Correlation Pattern Recognition

B. V. K. Vijaya Kumar, Abhijit Mahalanobis, and Richard D. Juday

CAMBRIDGE

CORRELATION PATTERN RECOGNITION

B. V. K. VIJAYA KUMAR
Department of Electrical and Computer Engineering
Carnegie Mellon University
Pittsburgh, PA 15213, USA

ABHIJIT MAHALANOBIS
Lockheed Martin Missiles & Fire Control
Orlando

RICHARD JUDAY
Formely of NASA Johnson Space Center
Longmont

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521571036

First published 2005

Printed in the United Kingdom at the University Press, Cambridge

Typeface Times 11/14 pt. *System* Advent 3B2 8.07f [PND]

A catalog record for this publication is available from the British Library

ISBN-13 978-0-521-57103-6
ISBN-10 0-521-57103-0 hardback

Preface

Mathematically, correlation is quite simply expressed. One begins with two functions $f(\bullet)$ and $g(\bullet)$, and determines their correlation as a third function $c(\bullet)$:

$$c(t) \triangleq \int_{-\infty}^{\infty} f(\tau)g^*(t+\tau)\,\mathrm{d}\tau$$

This simplicity is at the core of a rich technology in practical pattern recognition. For unit-energy signals (and images or higher-dimensional signals), the correlation output $c(t)$ achieves its maximum of 1 if and only if the signal $f(\tau)$ matches the signal $g(t+\tau)$ exactly for some t value. Thus, correlation is an important tool in determining whether the input signal or image matches a stored signal or image. However, the straightforward correlation operation (defined by the above equation) does not prove satisfactory in practical situations where the signals are not ideal and suffer any of the many distortions such as image rotations, scale changes, and noise. Over the last 20 years, the basic correlation operation has been improved to deal with these real-world challenges. The resulting body of concept, design methods, and algorithms can be aptly summarized as *correlation pattern recognition* (CPR).

Correlation pattern recognition, a subset of statistical pattern recognition, is based on selecting or creating a reference signal and then determining the degree to which the object under examination resembles the reference signal. The degree of resemblance is a simple statistic on which to base decisions about the object. We might be satisfied with deciding which class the object belongs to, or beyond that we might want more sophisticated information about which side we are viewing the object from – or conversely we might wish our pattern recognition to be quite independent of the aspect from which the object is viewed. Often it is critical to discriminate an object from classes that differ only subtly from the interesting class. Finally, the object may be embedded in (or surrounded by) clutter, some of whose characteristics may be similar to the interesting class. These considerations are at quite different levels, but the correlation algorithms create reference signals such that their correlation against the object produce statistics with direct information for those questions.

One of the principal strengths of CPR is the inherent robustness that results from its evaluating the whole signal at once. The signal is treated in a gestalt – CPR does not sweat the individual details. In contrast, feature-based techniques tend minutely to extract information from piecewise examination of the signal, and then compare the relationships among the features. By comparing the whole image against the template, CPR is less sensitive to small mismatches and obstructions.

For many years, the testing grounds for CPR have mainly been automatic target recognition (ATR) applications where correlation filters were developed to locate multiple occurrences of targets of interest (e.g., images of tanks, trucks, etc.) in input scenes. Clearly, processing speed is of interest in such applications, which has led to much interest in coherent optical correlators because of their ability to yield two-dimensional Fourier transforms (FTs) at the speed of light. However, the input and output devices in optical correlators have not progressed as fast as one would like and it is reasonable to say that today most image correlations are calculated digitally. Over the past few years, there has been a growing interest in the use of correlation filters for biometrics applications such as face recognition, fingerprint recognition, and iris recognition. In general, correlation filters should prove valuable in many image recognition applications.

Correlation can be implemented either in the time domain (space domain for images) or in the frequency domain. Because diffraction and propagation of coherent light naturally and conveniently produce the two-dimensional FT – and do so "at the speed of light" – early applications of coherent optical processing focused on correlation. This frequency domain approach is the reason for the use of the phrase "correlation filters." With the availability of the fast Fourier transform (FFT) algorithm and very high-speed digital processors, nowadays image correlations can be carried out routinely using digital implementations. In this book, we present both digital and optical processing approaches to correlation and have tried to indicate the differences and similarities. For example, in digital correlators, filter values may range more widely than in optical correlators where the optical devices impose constraints (e.g., that transmittance has to be a real value between 0 and 1). Another example is that the optical detectors detect only intensity (a real, positive value) whereas digital methods can freely produce and manipulate complex values. These differences have led to vigorous debates of the comparative advantages of digital and optical correlators and we hope that this energy has carried through to the book itself. We have enjoyed writing it.

Readers who are new to the correlation field may regard the superficial simplicity of the correlation paradigm to be anti-climactic and make no further attempt to grasp the versatility of the correlation pattern recognition techniques. Because the output from a matched filter is the cross-correlation of the received signal with the stored template, often correlation is simply misinterpreted as just matched filtering. We have sought to dispel this myth with a complete treatment of the diverse techniques for designing correlation filters that are anything but simple matched filters. It is well known that the filter theory finds widespread applications in controls, communications, adaptive signal processing, and audio and video applications. From a pattern recognition viewpoint, the same filtering concepts offer substantial benefits such as shift-invariance, graceful degradation, and avoidance of segmentation, not to mention computational simplicity (digitally or optically), and analytical closed-form solutions that yield optimal performance.

In putting together this book, our vision was to provide the reader with a single source that touches on all aspects of CPR. This field is a unique synthesis of techniques from probability and statistics, signals and systems, detection and estimation theory, and Fourier optics. As a result, the subject of CPR is rarely covered in traditional pattern recognition and computer vision books, and has remained elusive to the interested outsider.

The book begins with a practical introduction to CPR, and it ends with the current state of the art in computer-generated correlation filters. It discusses the sometimes seemingly abstract theories (e.g., detection theory, linear algebra, etc.) at the foundation of CPR, and it proceeds to applications. It presents the material necessary for a student to operate a first optical or digital correlator (aiming the level of the material at first-year graduate students in electrical engineering or optics programs). The book is intended to summarize recently published research and to put a usefully current overview of the discipline into the hands of the seasoned worker. In short, to take a line from Stuart L. Meyer, we are writing the book we would like to have owned as we began working in the field.

We believe that one of the main reasons that CPR is not used in more applications is that its practitioner must become familiar with some basic concepts in several fields: linear algebra, probability theory, linear systems theory, Fourier optics, and detection/estimation theory. Most students would not be exposed to such a mix of courses. Thus, Chapters 2, 3, and 4 in this book are devoted to providing the necessary background.

Chapter 2 reviews basic concepts in matrix/vector theory, simple quadratic optimization and probability theory, and random variables. Quadratic optimization will prove to be of importance in many correlation filter designs; e.g., when minimizing the output noise variance that is a quadratic function of the filter being designed. Similarly, basic results from probability theory, random variables, and random processes help us to determine how a filter affects the noise in the input.

As discussed before, correlation is implemented efficiently via the frequency domain. This shift-invariant implementation is based on ideas and results from the theory of linear systems, which is summarized in Chapter 3. This chapter reviews basic filtering concepts as well as the concept of sampling, an important link between continuous images and pixelated images. This chapter also introduces random signal processing, where a random signal is input to a deterministic linear, shift-invariant system.

The usual task of a pattern recognition system is to classify an input pattern into one of a finite number of classes (or hypotheses) and, if underlying statistics are known or can be modeled, we can use the results from detection theory to achieve goals such as minimizing classifier error rates or average cost. Another related topic is estimation theory, where the goal is to estimate an unknown parameter from the observations. One application of estimation is the estimation of a classifier error rate. Chapter 4 summarizes some basic concepts from detection and estimation theory.

Chapters 5 and 6 are aimed at introducing the various correlation filter designs. Chapter 5 introduces the basic correlation filters, which are aimed at recognizing a single image. It starts with the basic notion of matched filters and shows how its output is nothing but a correlation. But then the limitations of the matched filter are discussed and other alternatives such as optimal tradeoff filters (that tradeoff noise tolerance and correlation peak sharpness) are introduced. Performance metrics useful for characterizing correlation filters are introduced. Chapter 5 also introduces some correlation filter variants (e.g., binary phase-only filter) that were introduced because of optical device limitations.

Chapter 6 presents many advanced correlation filters (also called synthetic discriminant function or SDF filters), which are the correlation filters being used in many ATR and biometrics applications. In most of these advanced correlation filter designs, the main idea is to synthesize a filter from training images that exhibit the range of image distortions that the filter is supposed to accommodate. One breakthrough filter is the minimum average correlation energy (MACE) filter, which produces sharp correlation peaks and high discrimination. The MACE filter has been used with good success in ATR and biometrics applications. This and other advanced correlation filters are discussed in Chapter 6.

Chapters 7 and 8 are devoted to optical correlator implementations. Chapter 7 is aimed at introducing some basic optics concepts such as diffraction, propagation, interference, coherence, and polarization. This chapter also introduces the important topic of spatial light modulators (SLMs), which are the optical devices that convert electrical signals to optical signals. Historically, SLMs have been the limiting factors in the speed and capabilities of optical correlators. Nowadays, SLMs originally intended for the display industry are fueling a growth of small laboratory tinkering. For less than \$4000, a single color television projector provides three high quality (though slow) modulators of several hundred pixels on a side, along with their necessary drive electronics. Other SLMs and architectures are becoming available whose speeds are substantially higher than the 30 frames per second for conventional broadcast television. Conventional wisdom in optical filter computation does not make appropriate use of these modulators, as is now possible using the recent algorithmic advances. Many of these SLMs are potentially very powerful but are often improperly used. The algorithms now allow us to make productive use of SLM behavior that until very recently would have been regarded as difficult and inferior. These concepts are discussed in Chapter 7.

Chapter 8 provides the mathematical details as well as the algorithms for designing correlation filters that can be implemented on limited-modulation SLMs. Unlike digital designs, these designs must carefully consider the SLM constraints right from the start. Over the past few years, significant mathematical advances (in particular, applying the minimal Euclidean distance [MED] principle) have been made in the design of such limited modulation correlation filters, the topic of Chapter 8.

Finally, Chapter 9 provides a quick review of two correlation filter applications. First is the automatic recognition of targets in synthetic aperture radar (SAR) scenes and the second is the verification of face images. Some MATLAB® code is provided to illustrate the design and application of the correlation filters.

This book would not have been possible without the help of many. At the risk of offending many others who have helped, we would like to acknowledge a few in particular. B. V. K. Vijaya Kumar (BVKVK) acknowledges Professor David Casasent of Carnegie Mellon University (CMU) for introducing him to the topic of optical computers, various colleagues and students for the many advances summarized in this book, the Electrical and Computer Engineering Department at CMU for supporting this effort through a sabbatical leave, and the Carnegie Institute of Technology for the Phil Dowd Fellowship that has accelerated the completion of this book. BVKVK also acknowledges the profound positive influences of his late parents (Ramamurthy Bhagavatula and Saradamba Bhagavatula) and the immense patience and love of his wife Latha Bhagavatula. Abhijit Mahalanobis (AM) would like to acknowledge his mother and late father for their guiding hand, and his wife for her patience in not ceasing to believe in the fact that all good things must come to an end (although this book nearly proved her wrong). Richard Juday wishes to acknowledge the support that NASA's Johnson Space Center provided through a decade and a half of his work in this field, and also the contributions of literally dozens of students, visiting faculty, post-doctoral fellows, and external colleagues. Dr. Stanley E. Monroe has been a particularly steadfast contributor, advisor, critic, and friend to all whose work has touched the Hybrid Vision Laboratory.

The MathWorks, Inc., very kindly provided their state-of-the-art software, MATLAB®, which we have found very useful in developing algorithms and graphics for this book. MATLAB® is a trademark of The MathWorks, Inc., and is used with permission. The MathWorks does not warrant the accuracy of the text in this book. This book's use or discussion of MATLAB® software or related products does not constitute endorsement or sponsership by The MathWorks of a particular pedagogical approach, or particular use of the MATLAB® software.

Contents

1
Introduction

There are many daily pattern recognition tasks that humans routinely carry out without thinking twice. For example, we can recognize those that we know by looking at their face or hearing their voice. You can recognize the letters and words you are reading now because you have trained yourself to recognize English letters and words. We can understand what someone is saying even if it is slightly distorted (e.g., spoken too fast). However, human pattern recognition suffers from three main drawbacks: poor speed, difficulty in scaling, and inability to handle some recognition tasks. Not surprisingly, humans can't match machine speeds on pattern recognition tasks where good pattern recognition algorithms exist. Also, human pattern recognition ability gets overwhelmed if the number of classes to recognize becomes very large. Although humans have evolved to perform well on some recognition tasks such as face or voice recognition, except for a few trained experts, most humans cannot tell whose fingerprint they are looking at. Thus, there are many interesting pattern recognition tasks for which we need machines.

The field of machine learning or pattern recognition is rich with many elegant concepts and results. One set of pattern recognition methods that we feel has not been explained in sufficient detail is that of correlation filters. One reason why correlation filters have not been employed more for pattern recognition applications is that their use requires background in and familiarity with different disciplines such as linear systems, random processes, matrix/vector methods, statistical decision theory, pattern recognition, optical processing, and digital signal processing. This book is aimed at providing such background as well as introducing the reader to state-of-the-art in design and analysis of correlation filters for pattern recognition. The next two sections in this chapter will provide a brief introduction to pattern recognition and correlation, and in the last section we provide a brief outline of the rest of this book.

1.1 Pattern recognition

In pattern recognition, the main goal is to assign an observation into one of multiple classes. The observation can be a signal (e.g., speech signal), an image (e.g., an aerial view of a ground scene) or a higher-dimensional object (e.g., video sequence, hyperspectral signature, etc.) although we will use an image as the default object in this book. The classes depend on the application at hand. In automatic target recognition (ATR) applications, the goal may be to classify the input observation as either natural or man-made, and follow this up with finer classification such as vehicle vs. non-vehicle, tanks vs. trucks, one type of tank vs. another type.

Another important class of pattern recognition applications is the use of biometric signatures (e.g., face image, fingerprint image, iris image, and voice signals) for person identification. In some biometric recognition applications (e.g., accessing the automatic teller machine), we may be looking at a verification application where the goal is to see whether a stored template matches the live template in order to accept the subject as an authorized user. In other biometric recognition scenarios (e.g., deciding whether a particular person is in a database), we may want to match the live biometric to several stored biometric signatures.

One standard paradigm for pattern recognition is shown in Figure 1.1. The observed input image is first preprocessed. The goals of preprocessing depend very much on the details of the application at hand, but can include: reducing the noise, improving the contrast or dynamic range of the image, enhancing the edge information in the image, registering the image, and other application-specific processes.

A feature extraction module next extracts features from the preprocessed image. The goal of feature extraction is to produce a few descriptors to capture the essence of an input image. The number of features is usually much smaller than the number of pixels in that input image. For example, a 64×64 image contains 4096 numbers (namely the pixel values), yet we may be able to capture the essence of this image using only 10 or 20 features. Coming up with good features depends very much on the designer's experience in an application domain. For example, for fingerprint recognition, it is well known that features such as ridge endings and bifurcations called minutiae (shown in

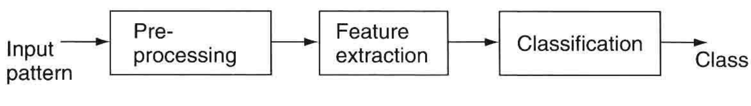

Figure 1.1 Block diagram showing the major steps in image pattern recognition

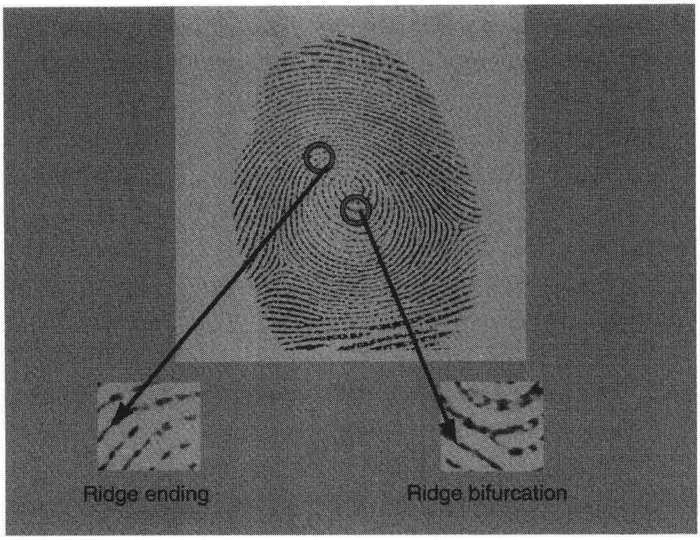

Figure 1.2 Some features used for fingerprint recognition: ridge ending (left) and ridge bifurcation (right)

Figure 1.2) are useful for distinguishing one fingerprint from another. In other pattern recognition applications, different features may be used. For example, in face recognition, one may use geometric features such as the distance between the eyes or intensity features such as the average gray scale in the image, etc. There is no set of features that is a universal set in that it is good for all pattern recognition problems. Almost always, it is the designer's experience, insight, and intuition that help in the identification of good features.

The features are next input to a classifier module. Its goal is to assign the features derived from the input observation to one of the classes. The classifiers are designed to optimize some metric such as probability of classification error (if underlying probability densities are known), or empirical error count (if a validation set of data with known ground truth[1] is available). Classifiers come in a variety of flavors including statistical classifiers, artificial neural-network-based classifiers and fuzzy logic-based classifiers. The suitability of a classifier scheme depends very much on the performance metric of interest, and on what a-priori information is available about how features appear for different classes. If we have probability density functions for various features for different classes, we can design statistical classification schemes. Sometimes, such probability density information may not be available and, instead, we may have sample feature vectors from different classes. In such a

[1] A term from remote sensing to denote the correct class of the object being tested.

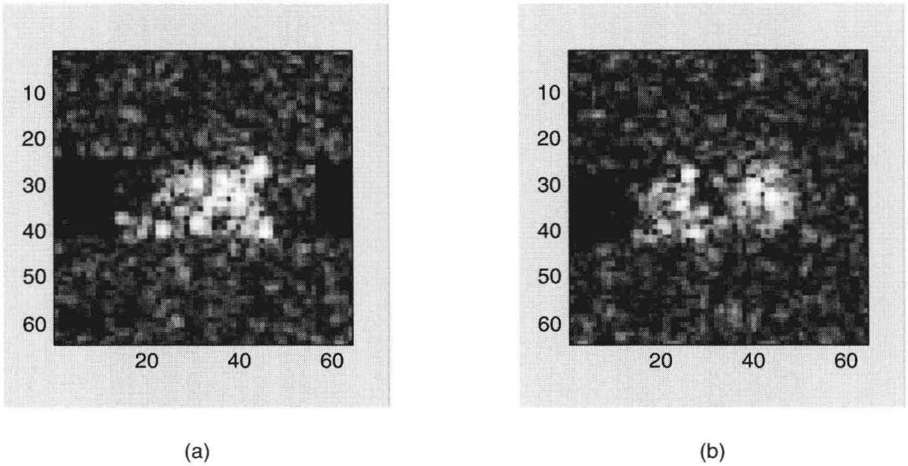

Figure 1.3 Synthetic aperture radar (SAR) images of two vehicles, (a) T72 and (b) BTR70, from the public MSTAR database [3]

situation, we may want to use trainable classifiers such as neural networks. In this book, we will not discuss these different pattern recognition paradigms. Interested readers are encouraged to consult some of the many excellent references [1, 2] discussing general pattern recognition methods.

Another important pattern recognition paradigm is to use the training data directly instead of first determining some features and performing classification based on those features. While feature extraction works well in many applications, it is not always easy for humans to identify what the good features may be. This is particularly difficult when we are facing classification problems such as the one shown in Figure 1.3, where the images were acquired using a synthetic aperture radar (SAR) and the goal is to assign the SAR images to one of two classes (tank vs. truck). Humans are ill equipped to come up with the "best" features for this classification problem. We may be better off letting the images speak for themselves, rather than imposing our judgments of what parts of SAR images are important and consistent in the way a target appears in the SAR imagery. Correlation pattern recognition (CPR) is an excellent paradigm for using training images to design a classifier and to classify a test image.

1.2 Correlation

Most readers are probably familiar with the basic concept of correlation as it arises in probability theory. We say that two random variables (RVs, the

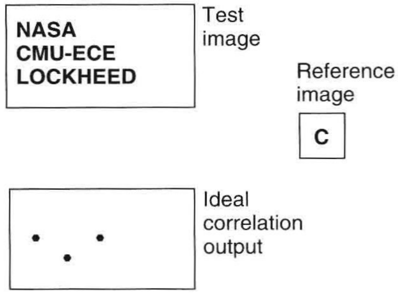

Figure 1.4 Schematic of the image correlation: reference image, test image, and ideal correlation output

concept to be explained more precisely in Chapter 2) are correlated if knowing something about one tells you something about the other RV. There are degrees of correlation and correlation can be positive or negative. The role of correlation for pattern recognition is not much different in that it tries to capture how similar or different a test object is from training objects. However, straightforward correlation works well only when the test object matches well with the training set and, in this book, we will provide many methods to improve the basic correlation and to achieve attributes such as tolerance to real-world differences or distortions (such as image rotations, scale changes, illumination variations, etc.), and discrimination from other classes.

We will introduce the concept of CPR using Figure 1.4. In this figure, we have two images: a reference image of the pattern we are looking for and a test image that contains many patterns. In this example, we are looking for the letter "C." But in other image recognition applications, the reference $r[m, n]$ can be an (optical, infrared, or SAR) image of a tank and the test image $t[m, n]$ can be an aerial view of the battlefield scene. In a biometric application, the reference may be a client's face image stored on a smart card, and the test image may be the one he is presenting live to a camera. For the particular case in Figure 1.4, let us assume that the images are binary with black regions taking on the value 1 and white regions taking on the value 0.

The correlation of the reference image $r[m, n]$ and the test image $t[m, n]$ proceeds as follows. Imagine overlaying the smaller reference image on top of the upper left corner portion of the test image. The two images are multiplied (pixel-wise) and the values in the resulting product array are summed to obtain the correlation value of the reference image with the test image for that relative location between the two. This calculation of correlation values is then repeated by shifting the reference image to all possible centerings of the reference image with respect to the test image. As indicated in the idealized correlation output in Figure 1.4, large correlation values should be obtained at the three locations where the reference matches the test image. Thus, we can locate the targets of interest by examining the correlation output for peaks and determining if those correlation peaks are sufficiently large to indicate the presence of a reference object. Thus, when we refer to CPR in this book, we are not referring to just one correlation value (i.e., one inner product of two arrays), but rather to a correlation output $c[m, n]$ that can have as many pixels as the test image. The following equation captures the cross-correlation process

$$c[m,n] = \sum_k \sum_l t[k,l]r[k+m,l+n] \tag{1.1}$$

From Eq. (1.1), we see that correlation output $c[m, n]$ is the result of adding many values, or we can say that the correlation operation is an integrative operation. The advantage of such an integrative operation is that no single pixel in the test image by itself is critical to forming the correlation output. This results in the desired property that correlation offers graceful degradation. We illustrate the graceful degradation property in Figure 1.5. Part (a) of this figure shows a full face image from the Carnegie Mellon University (CMU) Pose, Illumination, and Expression (PIE) face database [4] and part (b) shows the correlation output (in an isometric view) from a CPR system designed to search for the image in part (a). As expected, the correlation output exhibits a large value indicating that the test image indeed matches the reference image. Part (c) shows the same face except that a portion of the face image is occluded. Although the resulting correlation output in part (d) exhibits correlation peaks smaller than in part (b), it is clear that a correlation peak is still present indicating that the test image does indeed match the reference object. Some other face recognition methods (that rely on locating both eyes to start the feature extraction process) will not exhibit similar graceful degradation properties.

Another important benefit of CPR is the in-built shift-invariance. As we will show in later chapters, correlation operation can be implemented as a linear, shift-invariant filter (this shift-invariance concept will be made more precise in Chapter 3 on linear systems), which means that if the test image contains the reference object at a shifted location, the correlation output is also shifted by exactly the same amount. This shift-invariance property is illustrated in parts (e) and (f) of Figure 1.5. Part (e) shows a shifted and occluded version of the reference image and the resulting correlation output in part (f) is shifted by the same amount, but the correlation peak is still very discernible. Thus, there is no need to go through the trouble of centering the input image prior to recognizing it.