J. Patrick Fitch

Synthetic Aperture Radar

Springer-Verlag

J. Patrick Fitch

Synthetic Aperture Radar

C.S. Burrus, Consulting Editor

With 93 Illustrations

Springer-Verlag
New York Berlin Heidelberg
London Paris Tokyo

J. Patrick Fitch
Engineering Research Division
Electronics Engineering Department
Lawrence Livermore National Laboratory
University of California
Livermore, CA 94550
USA

Consulting Editor
Signal Processing and Digital Filtering
C.S. Burrus
Professor and Chairman
Department of Electrical and
Computer Engineering
Rice University
Houston, TX 77251-1892
USA

Library of Congress Cataloging-in-Publication Data
Fitch, J. Patrick.
Synthetic aperture radar.
Includes bibliographies.
1. Synthetic aperture radar. I. Title.
TK6592.S95F58 1988 621.36'78 87-32110

Camera-ready copy prepared by the author using LaTeX.
Printed and bound by Arcata Graphics/Halliday, West Hanover, Massachusetts.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-96665-X Springer-Verlag New York Berlin Heidelberg
ISBN 3-540-96665-X Springer-Verlag Berlin Heidelberg New York

Synthetic Aperture Radar

PREFACE

Radar, like most well developed areas, has its own vocabulary. Words like Doppler frequency, pulse compression, mismatched filter, carrier frequency, in-phase, and quadrature have specific meaning to the radar engineer. In fact, the word radar is actually an acronym for RAdio Detection And Ranging. Even though these words are well defined, they can act as road blocks which keep people without a radar background from utilizing the large amount of data, literature, and expertise within the radar community. This is unfortunate because the use of digital radar processing techniques has made possible the analysis of radar signals on many general purpose digital computers. Of special interest are the surface mapping radars, such as the Seasat and the shuttle imaging radars, which utilize a technique known as synthetic aperture radar (SAR) to create high resolution images (pictures). This data appeals to cartographers, agronomists, oceanographers, and others who want to perform image enhancement, parameter estimation, pattern recognition, and other information extraction techniques on the radar imagery.

The first chapter presents the basics of radar processing: techniques for calculating range (distance) by measuring round trip propagation times for radar pulses. This is the same technique that sightseers use when calculating the width of a canyon by timing the round trip delay using echoes. In fact, the corresponding approach in radar is usually called the pulse echo technique. The second chapter contains an explanation of how to combine one dimensional radar returns into two dimensional images. A specific technique for creating radar imagery which is known as Synthetic Aperture Radar (SAR) is presented. Chapter 3 presents an optical interpretation and implementation of SAR. There are many similarities between SAR and other image reconstruction algorithms; a summary of tomography and ultrasound techniques is included as Chapter 4. Although the full details of these techniques are not explained, an intuitive understanding of the physical properties of these systems is possible from having studied the radar imaging problem.

Any type of digital radar processing will involve many techniques used in the signal processing community. Therefore a summary of the basic theorems of digital signal processing is given in Appendix A. The purpose of including this material is to introduce a consistent notation and to explain some of the simple tools used when processing radar data. Readers unfamiliar with the concepts of linear systems, circular convolution, and discrete Fourier transforms should skim this Appendix initially and refer to it as necessary. Matched filters are important in both pulse echo radar and SAR imaging: Appendices B and C discuss the statistical properties and digital implementation strategies for matched filters.

The approach in these notes is to present simple cases first, followed by the generalization. The objective is to get your feet wet, not to drown in vocabulary, mathematics, or notation. Usually an understanding of the geometry and physics of the problem will be more important than the mathematical details required to present the material. Standard techniques are derived or justified depending on which approach offers the most insight into the processing. Of course there are many radar related techniques which were simplified for presentation or omitted entirely—existing books and articles containing this information should be within the grasp of readers who studiously complete these notes.

These notes were initiated as part of the documentation for a software-based radar imaging system at Lawrence Livermore National Laboratory (LLNL). The code runs on a supercomputer developed in-house under the S-1 project. Some of the material presented here was also used in a graduate course at the University of California to introduce particular imaging systems and techniques. Comments by Lab researchers, faculty, and students have been helpful and encouraging during preparation of the manuscript. It is a pleasure to acknowledge my collaborators at LLNL: Steve Azevedo of the tomography research project and Jim Brase of the non-destructive signal processing program. Several of the figures in Chapter 4 were produced through joint efforts. Finally, a note of special appreciation and thanks to my wife Kathy for her encouragement and assistance with every aspect of the preparation of this manuscript.

Contents

Chapter 1

Radar Processing

1.1 Radar: A Well Defined Problem

Distance and time are equivalent. But don't panic, this is not going to be a discussion about Einstein's theory of relativity or the use of sundials. Actually, a simple example in everyday words is sufficient to describe how time measurements can be used to determine distances. Performing these types of measurements is the purpose of a radar system.

Suppose you are on a farm which has several open wells. Out of curiosity you might drop a pebble into one of the wells. After a few moments a splash or a dull thud is heard. The type of noise heard makes it possible to determine whether the well is wet or dry. The time from when the pebble was released until the noise was heard is proportional to the depth of the well. Obviously the well which has the longest drop to splash time is the deepest well. If the pebble dropping experiment is performed at each well, a plot of well depths is accumulated. The results for a six well farm can be seen in Figure 1.1. Note that the elapsed time, displayed on the horizontal axis, is sufficient for ranking the depth of any well relative to the other wells. Additional information would be required to calculate the absolute depth of any of the wells.

In a radar system the pebble is replaced by an electro-magnetic wave transmitted from an antenna. Another antenna, or in many cases the same antenna, serves as the "ear" waiting to hear a "splash". The splash corresponds to the reflection of the wave off some object in front of the antenna. By recording what the antenna receives it is possible to determine the relative position of objects which reflect the wave. A sample radar signal is given together with its return in Figure 1.2. For this example there are

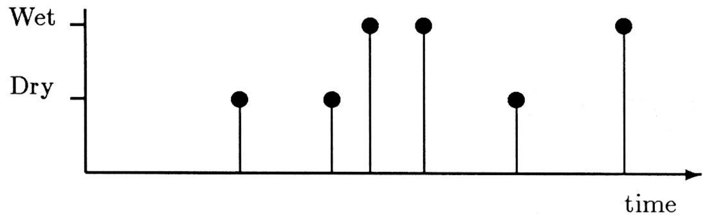

Figure 1.1: Pebble dropping experiment.

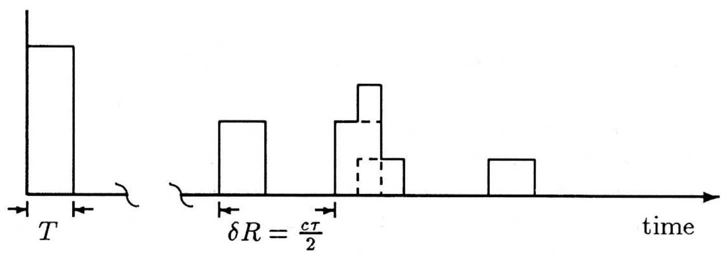

Figure 1.2: Sample radar reflection.

four reflecting objects or targets. Note that the reflections from two of the targets overlap to produce an irregularly shaped echo.

The vertical axis is a measure of how strongly the electro-magnetic wave was reflected off the objects and the horizontal axis displays the elapsed time from transmission of the pulse to reception. Some objects reflect radar signals very well and produce strong echoes—similar to the way a mirror reflects visible light. Because most objects do not reflect radar waves as efficiently as a mirror reflects light, only a portion of the incident radar energy is reflected. Because the electro-magnetic illumination is not from the visible region of the spectrum, the effect of the target on the radar wave (known as the target's signature) may not correspond with the human visual experience.

There are some important differences between the data obtained from the well experiment and the radar data. The pebble dropping experiment can be done one well at a time. In radar the number of wells is unknown and all the data is taken simultaneously. If two wells are found to have similar time delays, they would continue to be treated as separate while evaluating whether they are wet or dry. The format of the graphical display of Figure 1.1 may need to be changed, but all the information about every well is still obtained. Suppose the wells are close enough to permit simultaneously dropping a pebble in every well. Now an experiment which is similar to radar can be performed with a microphone and tape machine recording the splashes and thuds. This approach yields the same results as the sequential stone dropping experiment unless some of the splashes or thuds overlap during recording. Additional information would be required to distinguish between two simultaneous small splashes and one big splash. The problem becomes even more difficult when the possibility of a splash covering up a thud is considered. A radar antenna illuminating two objects with approximately the same time delay will create a similar problem. The reflections off the two objects may add and create a return signal that appears as one object at that time delay with a larger reflectivity. This problem is compounded by the random fluctuations in radar signals due to interfering radiation, atmospheric effects, thermal changes in the electronic components and other unpredictable degradations of the signal. The processes which contribute to the degradation from ideal transmission and reception of the radar signals are called noise. Under anything but ideal (noiseless) conditions it is impossible to determine the precise number or nature of objects in the radar return. The corresponding signal is therefore considered ambiguous—having at least two possible interpretations.

The reason ambiguous signals are received is the overlap of the returning radar pulse from closely spaced objects. Clearly, making the pulses shorter in duration, will reduce the ambiguity caused by overlapping reflections. However, as long as the pulses have some width there will be some minimum time delay between targets which is necessary to have unambiguous reception. In fact, to guarantee non-overlapping reflections, targets must be separated in time delay by at least the width of the transmitted pulse.

If the radar's transmitted pulse duration is T seconds, then the time delay between objects must be at least T seconds to prevent interference between the pulse echoes. Because radar waves are a form of electromagnetic radiation, they travel at a constant velocity c equal to the speed of light. This constant has been measured experimentally as approximately $3 \times 10^8 m/sec$. If the time delay between the reflection off two objects is τ, the light (radar pulse), had to travel an additional $c\tau$ meters for the farther object. Because $c\tau$ represents a round trip distance for the pulse, the actual physical separation of the objects is $c\tau/2$. For objects separated by less than $cT/2$, the reflections will overlap making it difficult to resolve where one target ends and the other begins. For this reason, $cT/2$ is usually called the range resolution of the radar system. These parameters were defined graphically in Figure 1.2.

Reducing the duration T of the radar pulse, improves the system's range resolution $cT/2$, which results in the radar being capable of discerning objects which are closer together. In the well experiment, the reduction in pulse width might correspond to using smaller pebbles which result in smaller (quieter and shorter duration) splashes. This creates a new problem: if the splash becomes too short, it will not create enough sound for the microphone. For radar signals, the loudness of the splash corresponds to the energy in the reflected pulse. The farther the targets are from the antenna the more energy is required in the pulse. Unfortunately it is more difficult and consequently more expensive to build a radar transmitter and receiver for a short pulse than for a long pulse of equal energy. What is needed is a transmitted pulse of sufficient duration to maintain the required energy levels together with a clever means of receiving and processing this pulse so that the data can be treated as if it were from a short pulse. In simpler terms: design a pulse so that overlapping returns from different time delays can be separated.

Just as the shape and size of the pebbles being dropped in the well can be controlled, the shape and energy in the transmitted radar light wave can be controlled. Let $u(t)$ be the function which describes the shape of the pulse. If the pulse is of length T, then $u(t) = 0$ for $t < 0$ and $t > T$. An object at time delay τ_1 will reflect a signal of the form $r_1(t) = \sigma_1 u(t - \tau_1)$, where σ_1 is some real-valued[1] number representing how strongly the object reflects the transmitted light. A second target might have a reflection of the form $r_2(t) = \sigma_2 u(t - \tau_2)$. Recall that the σ's correspond to the type of noise the pebble creates (splash or thud) and the τ's correspond to the elapsed time from drop to splash. The goal is to design a pulse shape $u(t)$ such that the returns are dissimilar for objects at different distances from the radar antenna—that is, when $\tau_1 \neq \tau_2$.

One possible measure of the similarity of the two waveforms r_1 and r_2 is the squared difference d^2 defined as

$$d^2 = \int [r_1(t) - r_2(t)]^2 dt. \tag{1.1}$$

Similar signals will have a relatively small squared difference. For example, the squared difference of a signal with itself is zero. Note that this equation defines what is meant by similar and dissimilar. There are many other

[1] In order to completely characterize the reflectivity σ of a target, a complex-valued representation with dependencies on illumination orientation and polarization is necessary. For simplicity, the orientation dependencies of σ are not considered in our analysis.

intuitively satisfying definitions for measuring similarity. For instance, the squared operator in the integral of Equation 1.1 could be replaced with an absolute value operator raised to an arbitrary power greater than one. For the mathematics which follows, however, the squared difference is perhaps the most convenient definition. Continuing with our design, we want to maximize the squared difference everywhere τ_1 does not equal τ_2. Expanding d^2 results in

$$\begin{aligned} d^2 &= \int [r_1(t) - r_2(t)]^2 dt \\ &= \int [\sigma_1 u(t-\tau_1) - \sigma_2 u(t-\tau_2)]^2 dt \\ &= \sigma_1^2 \int [u(t-\tau_1)]^2 dt + \sigma_2^2 \int [u(t-\tau_2)]^2 dt \\ &\quad - 2\sigma_1\sigma_2 \int u(t-\tau_1)u(t-\tau_2) dt. \end{aligned} \tag{1.2}$$

The first two terms which represent the energy in the reflections depend on $u(t)$ only through its energy $\int[u(t)]^2 dt$. Because the energy of the pulse can be controlled by scaling the pulse shape $u(t)$, these energy terms can be omitted from the optimization. The target reflectances, σ_1 and σ_2, are functions of the objects observed and cannot be predicted for designing the pulse shape. The maximization of d^2 has been reduced to minimizing the integral

$$A(\tau_1, \tau_2) = \int u(t-\tau_1)u(t-\tau_2) dt \tag{1.3}$$

for $\tau_1 \neq \tau_2$. For $\tau_1 = \tau_2$, the reflections overlap exactly and represent two targets at an identical time delay from the antenna. For now we will assume that these types of targets cannot be discriminated and their squared difference should be large to denote reflection from objects at the same time delay (distance from the antenna). The transmitted pulse shape $u(t)$ should therefore be selected to maximize $A(\tau_1, \tau_2)$ for $\tau_1 = \tau_2$ and should fall off quickly for $\tau_1 \neq \tau_2$. This function $A(\tau_1, \tau_2)$, which is called the autocorrelation function, equals the energy in $u(t)$ for τ_1 equal τ_2.

This analysis shows that in order to have good resolution, the autocorrelation should be small everywhere except when the time delay between signals is zero. Note that by a change of variables the autocorrelation can be written

$$A(\tau) = \int u(t)u(t+\tau) dt \tag{1.4}$$

where τ can equal $\tau_1 - \tau_2$ or $\tau_2 - \tau_1$. The limits on the integral would be over one cycle for periodic signals, all allowed values for finite length signals, and the limit of a symmetric average for signals of infinite extent. Integration determines the area under the product of $u(t)$ with $u(t+\tau)$, where $u(t+\tau)$ is a shifted version of $u(t)$. $A(\tau)$, therefore, is the value of the integral for a particular shift τ. The autocorrelation for several functions is shown in Figure 1.3. Rectangular pulses are especially helpful when developing an intuition for the shift and integrate function performed by the autocorrelation operation.

The correlation operation, which is similar to the autocorrelation, is used when two different functions are to be compared. For two possibly complex functions u and v, the correlation is defined mathematically as

$$C(\tau) = \int u^*(t)v(t+\tau)dt \tag{1.5}$$

where * is the complex conjugate operation. In order to simplify the mathematics, it is often convenient to represent radar signals as complex functions. The correlation of two functions can be considered as a plot of their average product as one function is slid past the other. An autocorrelation occurs when one function is used for both inputs to a correlation. Note that if the waveforms r_1 and r_2 in Equation 1.2 had not been scaled versions of the pulse shape u, then the maximization would have been reduced to minimizing the correlation of r_1 with r_2. When two signals which are known to be different are compared using a correlation operation, the output C is often referred to as the cross-correlation. The term correlation refers generically to the comparison of two arbitrary signals by integrating the product of the two signals for different relative shifts.

Functions of the form $\cos[2\pi(ft + .5at^2)]$ or more generally $e^{j2\pi(ft+.5at^2)}$ compress into very sharp autocorrelations. As an example, consider the autocorrelation of the complex exponential with phase $2\pi(ft + .5at^2)$. The first time derivative of this phase is $2\pi(f + at)$, which is the equation for a line of slope a and initial value f. This first derivative of the phase is called the frequency of the waveform. The 0.5 coefficient of a in the phase is used so that the frequency will be a line of slope a. This type of signal is therefore called a linear frequency modulated (linear fm) signal. The larger the value of a the faster the frequencies change. Because it is a burst of different frequencies, the linear fm signal is also called a chirp of rate a. The autocorrelation of a linear fm signal can be calculated as

$$\begin{aligned} A(\tau) &= \int e^{-j2\pi(ft+.5at^2)}\, e^{j2\pi(f(t+\tau)+.5a(t+\tau)^2)}dt \\ &= e^{j2\pi(f\tau+.5a\tau^2)} \int e^{j2\pi a\tau t}\, dt. \end{aligned} \tag{1.6}$$

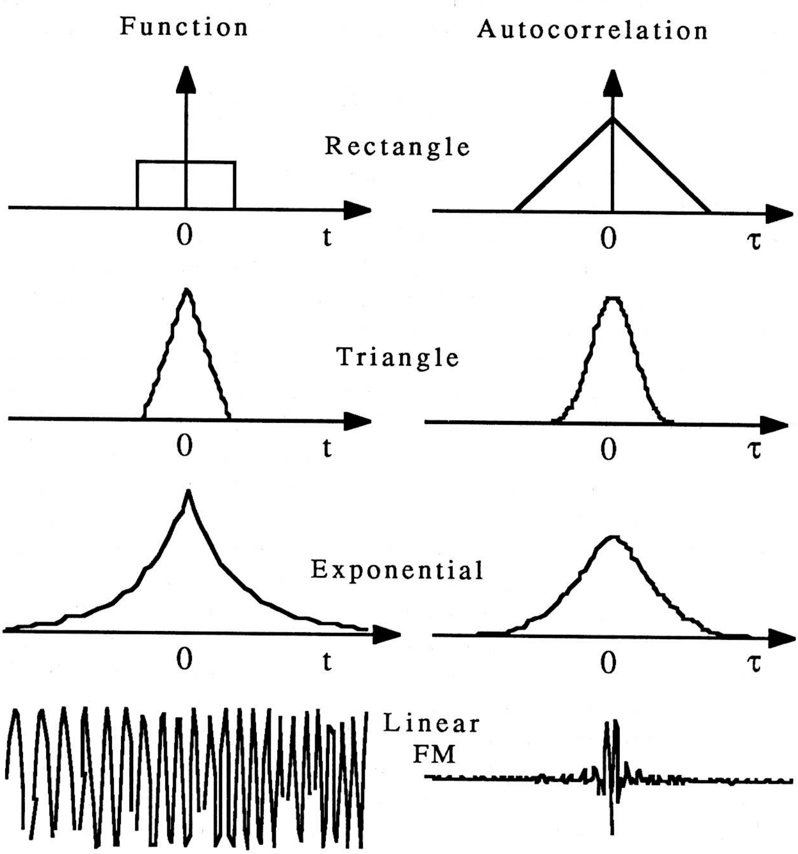

Figure 1.3: Autocorrelation of several functions.

In order to evaluate the integral, let the complex exponential be a pulse of duration T beginning at time T_0. This results in

$$A(\tau) = e^{j2\pi\tau[f+a(T_0+.5T)]} \frac{\sin[\pi a\tau(T-|\tau|)]}{\pi a\tau} \text{ for } -T \leq \tau \leq T. \quad (1.7)$$

The initial frequency $f + aT_0$ contributes to the autocorrelation as a phase factor. The shape of the magnitude is determined by a and T. The general shape of the autocorrelation function is perhaps easier to see when it is rewritten as

$$A(\tau) = \Phi \cdot (T - |\tau|) \cdot \frac{\sin[\pi a\tau(T-|\tau|)]}{\pi a\tau(T-|\tau|)} \quad (1.8)$$

where Φ is the unit magnitude phase contribution $e^{j2\pi\tau[f+a(T_0+.5T)]}$. The $(T - |\tau|)$ term is a triangle function weighting the $\sin(x)$ over x or sinc function which follows. The width of the main lobe of this function is approximately $2/aT$. This can be seen in Figure 1.3 where the magnitude of $A(\tau)$ is plotted or it can be derived by using approximations to solve for the first zero crossings of the function. Because the function falls off quickly the time duration of the autocorrelation of this signal is usually considered to be approximately $1/aT$. The spike in the autocorrelation function can be made narrower by increasing aT. Recall from the original equation for $u(t)$ that a is the rate frequencies change in the pulse and T is the pulse duration. This means that the product aT, which equals the bandwidth of the signal, is an indicator of the narrowness of the autocorrelation function. Signals where the product aT^2 is large are called large time-bandwidth functions (T is the time-duration and aT is the bandwidth). The number aT^2 also represents the compression ratio defined as the initial pulse length T divided by the final pulse width $1/aT$. The linear fm chirp was chosen for this example because it is often selected as the high time-bandwidth function used as the transmitted waveform in radar systems. The impulse like shape of the autocorrelation function and the ease of electronically generating this type of signal both contribute heavily to its popularity. Practical experience has also shown this pulse shape to be relatively insensitive to scale changes which occur during echoing. In addition, the chirp demonstrates the accepted radar axiom that the system range resolution can be improved by increasing the bandwidth of the transmitted pulse.

Selection of a pulse shape with an autocorrelation function which is large near zero and falls rapidly implies that returns which partially overlap can be separated as reflections from different objects. In fact, the autocorrelation function not only defines the measure of how similar partially overlapping returns will be, but also provides a method for separating these reflections. For a pulse shape $u(t)$ and received signal of the form

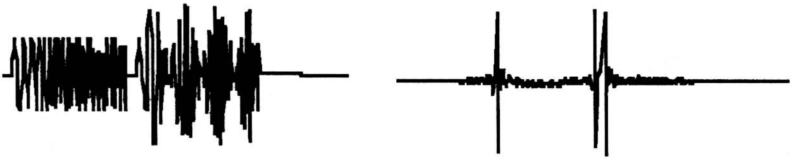

Figure 1.4: Detection of pulses using the autocorrelation.

$r(t) = \sigma u(t-\tau)$, the receiver that implements a correlation for complex signals is given by

$$y(t) = \int u^*(s)\ r(s+t)ds = \sigma \int u^*(s)\ u(s+t-\tau)ds = \sigma A(t-\tau) \quad (1.9)$$

The output $y(t)$ of the receiver will be large when t equals τ and will be small otherwise—assuming the pulse shape $u(t)$ was selected using the correlation criteria. This means that the output of this receiver will have spikes associated with time delays which correspond to reflecting objects. The transmitted pulse waveform $u(t)$ has been compressed to a spike which has the shape of its autocorrelation function $A(\tau)$. An example of a reflected signal with overlapping returns from targets with equal reflectivity is given in Figure 1.4. Because the correlation receiver performs a linear operation, pulses reflected with more energy will result in larger spikes after reception than spikes resulting from lower energy reflections.

In general, if there are N targets reflecting energy, then there will be N spikes output from the correlation receiver with each one scaled based on the reflectivity of the associated target. To summarize, if $r(t)$ is given by $\sum_{i=1}^{N} \sigma_i u(t-\tau_i)$, then the output of the correlation receiver is

$$\begin{aligned} y(t) &= \int u^*(s)\, r(s+t)\, ds \\ &= \sum_{i=1}^{N} \int u^*(s)\, u(s+t-\tau_i)\, ds \\ &= \sum_{i=1}^{N} \sigma_i\, A(t-\tau_i) \end{aligned}$$

The linearity of the correlation operation implies that this receiver can be expressed as the convolution of the received signal with an impulse