DIGITAL IMAGE PROCESSING

WILLIAM K. PRATT

DIGITAL IMAGE PROCESSING

WILLIAM K. PRATT

Professor of Electrical Engineering
Image Processing Institute
University of Southern California
Los Angeles, California

A WILEY-INTERSCIENCE PUBLICATION
JOHN WILEY & SONS
New York / Chichester / Brisbane / Toronto

Library of Congress Cataloging in Publication Data

Pratt, William K.
Digital image processing.

"A Wiley-Interscience publication."
Includes bibliographical references and index.
1. Image processing. I. Title.

TA1632.P7 1978 621.3815'42 77-20888
ISBN 0-471-01888-0

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

PREFACE

The field of digital image processing has grown considerably during the past decade with the increased utilization of imagery in myriad applications coupled with improvements in the size, speed, and cost effectiveness of digital computers and related signal processing technologies. Image processing has found a significant role in scientific, industrial, biomedical, space, and governmental applications. Such applications include the digital transmission of spacecraft imagery and personal telephone carrier television; the resolution improvement of electron microscope images and the compensation of sensor and transmission errors of pictures transmitted from deep-space probes; the automatic classification of terrain and detection of resources from earth resources satellite pictures; the formation and enhancement of biomedical imagery, including radiographs, thermograms, and nuclear scanned images; automatic map making from aerial photographs; and the detection of cracks and flaws in machine parts from industrial radiographs. In the future image processing will, no doubt, be utilized to a greater extent to aid medical practitioners in the detection and diagnosis of disease from biomedical images. Industrial applications should abound; image processing systems will analyze scenes from the "eyes" of industrial automatons to control their actions. Efficient image coding techniques offer the promise of scores of personal two-way television channels for homes and businesses. The list is limited only by imagination.

This book is intended to serve as a text for an electrical engineering or computer science graduate course in digital image processing, and as a reference for practicing engineers and scientists engaged in image processing research and development activities. Digital image processing is a broad subject encompassing studies of physics, physiology, electrical engineering, computer science, and mathematics. Readers are assumed to have an undergraduate technical background in one of these areas. Knowledge of linear system theory, vector algebra, probability, and random processes is certainly beneficial, but not absolutely necessary.

The book is divided into six parts. Part 1 contains three chapters concerned with the characterization of continuous images. Topics covered include the mathematical representation of continuous images, the psychophysical properties of human vision, and photometry and colorimetry. In Part 2 image sampling and quantization techniques are explored along with the mathematical representation of discrete images and image quality measures. Part 3 discusses two-dimensional signal processing techniques including general linear operators, pseudoinverse operators, superposition operators, convolution operators, and unitary transform operators such as the Fourier, Hadamard, and Karhunen–Loeve transforms. The final chapter in Part 3 analyzes and compares linear filtering techniques implemented by direct convolution, Fourier transform processing, and recursive filtering.

The last three parts of the book cover the three main application areas of digital image processing. Part 4 presents a discussion of image enhancement and restoration techniques. In image enhancement, picture manipulation processes are performed to provide a more subjectively pleasing image or to convert the image to a form more amenable to human or machine analysis. Image restoration is the task of improving the fidelity of an image in the sense of compensating for image degradations. Part 5, entitled image analysis, concentrates on the general subjects of scene analysis and picture understanding. Specific topics include the isolation and measurement of image features, the detection of objects within pictures, image registration, symbolic image description, and image understanding systems. Image coding, which is the subject of Part 6, involves methods for representing monochrome and color images with a minimal number of code bits for more efficient communication and storage.

Although readers should find this book reasonably comprehensive, many important topics allied to the field of digital image processing have been omitted to limit the size of the book. Among the most prominent omissions are the topics of pattern recognition and classification, digital holography, and image projection reconstruction. References to these topics are provided in Appendix 1.

WILLIAM K. PRATT

Malibu, California
January, 1978

ACKNOWLEDGMENTS

Image processing research at the University of Southern California began in 1962 on a very modest scale, but increased in size and scope with the attendant international interest in the field. In 1971 Dr. Zohrab A. Kaprielian, Dean of the School of Engineering and Vice President of Academic Research and Administration, announced the establishment of the USC Image Processing Institute. Presently, the Image Processing Institute comprises one of the world's largest research groups in image processing with excellent physical facilities. This environment has contributed significantly to this book. I am grateful to Dr. Kaprielian for his role in providing University support of image processing research. Also, I wish to acknowledge the following past and present members of the Institute's scientific staff who rendered invaluable assistance and advice in the preparation of the manuscript: Harry C. Andrews, Lee D. Davisson, Werner Frei, Ali Habibi, Ernest L. Hall, Ronald S. Hershel, Anil K. Jain, Richard P. Kruger, Nasser E. Nahi, Ramakant Nevatia, Guner Robinson, Alexander A. Sawchuk, William B. Thompson, and Lloyd R. Welch. In addition, I sincerely acknowledge the technical help of my graduate students during the writing of the book: Ikram Abdou, Behnam Ashjari, Wen-hsiung Chen, Faramarz Davarian, Michael Huhns, Clanton Mancill, Nelson Mascarenhas, John Roese, Clifford Reader, and Robert Wallis.

This book is an outgrowth of notes developed for the course Image Processing taught in the USC Electrical Engineering Department. I wish to thank the many students who suffered through the early versions of the notes for their valuable comments. Also, I sincerely appreciate the review of the notes provided by Professors Harry C. Andrews, Ernest L. Hall, and Werner Frei of USC and Dr. Ali Habibi, now of TRW Corp., who used the notes when they taught the Image Processing course.

The physical preparation of any book is a formidable task. I have been extremely fortunate to have the assistance of Florence Tebbets, Joyce Seguy, and Marilyn Chan of USC, who typed the manuscript. Mr. Ray Schmidt of the USC Image Processing Institute and his staff provided photographic services for the book, and Mr. Doyle Howland of USC was responsible for drafting and artwork. Mr. Scott Johnston and Mr. Edward Kasanjian assisted greatly in computer programming and preparation of computer-generated examples contained in the book. Many of the digital images have been photographically recorded by Dr. Clifford Reader of Ford Aerospace and Communications Corp. of Palo Alto, California. The gracious assistance of these individuals is sincerely appreciated and acknowledged.

Finally, words of appreciation for my sponsor, the Information Processing Techniques Office of the Advanced Research Projects Agency. This work was supported in part by ARPA and was monitored by the Air Force Range Measurements Laboratory under Contract No. F–08606–72–C–0008 and by the Wright-Patterson Air Force Base under Contract No. F–33615–76–C–1203.

W.K.P.

CONTENTS

Part 1. Continuous Image Characterization

Part 2. Digital Image Characterization

Part 3. Discrete Two-Dimensional Linear Processing

Part 4. Image Enhancement and Restoration

Part 5. Image Analysis

Part 6. Image Coding

Appendix

Index

1 CONTINUOUS IMAGE CHARACTERIZATION

Although this book is concerned primarily with digital, as opposed to analog, image processing techniques, it must be remembered that most digital images represent continuous natural images. Exceptions are artificial digital images such as test patterns that are numerically created in the computer and images constructed by tomographic systems. Thus it is important to understand the "physics" of image formation by sensors and optical systems including human visual perception. Another important consideration is the measurement of light in order quantitatively to describe images. Finally, it is useful to establish spatial and temporal characteristics of continuous image fields which provide the basis for the interrelationship of digital image samples. These topics are covered in the following chapters. The mathematical characterization of continuous images is the subject of Chapter 1. Chapter 2 discusses the psychophysical properties of vision. Photometry and colorimetry are described in Chapter 3.

1 MATHEMATICAL CHARACTERIZATION OF CONTINUOUS IMAGES

In the design and analysis of image processing systems it is convenient, and often necessary, mathematically to characterize the image to be processed. There are two basic mathematical characterizations of interest: deterministic and statistical. In the deterministic image representation a mathematical image function is defined and point properties of the image are considered. For a statistical representation the image is specified by average properties. The following sections contain a development of the deterministic and the statistical characterization of continuous images. Although the analysis is presented in the context of visual images, many of the results can be extended to general two-dimensional time-varying signals and fields.

1.1. CONTINUOUS IMAGE REPRESENTATION

Let $C(x, y, t, \lambda)$ represent the spatial energy distribution of an image source of radiant energy at spatial coordinates (x, y), at time t, and at wavelength λ. Because light intensity is a real positive quantity, that is, because intensity is proportional to the modulus squared of the electric field, therefore the image light function is real and non-negative. Furthermore, in all practical imaging systems, there is always a small amount of background light present. The physical imaging system also imposes some restriction on the maximum brightness of an image, for example, film saturation and cathode ray tube phosphor heating. Hence it is assumed that

$$0 \leq C(x, y, t, \lambda) \leq A \tag{1.1-1}$$

where A is the maximum image brightness. A physical image is necessarily limited in extent by the imaging system and image recording media. For mathematical simplicity all images are assumed to be nonzero only over a rectangular region for which

$$-L_x \leq x \leq L_x \tag{1.1-2a}$$

$$-L_y \leq y \leq L_y \tag{1.1-2b}$$

The physical image is, of course, observable only over some finite time interval. Thus let

$$-T \leq t \leq T \tag{1.1-2c}$$

The image light function $C(x, y, t, \lambda)$ is therefore a bounded four-dimensional function with bounded independent variables. As a final restriction, it is assumed that the image function is continuous over its domain of definition.

The brightness response of a standard human observer to an image light function is commonly measured in terms of the instantaneous luminance of the light field as defined by

$$Y(x, y, t) = \int_0^{\infty} C(x, y, t, \lambda) V_S(\lambda) \, d\lambda \tag{1.1-3}$$

where $V_S(\lambda)$ represents the relative luminous efficiency function, that is, the spectral response of human vision. Similarly, the color response of a standard observer is commonly measured in terms of some set of tristimulus values that are linearly proportional to the amounts of red, green, and blue light needed to "match" a colored light. For an arbitrary red-green-blue coordinate system, the instantaneous tristimulus values are

$$R(x, y, t) = \int_0^{\infty} C(x, y, t, \lambda) R_S(\lambda) \, d\lambda \tag{1.1-4a}$$

$$G(x, y, t) = \int_0^{\infty} C(x, y, t, \lambda) G_S(\lambda) \, d\lambda \tag{1.1-4b}$$

$$B(x, y, t) = \int_0^{\infty} C(x, y, t, \lambda) B_S(\lambda) \, d\lambda \tag{1.1-4c}$$

where $R_S(\lambda)$, $G_S(\lambda)$, $B_S(\lambda)$ are spectral tristimulus values for the set of red, green, and blue primaries. The spectral tristimulus values are, in effect, the tristimulus values required to match a unit amount of spectral light at wavelength λ. In a multispectral imaging system the observed image field is modeled as a spectrally weighted integral of the image light function. The ith spectral image field is then given as

$$F_i(x, y, t) = \int_0^{\infty} C(x, y, t, \lambda) S_i(\lambda) \, d\lambda \tag{1.1-5}$$

where $S_i(\lambda)$ is the spectral response of the ith sensor.

For notational simplicity a single image function $F(x, y, t)$ is selected to represent an image field in a physical imaging system. For a monochrome imaging system the image function $F(x, y, t)$ nominally denotes the image luminance, or some converted or corrupted physical representation of the luminance, while in a color imaging system, $F(x, y, t)$ signifies one of the tristimulus values, or some function of the tristimulus value. The image function $F(x, y, t)$ is also used to denote general three-dimensional fields such as the time-varying noise of an image scanner.

In correspondence with the standard definition for one-dimensional time signals, the time average of an image function at a given point (x, y) is defined as

$$\langle F(x, y, t)\rangle_T = \lim_{T\to\infty} \left\{\frac{1}{2T}\int_{-T}^{T} F(x, y, t)L(t)\,dt\right\} \tag{1.1-6}$$

where $L(t)$ is a time weighting function. Similarly, the average image brightness at a given time is given by the spatial average

$$\langle F(x, y, t)\rangle_S = \lim_{\substack{L_x\to\infty\\ L_y\to\infty}} \left\{\frac{1}{4L_xL_y}\int_{-L_x}^{L_x}\int_{-L_y}^{L_y} F(x, y, t)\,dx\,dy\right\} \tag{1.1-7}$$

In many imaging systems, such as image projection devices, the image does not change with time, and the time variable may be dropped from the image function. For other types of systems, such as movie pictures, the image function is time sampled. It is also possible to convert the spatial variation into time variation, as in television, by an image scanning process. In the subsequent discussion the time variable is dropped from the image field notation unless specifically required.

1.2. TWO-DIMENSIONAL SYSTEMS

A two-dimensional system, in its most general form, is simply a mapping of some input set of two-dimensional functions $F_1(x, y), F_2(x, y), \dots, F_N(x, y)$ to a set of output two-dimensional functions $G_1(x, y), G_2(x, y), \dots, G_M(x, y)$ where $(-\infty < x, y < \infty)$ denotes the independent, continuous spatial variables of the functions. This mapping may be represented by the operators $\mathcal{O}_m\{\cdot\}$ for $m = 1, 2, \dots, M$, which relate the input to output set of functions by the set of equations

$$\begin{aligned}
G_1(x, y) &= \mathcal{O}_1\{F_1(x, y), F_2(x, y), \dots, F_N(x, y)\}\\
&\vdots\\
G_m(x, y) &= \mathcal{O}_m\{F_1(x, y), F_2(x, y), \dots, F_N(x, y)\}\\
&\vdots\\
G_M(x, y) &= \mathcal{O}_M\{F_1(x, y), F_2(x, y), \dots, F_N(x, y)\}
\end{aligned} \tag{1.2-1}$$

In specific cases the mapping may be many-to-few, few-to-many, or one-to-one. The one-to-one mapping is defined as

$$G(x, y) = \mathcal{O}\{F(x, y)\} \tag{1.2-2}$$

For one-dimensional physical systems in which the independent variable is time, the system output is a function of the past and the present, but not the future. Such systems are called causal. Two-dimensional systems are generally noncausal; the spatial variables (x, y) may be negative with respect to some reference axis.

To proceed further with a discussion of the properties of two-dimensional systems, it is necessary to direct the discourse toward specific types of operators.

1.3. SINGULARITY OPERATORS

Singularity operators are widely employed in the analysis of two-dimensional systems, especially systems that involve sampling of continuous functions. The two-dimensional Dirac delta function is a singularity operator that possesses the following properties:

$$\delta(x, y) = \begin{cases} \infty, & x = 0, y = 0 \\ 0, & \text{otherwise} \end{cases} \tag{1.3-1a}$$

$$\delta(x - \xi, y - \eta) = \begin{cases} \infty, & x = \xi, y = \eta \\ 0, & \text{otherwise} \end{cases} \tag{1.3-1b}$$

$$\int_{-\varepsilon}^{\varepsilon}\int \delta(x, y)\, dx\, dy = 1 \qquad \text{for } \varepsilon > 0 \tag{1.3-1c}$$

$$\int_{-\infty}^{\infty}\int F(\xi, \eta)\delta(x - \xi, y - \eta)\, d\xi\, d\eta = F(x, y) \tag{1.3-1d}$$

In Property (c), ε is an infinitesimally small limit of integration. Property (d) is called the Sifting property of the Dirac delta function.

The two-dimensional delta function can be decomposed into the product of two one-dimensional delta functions defined along orthonormal coordinates. Thus

$$\delta(x, y) = \delta(x)\delta(y) \tag{1.3-2}$$

where the one-dimensional delta function satisfies one-dimensional versions of Eq. 1.3-1. The delta function also can be defined as a limit on a family of functions. General examples are given below (1, p. 275).

Rectangle

$$\delta(x, y) = \lim_{\alpha \to \infty} [\alpha^2 \operatorname{rect}\{\alpha x\} \operatorname{rect}\{\alpha y\}] \tag{1.3-3a}$$

Circle

$$\delta(x, y) = \lim_{\alpha \to \infty} \left[\frac{\alpha^2}{\pi} \operatorname{circ}\{\alpha\sqrt{x^2+y^2}\}\right] \tag{1.3-3b}$$

Gaussian

$$\delta(x, y) = \lim_{\alpha \to \infty} [\alpha^2 \exp\{-\alpha^2\pi(x^2+y^2)\}] \tag{1.3-3c}$$

Sinc

$$\delta(x, y) = \lim_{\alpha \to \infty} [\alpha^2 \operatorname{sinc}\{\alpha x\} \operatorname{sinc}\{\alpha y\}] \tag{1.3-3d}$$

Bessel

$$\delta(x, y) = \lim_{\alpha \to \infty} \left[\frac{\alpha J_1(2\pi\alpha\sqrt{x^2+y^2})}{\sqrt{x^2+y^2}}\right] \tag{1.3-3e}$$

where

$$\operatorname{rect}\{x\} = \begin{cases} 1, & |x| \le \frac{1}{2} \\ 0, & |x| > \frac{1}{2} \end{cases} \tag{1.3-4a}$$

$$\operatorname{circ}\{r\} = \begin{cases} 1, & r \le 1 \\ 0, & r > 1 \end{cases} \tag{1.3-4b}$$

$$\operatorname{sinc}\{x\} = \frac{\sin\{\pi x\}}{\pi x} \tag{1.3-4c}$$

Another useful formulation of the delta function is given by the identity (2, p. 99)

$$\delta(x-\xi, y-\eta) = \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} \exp\{i[u(x-\xi) + v(y-\eta)]\}\, du\, dv \tag{1.3-5}$$

where $i = \sqrt{-1}$.

1.4. ADDITIVE LINEAR OPERATORS

A two-dimensional system is said to be an additive linear system if the system obeys the law of additive superposition. In the special case of one-to-one mappings, the additive superposition property requires that

$$\mathcal{O}\{a_1F_1(x, y) + a_2F_2(x, y)\} = a_1\mathcal{O}\{F_1(x, y)\} + a_2\mathcal{O}\{F_2(x, y)\} \quad (1.4\text{-}1)$$

where a_1 and a_2 are constants that are possibly complex. This additive superposition property can easily be extended to the general mapping of Eq. 1.2-1.

A system input function $F(x, y)$ can be represented as a sum of amplitude weighted Dirac delta functions by the Sifting integral,

$$F(x, y) = \int_{-\infty}^{\infty}\int F(\xi, \eta)\delta(x - \xi, y - \eta)\, d\xi\, d\eta \quad (1.4\text{-}2)$$

where $F(\xi, \eta)$ is the weighting factor of the impulse located at coordinates (ξ, η) in the x-y plane as shown in Figure 1.4-1. Then, if the output of a general linear one-to-one system is defined to be

$$G(x, y) = \mathcal{O}\{F(x, y)\} \quad (1.4\text{-}3)$$

then

$$G(x, y) = \mathcal{O}\left\{\int_{-\infty}^{\infty}\int F(\xi, \eta)\delta(x - \xi, y - \eta)\, d\xi\, d\eta\right\} \quad (1.4\text{-}4a)$$

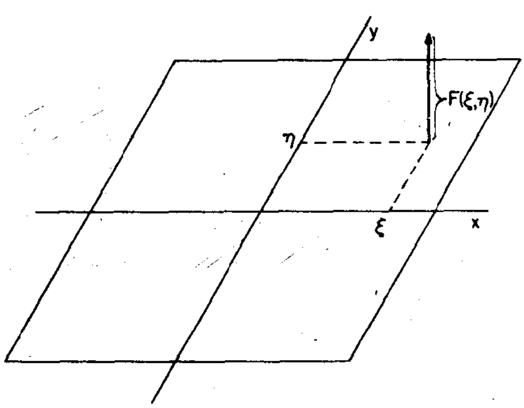

FIGURE 1.4-1. Decomposition of image function.