Advances in
ELECTRONICS and ELECTRON PHYSICS

Volume 87

ADVANCES IN ELECTRONICS AND ELECTRON PHYSICS

VOLUME 87

EDITOR-IN-CHIEF

PETER W. HAWKES

Centre National
de la Recherche Scientifique
Toulouse, France

ASSOCIATE EDITOR

BENJAMIN KAZAN

Xerox Corporation
Palo Alto Research Center
Palo Alto, California

Advances in
Electronics and Electron Physics

EDITED BY

PETER W. HAWKES

CEMES/Laboratoire d'Optique Electronique
du Centre National
de la Recherche Scientifique
Toulouse, France

VOLUME 87

ACADEMIC PRESS, INC.
Harcourt Brace & Company

Boston San Diego New York
London Sydney Tokyo Toronto

This book is printed on acid-free paper. ♾

ACADEMIC PRESS, INC.
1250 Sixth Avenue, San Diego, CA 92101-4311

United Kingdom Edition published by
ACADEMIC PRESS LIMITED
24–28 Oval Road, London NW1 7DX

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 49–7504
ISSN 0065-2539
ISBN 0-12-014729-7

PRINTED IN THE UNITED STATES OF AMERICA

93 94 95 96 BC 9 8 7 6 5 4 3 2 1

CONTENTS

Parallel Processing Methodologies for Image Processing and Computer Vision

S. YALAMANCHILI AND J. K. AGGARWAL

CONTRIBUTORS

Numbers in parentheses indicate the pages on which the authors' contributions begin.

J. B. ABBISS (1), Spectron Development Laboratories, Inc., 3535 Hyland Avenue, Tustin, California 92626

J. K. AGGARWAL (259), Department of Electrical and Computer Engineering, University of Texas, Austin, Texas 78712

M. A. FIDDY (1), Department of Electrical Engineering, University of Massachusetts Lowell, Lowell, Massachusetts 01854

U. HARTMANN (49), Institute of Thin Film and Ion Technology, KFA-Jülich, P.O. Box 1913, D-5170 Jülich, Federal Republic of Germany

B. K. JONES (201), School of Physics and Materials, Lancaster University, Lancaster LA1 4YB, United Kingdom

R. STERITI (1), Department of Electrical Engineering, University of Massachusetts Lowell, Lowell, Massachusetts 01854

S. YALAMANCHILI (259), School of Electrical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0250

This Page Intentionally Left Blank

PREFACE

Images, image formation and signal processing are the themes of this volume. The first chapter is concerned with a rapidly growing topic, neural networks for image processing. The authors, who have contributed extensively to the literature of image restoration, explain at length why neural networks are promising for restoration and how these new ideas can be implemented in practice.

A class of instruments for which some kind of image processing is indispensable consists of the near-field microscopes. In the second chapter, U. Hartmann explains the principles of scanning force microscopy and explores in detail the design problems of these instruments.

This is followed by a discussion by B.K. Jones on a topic that is endemic to all electronics and electron physics: noise. Here the emphasis is on the exploitation of noise measurements to give information about quality and reliability in electronic devices.

The volume concludes with another topic that is of the highest interest today, namely, the best ways of using parallelism in computing for image processing and computer vision. This is a relatively new subject and it is obvious that new architectures will continue to require new approaches. This survey, in which the problems, possibilities and constraints are presented very clearly, will certainly be found helpful in confronting the newest developments.

It only remains for me to thank all the contributors and to list material promised for future volumes.

FORTHCOMING ARTICLES

Electron holography	G. Ade
Image processing with signal-dependent noise	H. H. Arsenault
Parallel detection	P. E. Batson
Microscopic imaging with mass-selected secondary ions	M. T. Bernius
Magnetic reconnection	A. Bratenahl and P. J. Baum
Sampling theory	J. L. Brown
ODE methods	J. C. Butcher
Interference effects in mesoscopic structures	M. Cahay
Integer sinusoidal transforms	W. K. Cham
The artificial visual system concept	J. M. Coggins

Projection methods for image processing	P. L. Combettes
Minimax algebra and its applications	R. A. Cuninghame-Green
Corrected lenses for charged particles	R. L. Dalglish
Data structures for image processing in C	M. R. Dobie and P. H. Lewis
The development of electron microscopy in Italy	G. Donelli
Electron crystallography of organic compounds	D. L. Dorset
The study of dynamic phenomena in solids using field emission	M. Drechsler
Gabor filters and texture analysis	J. M. H. Du Buf
Amorphous semiconductors	W. Fuhs
Median filters	N. C. Gallagher and E. Coyle
Bayesian image analysis	S. and D. Geman
Theory of morphological operators	H. J. A. M. Heijmans
Electrostatic energy analysers	S. P. Karetskaya, L. G. Glikman, L. G. Beizina and Y. V. Goloskokov
Applications of speech recognition technology	H. R. Kirby
Spin-polarized SEM	K. Koike
High-definition television	M. Kunt
Fractal signal analysis using mathematical morphology	P. Maragos
Electronic tools in parapsychology	R. L. Morris
Image formation in STEM	C. Mory and C. Colliex
Phase-space treatment of photon beams	G. Nemes
Fuzzy tools for image analysis	S. K. Pal
Z-contrast in materials science	S. J. Pennycook
Electron scattering and nuclear structure	G. A. Peterson
Edge detection	M. Petrou
The wave-particle dualism	H. Rauch
Electrostatic lenses	F. H. Read and I. W. Drummond
Scientific work of Reinhold Rüdenberg	H. G. Rudenberg
Electron holography	D. Saldin
X-ray microscopy	G. Schmahl

Accelerator mass spectroscopy	J. P. F. Sellschop
Applications of mathematical morphology	J. Serra
Texture analysis	H. C. Shen
Focus-deflection systems and their applications	T. Soma
The suprenum project	O. Trottenberg
Knowledge-based vision	J. K. Tsotsos
Electron gun optics	Y. Uchikawa
Spin-polarised SEM	T. R. van Zandt and R. Browning
Morphology on graphs	L. Vincent
Cathode-ray tube projection TV systems	L. Vriens, T.G. Spanjer and R. Raue
Thin-film cathodoluminescent phosphors	A. M. Wittenberg
Diode-controlled liquid-crystal display panels	Z. Yaniv
Signal description	A. Zayezdny and I. Druckmann
The Aharonov-Casher effect	A. Zeilinger, E. Rasel and H. Weinfurter

This Page Intentionally Left Blank

ADVANCES IN ELECTRONICS AND ELECTRON PHYSICS, VOL. 87

Image Restoration on the Hopfield Neural Network

J. B. ABBISS

Spectron Development Laboratories, Inc., Tustin, California

M. A. FIDDY and R. STERITI

Department of Electrical Engineering, University of Massachusetts Lowell, Lowell, Massachusetts

I. Introduction

Images generated by an optical or electro–optical system can be degraded for a number of reasons. An intrinsic limitation is that the finite extent of

ISBN 0-12-014729-7

the entrance pupil imposes a finite upper bound on the system's spatial frequency response. The image quality of most operational imaging systems will not, however, approach this theoretical limit very closely. It is possible that the design or construction will be flawed, as in the case of the Hubble telescope, through defective manufacture, assembly, or quality assurance procedures. The detector itself may impose limitations; for example, where a CCD array is used, information is lost in the interpixel areas and image energy is integrated over the active area of each pixel. Other degrading factors will include defective pixels and noise in the CCD array and electronic subsystems.

The image restoration algorithms considered in this chapter were originally aimed at achieving performance beyond the diffraction limit, but are in fact capable of compensating simultaneously or separately for aberrations induced by the optical components and for the limitations of the detector. They are inherently robust and possess valuable noise-suppressing properties. The mathematical foundation for the methods is the fact that the spatial frequency spectrum of an object of finite extent is bandlimited. The spectrum is the Fourier transform of the object, in the coherent case, or its intensity, in the incoherent case. Image quality is a function of the way in which these spectral components have been truncated or modified. Image restoration is concerned with techniques for correcting or extrapolating the spectrum, thus recovering a closer approximation to the original object.

We describe some image restoration or superresolution algorithms that can be implemented on an artificial neural network. The motivation for this is that the kind of restoration algorithms of interest can be formulated as optimization problems, which are well suited to solution in this way. Neural network solutions to problems offer a degree of redundancy or "fault tolerance" (Rumelhart and McClelland, 1986; Kosko, 1991). From a computer science point of view, one could argue that we are really talking about massively parallel interconnections between processing units, or connectionism, and that the term *neural* has unnecessary or unhelpful additional connotations. Nevertheless, since the term neural net is so pervasive these days, we shall also use this name. There is also the expectation that either electronic or optical hardware will become increasingly available which will permit these algorithms to be executed at high speed and in parallel. We consider both types of hardware in Section VIII. Indeed, one could interpret the content of this chapter as a discussion of the types of algorithms that could be successfully implemented on fully parallel neural hardware once it is available; i.e., we can stipulate how the network can be successfully "trained" to solve this kind of problem.

The subject of neural networks is very broad, and we confine ourselves here to one type of network known as a Hopfield network (Hopfield, 1982, 1984). It is a particular example of a fully connected processing architecture in the sense that each processing element is connected to every other element by a weighted link. We describe the basic properties of this network and show how it provides a framework for interpreting a variety of optimization procedures useful in image restoration, by relating the energy function for the Hopfield network to that of a specific optimization problem. In this way *learning* as such, e.g., by example or exposure to a training set, is not necessary. More complex neural networks can contain many layers of processing nodes, which are fully interconnected between layers. Training these networks via iterative learning algorithms can be time-consuming and may be only partially successful. For these networks, however, it is still possible to specify the interconnection weights between processing nodes (Poggio and Girosi, 1989).

The image restoration problem we consider is that of improving the resolution of a low-pass filtered image or retrieving a high-resolution image from limited noisy and/or distorted spectral data. In all inverse problems of this kind, the practical constraints of real data result in a fundamental lack of uniqueness. As a result, it is necessary to adopt some kind of appropriate model for the specific problem in hand. One must recognize its biases, if any, and then solve an optimization problem, in order to determine the *best* solution for the problem consistent with the imposed constraints. It is the energy function defined for this purpose that dictates the architecture of the neural network.

For optimization problems, the question of uniqueness and the identification of algorithms that can arrive at the desired solution in a stable and repeatable fashion, despite the presence of noise in the data samples, are significant concerns. Algorithms that can effectively reach a solution but not necessarily *the* solution are of limited practical use. Also, solutions that are optimal in some prescribed sense may not be the *best* reconstructions, in terms of the quality of that image or the fidelity of the features of interest. For the restoration problems described later, it can be shown that the energy functions used do possess a unique minimum. Consequently, any procedure that reduces this energy, as a function of the image parameters, should ultimately provide a unique estimate of the reconstructed image.

The earliest example of using a Hopfield network to obtain an approximate solution to an optimization problem, the traveling salesman problem, was given by Hopfield and Tank (1985). However, there are some important optimization problems for which there is no single minimum associated with the energy function used. Under these circumstances, the procedure may stagnate at some local minimum of the energy surface. An important problem of this type is the Fourier phase retrieval problem (Fienup, 1982; Fiddy, 1987). This arises in many applications, such as imaging through turbulent or random media, intensity interferometry (for very high resolution imaging), and high-frequency scattering experiments such as x-ray diffraction studies of materials. There is, as yet, no satisfactory solution to this problem. Methods such as simulated annealing, one of the few methods proposed for locating global minima, are notoriously slow and difficult to accelerate; some kind of reliable algorithm must be found. The Hopfield artificial neural network operates in an iterative fashion, and it can be shown that the network converges to a state with the lowest local energy. As will be seen later, training a Hopfield network can lead to an associated energy function for that network with many local minima, some entirely spurious if the number of memorized states is excessive or if the self-feedback to a processing element is nonzero. These difficulties are not an issue in the signal processing applications described here, because the network is designed to have only one energy minimum.

The parameters describing the image restoration problem define an energy function that can incorporate prior knowledge about the object. This energy function can be directly mapped onto the connection strengths of a Hopfield network. Thus, once the (Hopfield) hardware is realized (or simulated), the network architecture proceeds to update neural values until a stable state is reached. In our case, that also corresponds to the image restoration problem's solution. In this way, new algorithms can be developed that allow image reconstruction to be carried out on fully parallel hardware. The hardware we consider here is a programmable Hopfield net, which can be updated *synchronously* (i.e., with simultaneous updating of all neural states) to provide a reconstructed image at high speed.

Because of the ill-posed nature of the problem, restoration methods always require some degree of prior knowledge about the image to be available. Examples of prior knowledge include low-resolution image features or edge locations; we therefore envisage this approach to be particularly suitable for remote sensing and monitoring applications, as in quality control. However, image restoration methods are well known to be ill-conditioned – hence the need to employ regularization techniques.

A. Artificial Neural Processors

There are several models of neural networks, each of which has a structure based loosely on biological nervous-system hardware (Rumelhart and McClelland, 1986). A neural network architecture consists of a very large number of simple processing elements densely interconnected by a set of weighted links. Each processing element updates its state by comparing the sum of its inputs with a prescribed threshold. The study of the properties of neural networks is a subject still somewhat in its infancy (Zornetzer *et al.*, 1991; Zurada, 1992). It is also difficult to present many concrete applications based on neural networks, since current hardware limitations reduce their practical impact. It has been suggested by Anderson and Rosenfeld (1987) that they may not become useful until cheap special-purpose parallel hardware is available. It is expected that they will prove useful in solving computationally intensive, difficult, or nonlinear problems such as those in robotic control, pattern recognition, modeling plant dynamics, etc. (Eckmiller and Malsburg, 1987; Pao, 1989). Should neural hardware become available, the question remains as to how one would make best use of a neural computer – i.e., how one should program or "train" it to perform the tasks required. The hope is that some problems for which it is difficult to find satisfactory algorithmic solutions might be amenable to solution on this kind of computing architecture, which can somehow organize itself and learn what it is expected to accomplish. In all cases, the behavior of an artificial neural network, after appropriate training, can be expressed in terms of the minimization of some appropriate energy or cost function.

For our purposes, one can describe the recovery or restoration of an image as a deconvolution exercise. It may be necessary to remove systematic degradations such as blurring or low-pass filtering effects, as well as noise. For many years, methods designed to achieve this deconvolution have been based primarily on inverse filtering, which requires high signal-to-noise ratio images (Andrews and Hunt, 1977). These methods can be computationally intensive, and techniques for speeding them up are necessary. An artificial neural network promises this possibility because of its programmable parallel-processing potential. This is not to say that other parallel-processing architectures could not successfully compete with artificial neural networks. The differences between the two options lie in the way in which the solution is computed. Our task is to find a procedure that minimizes a well-defined energy function. A conventional parallel computer relies on the execution of a search algorithm to do this, and there might be several ways in which the processors could be organized in order to obtain the result; how to partition the processors to effectively compute the solution becomes an issue. However, for the case of hardware representing a fully connected network of processors, the connection weights are modified in order to execute the minimization. Such a network is synonymous with a Hopfield neural network. If the network dynamics permit synchronous updating of the network, then rapid computations are possible.

Any deconvolution procedure that is based on a least squares approach can be formulated for high-speed processing on a fully connected computing architecture. In the following sections, we describe the mathematical basis of these restoration schemes and suggest different methods of implementation and hardware.

B. Image Deconvolution

Deconvolution is a problem that arises in many areas of imaging as well as signal processing. It is a difficult problem to solve algorithmically because it is ill-posed and can be computationally intensive; by ill-posed we mean that a solution may not exist, or it may not be unique, or it may depend discontinuously on the data. Here we will confine our discussion to the study of two-dimensional image restoration. Typical constraints that might be available to assist with the restoration are, for example, prior knowledge that the image should be real positive and bounded by some support shape. The positivity constraint can be unsuitable in the case of low-pass filtered image data; if the spectral extrapolation does not extend to infinity, the restored image will still exhibit negative side-lobes.

Deconvolution, viewed as an optimization problem, can be solved in at least two distinct ways: either directly via a matrix inverse, or iteratively. The former leads to the need to implement an algorithm based, for example, on Gaussian elimination or singular value decomposition (SVD) in order to solve a system of equations. Numerical deconvolution is an ill-conditioned procedure; large changes in the solution can result from small changes in the input data. The ill-conditioning is a manifestation in the discrete numerical case of the ill-posed nature of the problem. Steps must be taken to stabilize or regularize the solution, i.e., to ensure existence, uniqueness, and continuous dependence of the solution on the data. A different and equally robust approach to deconvolution is to solve the problem with a regularized iterative procedure, which can be shown to converge to the same solution; we cite, for example, regularized Gerchberg–Papoulis-type algorithms, which can be used for deconvolution and spectral extrapolation. The processing steps required for this are matrix operations involving the imposition of constraints between Fourier transformations.

Image restoration, by virtue of its generally multidimensional character, is inherently more suited to parallel processing architectures. Parallel processing on neural electronic hardware is in its infancy, and a few semiconductor devices have only recently become available; examples are briefly reviewed in Section VIII.A. Optical implementation offers a competing technology with potentially higher speeds and higher degrees of parallelism, because interconnections need not be physically hardwired. The concept of using optics for parallel processing is far from new, and much effort over the years has been invested in the development of such systems; see Section VIII.B. However, several problems arise when using optical hardware, the most important being the transfer of the required information onto the optical carrier. The available spatial light modulators (SLMs) that could be used have traditionally suffered from limitations in terms of speed, dynamic range, resolution, or cost.

The use of neural networks in image (or signal) classification, recognition, or understanding is steadily increasing. These are applications that the human brain is particularly good at, while current algorithms implemented largely on serial machines still leave much to be desired.

II. Neural Networks

As mentioned earlier, the Hopfield network is a fully connected network in the sense that any one of the processing elements is connected to every other one. This contrasts with layered networks, such as a multilayered perceptron, MLP (Kosko, 1991), in which processing elements are arranged in layers with connections only between neighboring layers. This difference in topology is accompanied by differences in the thresholding functions and in the procedures to find the connection strengths. The Hopfield network is implemented iteratively; the connection strengths are assigned and specify a cost function that the iterative procedure minimizes. The MLP is a one-pass network once the connection strengths have been "learned" by the minimization of an error function that quantifies the difference between current and desired output states.

Such a multilayered network is more versatile in its performance than the Hopfield model. The price paid for this is that there is no rule that is both simple and reliable for "learning" the connection strengths, i.e., by calculation of the outer product that gives the connection strengths. Usually, to determine the connection strengths in the MLP case, an iterative error backpropagation scheme is required, whose convergence properties are uncertain, but which does generally perform, eventually, in a satisfactory manner.

Backpropagation (and other learning) algorithms use error signals based on a system's response to update an initially random set of connection strengths. A comparison between the actual network output for a given input and the training example is made, and the simple difference is used to modify internal connection strengths to reduce this error. Over all of the output neurons, the mean-squared error for the example is reduced iteratively in this way. The process is computationally intensive, and learning times can be very long, almost unacceptably long, for all but