principles we shall study include or adjoin the theory of classical radar and sonar systems, electromagnetic propagation, tomography, and physical optics, as well as estimation and detection theory.

1.1 Remote image-formation systems

Mankind has designed a variety of devices that are used to observe the environment by processing electromagnetic radiation in the radio and microwave frequency bands, the infrared band, and the optical or X-ray band; or by processing acoustic, pressure, or magnetic variations. The many varieties of radar and sonar systems are examples of

Figure 1.1 Elements of a remote surveillance system

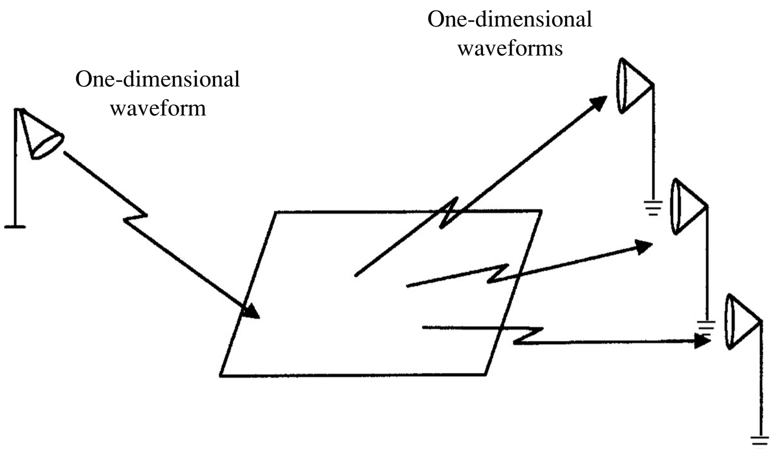

Figure 1.2 Probing a scene with waveforms

such systems. A surveillance system may be active, using its own energy to illuminate the environment, or it may be passive, relying on signals already in the environment.

Surveillance theory studies the design of signals to probe the environment as well as the design of computational procedures for the extraction of information from received signals within which that information may be deeply buried. As such, surveillance theory is that branch of information theory that is explicitly concerned with the design of systems to observe the environment and with the performance of those systems. The theory is concerned specifically with the mathematical structure of the algorithms, such as image-formation algorithms, needed to extract information from received signals.

A precise definition of the term "remote surveillance system" is difficult to formulate. Loosely, we mean any system that collects signals and creates an observable image by processing those signals. Figure 1.1 illustrates how a remote surveillance system can be partitioned into the "sensors" and the "algorithms." We shall be concerned in detail with the image-formation algorithms and with the performance, but not with the detailed physics of the sensors or with the implementation of the algorithms.

The "image," which often is the end product of the surveillance system, is always some kind of depiction of an "actual" scene, usually a two-dimensional or three-dimensional scene, which we denote as $\rho(x, y)$ or $\rho(x, y, z)$. The scene may emit its own signals that the surveillance sensors intercept, or it may be probed by signals generated within the surveillance system. Figure 1.2 shows a representative configuration in which the scene $\rho(x, y)$ is probed by signals generated as one-dimensional waveforms. In this case, the sensors collect one or more reflected one-dimensional waveforms, $s_m(t)$, and from these, the computational algorithms must form a suitable image of the scene. Thus the computational task is to estimate a two-dimensional function, $\rho(x, y)$ (or a three-dimensional function, $\rho(x, y, z)$) when given a set of one-dimensional scattered waveforms, $s_m(t)$, that depend on $\rho(x, y)$ (or $\rho(x, y, z)$). This kind of task is sometimes called an *inverse problem*. Among the most useful mathematical tools that we shall develop for this task are the two-dimensional Fourier transform, the projection-slice theorem, and the ambiguity function.

Another important topic that will be studied is the relationship between the signal at the input aperture of a transducer, such as an antenna, and the wavefront radiated by the antenna. The reflection of waves, however, will be modeled in a simple way. The detailed relationship between the incident wave on a reflecting object and the reflected wave, which is called the *forward problem*, is of interest in this book only insofar as it affects the inverse problem. Simple models for the reflection will be adequate for most of our purposes.

Radar and sonar are among the surveillance systems that will be studied. Originally, radar and sonar systems used simple waveforms and simple processing techniques that could be implemented with simple devices such as filters and threshold detectors. But, over the years, a new level of sophistication began to find its way into many of these systems. By maintaining a precise phase record of long-duration signals, and processing the signals phase-coherently, one can obtain new levels of system performance. Systems that depend on phase coherence over time are called *coherent surveillance* systems. Some early coherent systems in the radar bands were designed to use optical processing. More recently, digital processing of coherent radar waveforms has become practical.

Our goal is to develop the theory of the various kinds of remote surveillance systems in a common mathematical setting. We shall be concerned with both coherent processing techniques, such as are used for forming the images of radar reflectors, for detecting moving objects, and for passively locating radiation sources, and with other processing techniques, such as are used in X-ray tomography, that do not employ phase coherence. The scope of the treatment is illustrated by these examples.

The many forms of remote surveillance imagery, such as radar imagery, may often look very different from visual imagery or conventional photographs of that same scene or object. This means that the user of that sensor may need training and experience in interpreting the output imagery. To the novice, it may seem to be a limitation of the sensor, but a more sophisticated view is that a new sensor opens a new window in our way of perceiving reality. A bat or a dolphin lives in a world that is perceived in large measure by means of acoustic or sonar data. This kind of sensor has nothing like the high angular resolution of our optical world, yet it does have other attributes, such as a strong doppler shift and the ability to resolve objects instantly by their velocities. Because it has a different kind of input data, the dolphin or the bat undoubtedly perceives the world differently from the way in which we do.

One way of defining the kind of surveillance system we shall study is as a system in which raw signals in the environment that the human cannot sense directly are turned into processed signals that are compatible with one of the human senses. Thus a radar receiver converts an electromagnetic wave into a visual image compatible with the human eye, and a tomographic medical scanner turns an X-ray signal into another kind of visual image. Again, the novice may resent the fact that the image is not an exact replica of a photograph. An X-ray image of the human body does not look like a photograph of a human skeleton, but, to the diagnostician, it will be preferred because it contains other kinds of useful information.

1.2 The history of image formation

The subject of remote image formation consists of the common overlap of a number of well-developed subjects such as physical optics and signal processing, and, from a broad point of view, its historical roots go back to the roots of these various subjects. We are interested here in a narrower view of the history of remote surveillance, especially of coherent surveillance systems and tomography. Our brief treatment will serve only to sketch the historical background of the material in this book.

There are many kinds of remote surveillance systems that were developed independently, but which share common fundamentals of signal processing and a common mathematical framework. These include: imaging radars, moving-target detection radars, optical imaging, holography, radio astronomy, sonar beamforming, microscopy, diffraction crystallography, as well as more recent topics such as seismic processing, tomography, and passive source location.

Optical image-formation systems are the earliest, the most developed, and the most readily accepted by the user. Optical images are usually quite sharp, and with high resolution and excellent color contrast. Optical imaging systems may be passive, usually using reflected light and occasionally radiated light, or may be active, using an illumination source. Many such systems are adequately described using geometrical optics. Image-formation systems using radiation in the infrared bands, which are similar to optical systems, form images of temperature variations in a scene because an object emits infrared radiation according to its temperature.

Early radars used simple techniques for processing the received data, while modern radars can use processing that is quite sophisticated. Sophisticated radar signal processing first appeared in the development of those imaging radars known as *synthetic-aperture radars*. Wiley, in 1951, suggested the principle of such radars, although he did not publish his ideas nor did those ideas then result directly in the construction of such a radar. Wiley observed that, whereas the azimuthal resolution of a conventional radar is limited by the width of the antenna beam, each reflecting element within the antenna beam from a moving radar has a doppler frequency shift that depends on the angle between the velocity vector of the radar and the direction to the reflecting element. Thus he concluded that a precise frequency analysis of the radar reflections would provide finer along-track resolution than the azimuthal resolution defined by the antenna beamwidth. The following year, a group at the University of Illinois arrived at the same idea independently, based upon frequency analysis of experimental radar returns. During the summer of 1953, these ideas were reviewed by the members of a summer study, "Project Wolverine," at the University of Michigan and plans were laid for the development of synthetic-aperture radar. It was recognized that the processing requirements placed extreme demands on the technology of the day. Many kinds of analog processors (filter banks, storage tubes, etc.) were tried. Meanwhile, Emmett Leith, at the University of Michigan, turned to the processing ideas of holography and modified the optical processing techniques to satisfy the processing requirements for radar. In 1957, by using optical processing, the first synthetic-aperture radar was successfully demonstrated. Later, Green (1962) proposed the use of range-doppler techniques for remote radar imaging of the surface of rotating planetary objects, a method closely related to synthetic-aperture radar. High-resolution radar images of Venus from the earth gave us our first view of the surface of that planet.

Optical processors[1] are analog processors based on the Fourier transforming property of a lens. These have been the processors of choice for imaging radars because of the sheer volume of data that can be handled. However, optical processors for imaging are slow and may require developing the photographic film twice within the processing. Optical processors are very sensitive to vibration, and they are limited in the form of computation that can be included. Hence attention now has turned to other methods for processing. The advent of high-speed, digital array processors has had a large impact on the processing of synthetic-aperture imagery, and optical processing now plays a diminished or vanishing role.

The development of search radars for the detection of moving targets is spread more broadly, and individual contributions are not as easy to identify. From the first use of radar, it was recognized that the need to detect moving targets could be satisfied by using the doppler shift of the return. A moving reflector causes a doppler-shifted echo. However, the magnitude of the doppler shift is only a very small fraction of the transmitted pulse bandwidth. At that time, the technology did not exist to filter a faint, doppler-shifted signal from a strong background of signals echoed from other stationary emitters. Hence the development of search radars did not depend so much on invention at the conceptual level as it did on the development of technology to support widely understood requirements. By the end of World War II, radars had been developed that used doppler filters to suppress the clutter signal reflected from the stationary

[1] Not to be confused with photonic processors.

background. These early radars used simple delay lines to cancel the stationary return from one pulse with the (nearly identical) return from the previous pulse, thereby rejecting signals with zero doppler shift. In this way, large, rapidly moving objects could be detected from stationary radar platforms.

Later, the requirements for search radars shifted to include moving, airborne radars for observing small, slowly moving target objects at long range. It then became necessary to employ much more delicate techniques for finding a signal return within a large clutter background. These techniques employ coherent processing with the aid of large digital computers.

Meanwhile, astronomers had come to realize that a large amount of astronomical information reaches the earth in the microwave bands. Astronomers are well grounded in optical theory where beamwidths smaller than one arc second are obtained. In the microwave band, a comparable beamwidth requires a reception antenna that is many miles in diameter. Under the impact of wind, ice, and temperature gradients, such an antenna would need to be mechanically rigid to a small fraction of an inch. Clearly, such antennas are impractical. Around 1952, Martin Ryle, at the University of Cambridge, began to study methods for artificially creating such an aperture by combining individual antenna elements, or by allowing the earth's rotation to sweep an array of fixed antenna elements through space. In retrospect, this development of radio astronomy may be viewed as a passive counterpart to the development of active synthetic-aperture radar. The aperture is synthesized by recording the radio signal received at two or more antenna elements and later processing these records coherently within a digital computer. The first such radio telescope was the Cambridge One-Mile Radio Telescope completed in 1964, followed by the Cambridge Five-Kilometer radio telescope in 1971. More recently, other synthetic-aperture radio telescopes have been built and put into operation throughout the world. (The continent-sized Very Large Baseline Array has an angular resolution of 0.0002 arc second.) For the development of synthetic-aperture radio telescopes, Ryle was awarded the 1974 Nobel prize in physics (jointly with Hewish who discovered pulsars with the radio telescope).

The diffraction of X-rays by crystals was demonstrated in 1912 by Max von Laue as a proof of the wave properties of X-rays. Sir William H. Bragg immediately inverted the point of view to turn this diffraction phenomenon into a way of probing crystals, which has since evolved into a sophisticated imaging technique. The 1914 Nobel prize in physics was awarded to von Laue, and the 1915 Nobel prize in physics was awarded to Bragg and his son, Sir W. Lawrence Bragg, who formulated the famous Bragg law of diffraction. This early work was directed toward finding the lattice structure of the crystal as a whole, but was not much concerned with the structure of the individual molecular cell. More recently, attention has turned to the finer question of finding the scattering structure within an individual cell. A difficulty of the task is that, because of the small wavelength of X-rays, the phase of the diffracted X-ray wavefronts cannot be measured. Herbert Hauptman and Jerome Karle (1953) showed how to bypass this problem of missing phase by using prior knowledge about the molecules composing the crystal, for which they received the 1985 Nobel prize in chemistry. Earlier, James Watson and Francis Crick – using the diffraction images produced by Rosalind Franklin – discovered the structure of the DNA molecule, for which they shared the 1962 Nobel prize in medicine.

Closely related to the methods of the Fourier transform and signal processing are many kinds of optical processing, many of them using diffraction phenomena that are describable in terms of the two-dimensional Fourier transform. A method known as the *schlieren method* was proposed by Jean Foucault in 1858 as a way to image air density variations. Frits Zernike in 1935 developed phase-contrast methods to improve microscopy images, for which he was awarded the 1953 Nobel prize in physics. Aaron Klug (1964) developed methods for the imaging of viruses using the diffraction of laser light by electron microscope images, for which he won the 1982 Nobel prize in chemistry.

Dennis Gabor, influenced by the techniques used in microscopy and crystallography, proposed holography in a series of papers in 1948, 1949, and 1951, originally for microscopy but later as a replacement for photography. The work earned Gabor the 1971 Nobel prize in physics. Gabor realized that, whereas conventional photography first processes the optical information to form an image which is then recorded on film, it is also possible to record the raw optical data on the photographic film directly and place the processing in the future with the viewer. He called his method for the photographic recording of the raw optical data a *hologram*. Because the raw optical data contain more information than a final photographic image, in principle the hologram can be used to create images superior to a photograph. Most striking in this regard is the creation of three-dimensional images from a two-dimensional hologram. Holography is technically much more difficult than photography because recording the raw optical data requires precision on the order of optical wavelengths. For this reason, the idea of holography did not immediately draw the attention it deserved. Holography became more attractive after the invention of the laser and also after the more practical reformulation by Leith and Upatnieks (1962), which was strongly influenced by Leith's work on synthetic-aperture radar.

Imaging was introduced into medical diagnostics by Röntgen in 1895 with the invention of X-ray radiography, which exposes a photographic film to transmission X-rays. Edison, by introducing an X-ray-sensitive fluorescent screen or fluoroscope in 1896, eliminated the delay required to process the film. The development of X-ray tomography in the modern sense of computerized image reconstruction for medical applications began in Great Britain. The key feature, based on the projection-slice theorem and the Radon transform, is the algorithmic reconstruction of images from their X-ray projections, first developed by Cormack in 1963 and reduced to practice by Hounsfield in 1971. The 1979 Nobel prize in physiology and medicine was awarded to Hounsfield and Cormack for the development of computerized tomography. The ideas of tomography are closely related to similar methods used in radio astronomy, especially the formulation of reconstruction algorithms by Bracewell (1956). Other kinds of computerized tomography are now in use for medical diagnostic imaging systems. In addition to the method of projection tomography based on X-ray projections, there are the methods of emission tomography and diffraction tomography. Emission tomography based on radioisotope decay was proposed by Kuhl and Edwards (1963).

Magnetic-resonance imaging (MRI) is yet another kind of tomographic imaging system based on magnetic excitation of atomic nuclei and their subsequent radiation. The ground-breaking idea that enables magnetic-resonance imaging, for which Lauterbur and Mansfield (1973) shared the 2003 Nobel prize in medicine, is to use gradient magnetic fields to encode spatial information into the transient response of a spin system after magnetic pulse excitations. Whereas X-ray tomography gives an image of electron density, MRI gives an image of the distribution of hydrogen nuclei (protons), though in principle it may be tuned to observe instead the distribution of other species of nuclei. The physical phenomenon of nuclear magnetic resonance had been observed in 1946 by Bloch and Purcell, independently, for which they received the 1952 Nobel prize in physics. It was later realized that magnetic resonance effects varied with the kind of tissue excited, but it was not known how to use this effect to make images. Lauterbur conceived and demonstrated his method of using the magnetic resonance phenomenon to form images by spatially encoding the magnetic field, for which he received the 2003 Nobel prize in medicine. Since then, MRI has become an important modality in medical diagnosis. By using both static and time-varying magnetic excitation fields, a magnetic-resonance imaging system causes all nuclei of a given kind, selected by resonance frequency, to oscillate, but with an amplitude and frequency modulation that depend on position as determined by the magnetic excitation at that position. The energy that is radiated by the selected species of nuclei, usually hydrogen nuclei, is measured and sorted by frequency analysis. Because of the spatially-varying magnetic excitation, the frequency distribution of the radiated energy corresponds to the spatial distribution of the sources of radiation, which equates to the spatial density distribution of the target nuclei. Sophisticated algorithms based on the mathematics of tomography have been developed to so extract the image from the frequency distribution of multiple projections of measured data.

1.3 Radar and sonar systems

A radar obtains information about an object or a scene by illuminating the object or scene with electromagnetic waves, then processing the echo signal that is reflected from that object and intercepted by the radar receiving antenna. A sonar obtains information

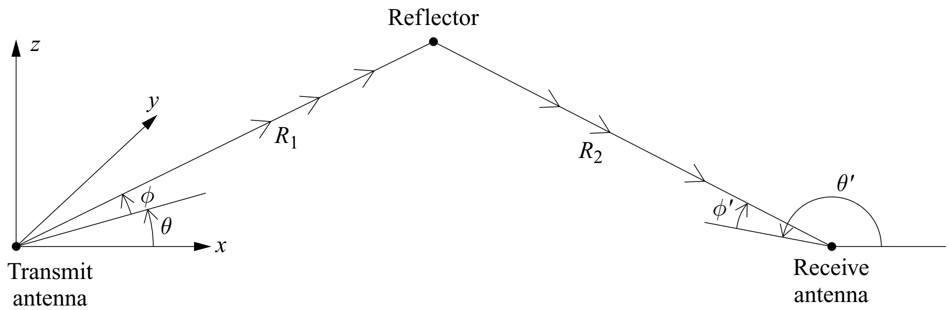

Figure 1.3 Elementary radar/sonar geometry

about an object or a scene by illuminating the object or scene with acoustic waves, then processing the echo signal that is reflected from that object and intercepted by the sonar hydrophones. The geometry of a radar or sonar system is shown in Figure 1.3. The transmitting antenna and the receiving antenna are shown at separate locations although, in most instances, they will be together; they may even be the same antenna, as is the usual case.

By using electromagnetic waves in the microwave bands, a radar is able to penetrate optically opaque media such as clouds, dust, or foliage. In this way, it is possible to form radar images of objects hidden by dust, soil, or snow. Similarly, a sonar or ultrasound system can form images of objects that are optically masked by opaque media. In some cases, the broad beamwidth of a radar antenna or a sonar hydrophone is attractive as a way of viewing a large region of space. For reasons such as these, radar and sonar have long been popular as surveillance systems.

While there may be a great deal of difference between the propagation of electromagnetic waves and the propagation of pressure waves, there is also a great deal of similarity. This similarity carries over to radar and sonar systems. From our point of view, each is a system that forms a complex baseband pulse, $s(t)$, which is transmitted as the amplitude and phase modulation of the passband pulse $\widetilde{s}(t)$, and receives an echo pulse, $\widetilde{v}(t)$, which is a delayed and frequency-shifted copy of the passband pulse $\widetilde{s}(t)$ and is also contaminated by noise. The transmitted pulse $\widetilde{s}(t)$ propagates at a velocity of c over a path of length R_1 from the transmitter to the reflector, and then over a path of length R_2 from the reflector to the receiver. The received pulse $\widetilde{v}(t)$ is an echo of the transmitted pulse but, because it is contaminated by noise and other impairments, it may be difficult to recognize or use. We shall be interested in methods of processing the received pulse to extract useful information from it. The same fundamental ideas apply equally to radar pulses and to sonar pulses.

Figure 1.4 shows an elementary block diagram of a radar system. The baseband pulse $s(t)$ is "up-converted" to form the passband pulse $\widetilde{s}(t)$ that is transmitted by the antenna. The signal transmitted at angle (ϕ, ψ) is $E_t(\phi, \psi)\widetilde{s}(t)$, where $E_t(\phi, \psi)$ is a function of the angles ϕ and ψ and is called the *antenna pattern* of the transmitting antenna. The

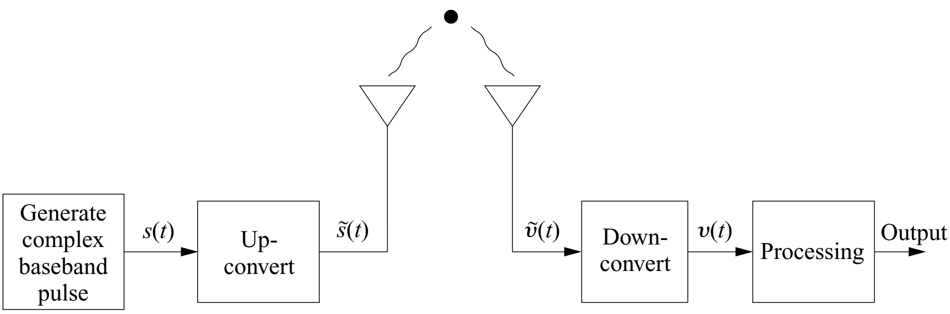

Figure 1.4 Elementary radar block diagram

signal reflected from the reflector is (proportional to[2]) $\rho E_t(\phi, \psi)\widetilde{s}(t - R_1/c)$ where ρ is a parameter called the *reflectivity* of the reflector and is usually dependent on the incident angle and the scattered angle of the signal. The echo signal intercepted by the receiving antenna is

$$\widetilde{v}(t) = E_r(\phi', \psi')\rho E_t(\phi, \psi)\widetilde{s}(t - (R_1 + R_2)/c) + n(t)$$

where $n(t)$ is additive noise, and $E_r(\phi', \psi')$ is the antenna pattern of the receiving antenna. The angular coordinates of the receiving antenna are distinguished by primes. The received echo signal $\widetilde{v}(t)$ is "down-converted" to the complex baseband representation $v(t)$. The signal-processing task is to process the received baseband signal $v(t)$ to obtain information about the reflector such as its position, velocity, size, or even its detailed shape.

The received signal is distributed both in space across the aperture of an antenna, and in time. The distribution in space may be processed, usually by the antenna system, to gather all of the received spatially distributed signal into a single, time-dependent signal. Simple, linear processing of the signal across the aperture is usually summarized by referring to the shape and width of an antenna "beam." The time variations of the received signal are processed so as to determine the time-varying range of the reflecting objects. The space distribution may be processed in other ways to determine the direction of arrival of the signal. We shall study the processing of the space distribution of the signal in Chapter 5 (and Chapter 13) and the processing of the time distribution of the signal in several later chapters.

We define a coherent surveillance system loosely as any system for which coherent processing is fundamental to its operation. We further define coherent processing as the processing of a received passband signal (or signals) by procedures that exploit the carrier phase structure of the waveforms to extract the desired information more effectively. We shall see that the extraction of maximum information from a received

[2] A proportionality constant associated with spherical wave propagation is usually suppressed in a discussion of this kind and is treated separately in a power budget calculation.