The Past, Present, and Future of Image and Multidimensional Signal Processing

The Image and Multidimensional Signal Processing Technical Committee

Edited by
Rama Chellappa,
Bernd Girod,
David C. Munson Jr.,
A. Murant Tekalp, and
Martin Vetterli

50 Years of Service to Signal Processing Professionals

I have been delighted to hear positive feedback from our readers about this special series of articles prepared by technical committees of the Signal Processing Society. On one hand, these articles present a historical review of what has been happening in the field of signal processing. On the other hand, they provide an insightful guide to the research directions that may become popular in the coming years. Many readers have mentioned that they found the latter the most exciting reason to read these articles.

In this issue, we present the contribution of the Image and Multidimensional Signal Processing Technical Committee, which covers one of the most active areas in signal processing. The field of image and multidimensional signal processing began as a field of strong theoretical framework based on mathematics, statistics, and physics. Later, with advances in computing, memory, and image-sensing technology, techniques developed for image enhancement, still and moving image compression, and image understanding (just to name a few) gave this field a solid base of practical applications. Furthermore, with the recent exploding growth of the Internet and the ubiquity of images and video, the field of image and multidimensional signal processing is becoming more and more exciting.

In this article, several current and former members of the Image and Multidimensional Signal Processing Technical Committee present an insightful review of the exciting developments in the field and report their views on various aspects of this field. Topics covered include: multidimensional signal-processing theory, image acquisition, image transforms, image modeling, image enhancement and restoration, image and video analysis, processing, coding, hardware and software implementation issues, and computed imaging. A comprehensive list of references at the end of the article provides readers with excellent pointers to further study of these topics. The readers will also find a number of special "sidebars" by experts that provide thought-provoking, some even controversial, views on emerging topics including image quality, digital cameras, fractals, model-based coding, and standards.

Now, I invite you to read this article prepared by experts in image and multidimensional signal processing—enjoy the excitement that has already been created as well as the excitement that will be created in this active research field.

Tsuhan Chen, Guest Editor
Carnegie Mellon University

1053-5888/98/$10.00©1998IEEE

Within seconds of entering the world, those of us who are blessed with the gift of sight start acquiring images of what lies around us. Processing, analyzing, and understanding of images by humans then becomes almost routine. Prior to the advent of digital computers on campuses, in research centers, and in industry, machine processing of visual and other sensory images was a daunting task. During the 1970s and 1980s, the focus was on image representation using transforms and models, image filtering and restoration, still and video compression, and image reconstruction. Although mainframes were originally used, affordable minicomputers became popular. This progress in computer hardware as well as in image acquisition and display devices enabled image processing research groups to emerge around the world. Since the mid-80s, powerful workstations and personal computers have made desktop or even laptop image processing research and technology possible. With universal access to the Internet, the future promises to be even more exciting!

From the applications point of view, digital photography, electronic imaging, digital TV, image libraries, and multimedia have given the field "a kick in the pants," so to speak, so that digital image processing has become a prime area of research not only in electrical engineering but also in other academic disciplines such as computer science, geography, health sciences, criminal justice, and remote sensing. In addition to processing images in the visible spectrum, over the last 20 years processing of radio astronomy, infrared, synthetic aperture radar (SAR), and medical images has also become very active. In particular, inventions such as computed tomography and magnetic resonance imaging have greatly nourished the field. Image processing has taken a central place in numerous applications, including, but not limited to, document image processing, optical character recognition, automatic target recognition, mapping of the earth's surface, terrain mapping, face and fingerprint recognition, teleconferencing, telemedicine, healthcare, and remote sensing. As a result, in addition to the traditional research areas mentioned above, image processing researchers are active in areas such as texture and shape analysis, image recognition, motion detection and estimation, parallel processing, hardware, and software.

One can cite many reasons for the success of the field. There is a strong underlying analytical framework based on mathematics, statistics, and physics. One can thus design well-founded, robust algorithms that eventually lead to consumer applications. As mentioned earlier, the field has also been helped immensely by the impressive advances in computer and memory technology, enabling faster processing of images, as well as in scanning and display. The field is maturing in the sense that standards are in place and new ones are being created for common utilization of technologies. More interestingly, the field is at its best stage for explosive growth in the next decade and beyond, thanks to the ubiquitous availability of video, emerging digital cameras, and inexpensive sensors. Although the field has existed for over 30 years, enthusiasts are tempted to say, "We haven't seen anything yet!"

When the responsibility of preparing this article was thrust on us, we were awed by the complexity of the task. As the field has grown so dramatically in depth and breadth, it has become increasingly difficult to be an expert. As editors of this article, we solicited and obtained contributions from leading researchers on critical topics that make up the field. It is important to acknowledge that all of us who are in this field have benefited immensely from and built upon the results of efforts by numerous researchers over the past 30 years. To them—too numerous to list here—and future generations we are indebted!

We begin this article with a section on multidimensional signal processing written by Russell Mersereau and James McClellan. This section summarizes research advances in multidimensional discrete operations, filter de-

List of Current IMDSP TC Members

Charles Bouman, *Purdue University*
Alan Bovik, *The University of Texas at Austin*
Yoram Bresler, *University of Illinois, Urbana-Champaign*
Reha Civanlar, *AT&T Labs*
Edward Delp, *Purdue University, Chair*
Eric Dubois, *Université du Québec*
Robert Gray, *Stanford University, Vice Chair*
Aggelos Katsaggelos, *Northwestern University*
Jelena Kovacevic, *Lucent Technologies*
Murat Kunt, *Swiss Federal Institute of Technology*
Reginald Lagendijk, *Delft University of Technology*
Richard Leahy, *University of Southern California*
Bernard Levy, *University of California, Davis*
Henri Maitre, *Ecole Nationale Supérieure des Télécommunications (ENST)*
Petros Maragos, *Georgia Institute of Technology*
Sanjit Mitra, *University of California, Santa Barbara*
Yrjo Neuvo, *Nokia Corporation*
Michael Orchard, *Princeton University*
Thrasyvoulos Pappas, *Lucent Technologies*
Jerry Prince, *Johns Hopkins University*
Kannan Ramchandran, *University of Illinois, Urbana*
Amy Reibman, *AT&T Labs*
Robert Safranek, *Lucent Technologies*
Ibrahim Sezan, *Sharp Laboratories of America, Inc.*
Murat Tekalp, *University of Rochester*
Michael Unser, *Swiss Federal Institute of Technology*
John Villasenor, *University of California, Los Angeles*
Andrew Yagle, *University of Michigan*
Hiroshi Yasuda, *The University of Tokyo*
Avideh Zakhor, *University of California, Berkeley*
Josiane Zerubia, *INRIA Sophia-Antipolis*

sign, and spectral estimation. Key steps at the beginning and end of the image processing algorithm chain are image acquisition and display. The next section by Jan Allebach and Joel Trussell discusses advances and future possibilities related to these steps and the role played by color. One of the key issues that impacts the evaluation of acquired and processed images is the notion of image quality. Bernice Rogowitz discusses the general notion of image quality from processing and human visual systems points of view. An article on emerging digital cameras by Majid Rabbani and Ken Parulski completes the discussion on image acquisition.

One of the most commonly performed operations on sampled images is the application of a multidimensional discrete transform. Jelena Kovacevic and Martin Vetterli trace the history of the development of numerous unitary transforms and the more recently invented transforms such as the wavelet transform. In addition to discrete transforms, statistical, stochastic and structural models have been proposed for image representation. The next section by Rama Chellappa, Rangaswamy L. Kashyap, and Azriel Rosenfeld summarizes recent progress and future research promises in the area of image models. Often, images acquired by sensors suffer from low contrast or are corrupted by sensor noise (additive or multiplicative and also signal dependent) or structural noise. To enhance the quality of images a set of algorithms known as image enhancement or filtering algorithms are sometimes applied. In the early years, mostly linear filtering methods were used. Over the last 15 years, nonlinear approaches have been found to be effective. Ed Coyle presents a succinct summary of nonlinear filters for image filtering. On many occasions, processes that introduce degradations due to motion or atmospheric blur as well as noise can be analytically modeled. One can then design algorithms to invert these processes so that restored images can be obtained. Jan Biemond and Aggelos Katsaggelos summarize nearly 30 years of work in this area.

Until the mid-70s most of the work done in image processing focussed on image filtering, enhancement, restoration, and compression. Since then, analyzing and understanding of images has become a significant area of research. Although many of the problems in image analysis and understanding (shape from shading, stereo, texture, 3-D motion and structure estimation, edge detection, and segmentation) can be formulated and solved analytically, many other problems such as feature matching and generic object recognition have been addressed by artificial intelligence based approaches. The section by Alan Bovik and Rama Chellappa summarizes analytical approaches to image segmentation and image recognition. One of the advantages of using analytical approaches is that the methodology is able to be generalized to other sensors, such as SAR, by properly accounting for the sensor phenomenology. Heuristic techniques often fail when such generalizations are attempted.

Analyzing sequences of images has emerged as a major area of research since the mid-70s. Major problems in this area are motion detection, 2-D motion estimation and 3-D motion and structure recovery. Eric Dubois presents a summary of work done in motion detection and 2-D motion estimation. Methods based on optical flow are primarily covered. In the next section, Thomas Huang details the progress made in the recovery of 3-D structure and motion estimation using discrete features such as points and lines. This section includes discussions pertinent to the motion of rigid, articulated and nonrigid objects. Given the overwhelming presence of video, several interesting problems have arisen. Retrieval of information or segments from video, segmentation of video into meaningful shots, and nonlinear editing are next discussed by A. Murat Tekalp.

The next major section is on all aspects of image and video compression by Bernd Girod, Robert Gray, Jelena Kovacevic, and Martin Vetterli. After a general introduction to source-coding principles, they cover the major compression techniques. Rate-distortion trade-offs are discussed throughout, as well as considerations and open issues that result when such techniques are employed as part of image communication systems. This section is accompanied by four brief essays on fractals (by Geoff Davis), model-based video coding (by Don Pearson), as well as a debate about coding standards with Gary Sullivan and Michael Orchard presenting opposing views. This section is followed by a brief summary by Ed Delp of the advances in hardware and software related to image processing.

We finally complete the article with an account of the progress made in computed imaging, which includes medical and SAR imaging. Although physics plays a subtle role in image formation and image restoration, it plays a very significant role in all aspects of computed imaging. This section by Richard Leahy and David Munson Jr. covers fundamental aspects of computed tomography, image reconstruction algorithms, magnetic resonance imaging, and positron emission tomography, as well as radio astronomy and SAR.

Multidimensional Digital Signal Processing

Russell M. Mersereau and James H. McClellan,
Georgia Institute of Technology

Multidimensional digital signal processing (MDSP) can be defined as the theoretical underpinnings for processing multidimensional (m-D) signals independent of a specific application. This field was most active during a 10-year period beginning in 1975. The high level of activity resulted in the publication of several textbooks [1- 5] and a collection of selected papers [6]. The operations of filtering, sampling, spectrum analysis, and signal representation are basic to all of signal processing. Understanding these operations in the m-D case formed a major activity during this time period. Most key results since that time have been directed more at the specific applications of image and video processing, medical imaging and array processing, but there remains considerable cross-fertilization among the application areas.

Algorithms for processing m-D signals can be grouped into four categories: (1) separable algorithms that use one-dimensional (1-D) operators to process the rows and columns of a multidimensional array, (2) nonseparable algorithms that borrow their derivation from their 1-D counterparts, (3) m-D algorithms that are significantly different from their 1-D counterparts, and (4) m-D algorithms that have no 1-D counterparts. The first three categories are listed in increasing order of their mathematical and computational complexity and in decreasing order of their popularity.

Separable algorithms operate on the rows and columns of an m-D signal sequentially. They have been widely used for image processing since the 1960s because they invariably require less computation than nonseparable algorithms. Examples of separable procedures include m-D discrete Fourier transforms (DFTs), discrete cosine transforms (DCTs), and fast Fourier transform (FFT)-based spectral estimation via the periodogram. In addition, separable finite impulse response (FIR) filters can be used in separable filter banks, wavelet representations for m-D signals, and decimators and interpolators for changing the sampling rate.

The second category contains algorithms that are uniquely m-D in that they cannot be decomposed into a repetition of 1-D procedures. Nonetheless, they are (with hindsight) straightforward generalizations of 1-D techniques. These can usually be derived by repeating the corresponding 1-D derivation in an m-D setting. Sampling and downsampling are one example. As in the 1-D case, bandlimited multidimensional signals can be sampled on periodic lattices [7] with no loss of information. However, unlike the 1-D case, where the minimum sampling rate depends only on the signal bandwidth, in the m-D case the geometry of the sampling lattice is also a factor. Most 1-D FIR filtering and FFT-based spectrum analysis algorithms also generalize straightforwardly to any m-D lattice [8]. Convolutions can be implemented efficiently using the m-D DFT, either on whole arrays or on subarrays. The window method for FIR filter design can be easily extended [9] and the FFT algorithm can be decomposed into a vector-radix form, which is slightly more efficient than the separable row-column approach for evaluating multidimensional DFTs [10, 11]. Nonseparable decimators and interpolators have also been derived [12, 13] that may eventually be used in subband image and video coders.

Some familiar 1-D methodologies do not extend to the m-D case or extend with so many complications that they are avoided whenever possible. One difficulty here is the absence of a factorization theorem for m-D polynomials; another is that the assumption of causality, which is commonly made for 1-D recursive systems with an implied temporal variable, is usually not appropriate for image processing applications where the variables are spatial. Taken together, both of these difficulties imply that recursive infinite impulse response (IIR) filtering is tricky except in the separable case. Conditions guaranteeing the stability of a recursive filter are much more difficult to apply than in the 1-D case [14-17], and the standard technique for designing IIR filters in the 1-D case (designing a rational squared-magnitude function and then performing a spectral factorization) is also difficult and amounts to a constrained approximation problem [18]. One alternative is to design a recursive filter in the spatial domain. This tends to produce a stable design (although stability is not guaranteed) and gives indirect control over the phase response of the filter [16].

A secondary effect of the absence of a factorization theorem is on the design of equiripple 2-D filters. The Parks-McClellan algorithm, which is a highly efficient method for designing 1-D linear phase FIR filters, does not extend. Instead, procedures similar to linear programming can be used [19], least *pth* norm methods [20] can be substituted, or equiripple filters can be designed using projections onto convex sets [21]. A widely popular alternative, which is nearly optimal in the Chebyshev sense, known as the McClellan transformation method, can be used to transform a 1-D zero-phase FIR filter into a 2-D zero-phase FIR filter [22]. Filters designed by this method also have efficient implementations [23].

Another major area of research has been in spectral estimation. Most of the "modern" spectral estimators, such as the Maximum Entropy Method [24-26] and Pisarenko's Method [24], require a new formulation based on constrained optimization. This is because their 1-D counterparts depend on factorization properties of polynomials [24]. An interesting case is the MLM (maximum likelihood method) estimate [27], where the 2-D version was developed first and then adapted to the 1-D situation.

There are also m-D algorithms that have no 1-D counterparts, especially algorithms that perform inversion and computed imaging. One of these is the operation of recovering an m-D distribution from a finite set of its projections [28], equivalently inverting a discretized Radon transform. This is the mathematical basis of computed tomography and positron emission tomography. Similar operations are exploited in spotlight-mode SAR [29]. Another imaging method, developed first for geophysical applications, is Fourier migration [30]. This efficient algorithm for image formation is now applied in wide-angle SAR systems. Finally, signal recovery methods unlike the 1-D case are possible: m-D signals with finite support can be recovered from the amplitude of their Fourier transforms [31] or from threshold crossings [32].

Image Acquisition: Hardcopy and Color

Jan P. Allebach, Purdue University, and H. Joel Trussell, North Carolina State University

Color scanners, cathode ray tube (CRT) displays, and printers are now a ubiquitous part of our everyday lives. In order to meaningfully record and process color images, it is essential to understand the mechanisms of color vision and the capabilities and limitations of color imaging devices. It is also necessary to develop algorithms that minimize the impact of device limitations and preserve color information as images are exchanged between devices. A recent review covers both of these topics and contains an extensive bibliography [33]. This section will mostly concentrate on signal-processing aspects.

In the human eye, three types of cones that govern color sensation are embedded in the retina. Each type is maximally sensitive to a different spectral band defined by the short (blue), medium (green), and long (red) wavelengths. This is the physical basis for *trichromacy*, the fact that any additive color mixture can be "matched" by proper amounts of three primary stimuli. There is a great deal of evidence, both physiological and psychophysical, for a color opponent stage in the early human visual system (HVS) in which the cone responses are linearly combined to yield three new channels: a lightness channel and two color opponent channels—red-green and blue-yellow. This provides a linear model for color systems.

However, perceptual processes, including vision, share the common property that the minimum detectable change in stimulus level is proportional to the overall magnitude of the stimulus. This principle, commonly known as Weber's Law, suggests a nonlinear stage in our model for the HVS in which signals are passed through a logarithmic transformation. Finally, the HVS also has limited ability to resolve spatial detail. This limitation is due to the optics of the ocular lens, the discrete nature of the photoreceptors on the retina, and early neural processing. The spatial contrast sensitivity function, obtained by measuring, as a function of spatial frequency, the contrast required to see a sinusoidal stimulus, provides a basis for modeling this aspect of the HVS.

Aspects of these elements of the HVS are seen in the color spaces that are commonly used in color imaging systems, such as YCrCb and CIELAB, and in the algorithms that have been developed for image rendering. Unfortunately, representation of more complex effects must consider adaptation and the colors of surrounding areas, as well as the existence of spatially tuned receptor units. The development of models that account for these factors is currently an area of active research and has resulted in new models for *color appearance*, as well as multichannel and multiresolution models for spatial vision.

Color images exist in the physical world as spatially varying spectral radiance or reflectance distributions. Color information needs to be recorded in order to process it. The processed image then needs to be physically realized by the synthesis of spatially varying spectral distributions. Both of these tasks, input and output, require substantial signal processing in their own right, regardless of the processing done on the digital image.

Image Capture

The most accurate devices for measuring large patches of color are *spectroradiometers* and *spectrophotometers*, which measure the radiant and reflectance spectra, respectively. Color image input devices, e.g. cameras and scanners, cannot now store data for more than a few spectral bands. For research purposes, 31-band (400-700 nm at 10 nm sampling) spectral images have been obtained [34]. Digital color cameras capture color images by recording electronically instead of using film. They typically have 2-D CCD arrays that capture the image in a single electronically controlled exposure. An arrangement for high-resolution spatial sampling and color filtering uses three CCD arrays with red, green, and blue color filters. For economy and in order to avoid the problems of registering multiple images, another arrangement used for low resolution is a color filter mosaic, which is overlaid on the CCD array during the semiconductor fabrication steps. Since the green region of the spectrum is perceptually more significant, such mosaics are laid out so as to have green, red, and blue recording pixels in the ratio *2:1:1* or *3:1:1*. Image-restoration techniques are then used to reconstruct the full images for each of the channels [35, 36].

Scanners are usually designed for digitizing images reproduced on paper or transparencies. For desktop scanners, speed is of greater importance than colorimetric accuracy, and therefore they usually employ an array of three linear CCD sensors with red, green, and blue color filters. In another variation on these devices, three different lamps are used in conjunction with a single linear CCD array to obtain a three-band image from three successive measurements. In actual scanners, the scanner measurement of the small area corresponding to a sampling unit is influenced by the color of the surrounding areas. Ideally, restoration schemes should be used to remove the blur from the recorded image, but because of computational limitations of scanners, this is not done on inexpensive devices.

For both cameras and scanners, the filters are usually the determining factor in colorimetric accuracy. Signal-processing methods have been applied to the problem of optimal filter design and in the choice of filter sets from off-the-shelf candidate filters [37-41]. These methods include the effects of estimation methods and noise. An interesting result is that very good colorimetric images can be obtained with only four to seven bands [42-45]; however, the bands are not obtained using the usual red, green, and blue narrowband filters.

Image Quality

Bernice Rogowitz, IBM T.J. Watson Research Center

The scientific study of perceptual image quality began with Otto Schade and his colleagues at RCA in the 1950s. Using the human contrast sensitivity function as a model of the HVS allowed technologists to treat the human observer as a simple component in the linear system analysis of the systems. A whole generation of image quality metrics emerged, providing variations on how to compute the relationship between the spatial frequency modulation of the display and the contrast sensitivity of the human observer (e.g., MTFA, SQRI).

This technique was later extended to evaluate the effectiveness of image-compression algorithms. In this case, the idea was that the image-compression algorithm created barely-detectable distortion components, and the perceptual impact of these components could be evaluated with respect to the threshold sensitivity function. Not surprisingly, even this rudimentary model provided a better characterization than stimulus-based calculations such as root mean square (RMS) error.

Over the past decade, these models have become more sophisticated, drawing increasingly on the literature on human vision. One evolutionary direction has been to include sensitivity to other characteristics of the visual stimulus (e.g., size, angle of view, color) into the parameterization of this threshold-sensitivity curve. Another enhancement of this basic image-quality model, based on vision research, has been to model the threshold response not as a single, parameterizable function, but as a set of functions, each representing a visual mechanism tuned to a particular spatial/temporal/chromatic range of visual information over a finite visual area. These are the spatial-frequency-channel, Gabor, and multiresolution models. These models have had widest impact in application to the visibility of suprathreshold artifacts, such as those produced by halftoning and low-bit-rate compression. These artifacts may not be uniformly distributed across the image, so a model with multiple sensors across space provides a good match to this problem.

One method for modeling image quality is to compute a map of how a family of visual mechanisms responds to an "original" stimulus, compare it to a map of how that same family of mechanisms responds to a processed stimulus, and use that difference as a measure of image quality. This process can be improved by comparing the results to human judgments of image quality, or even tuning the parameters of the model based on human judgments.

Psychophysical measurements of image quality, however, are themselves a complex field of study. It is relatively simple to have an observer judge one feature of an image, such as which of two images is embedded in more highly modulated noise. But for most suprathreshold artifacts and distortions, many physical attributes can contribute to the quality of the image, making it difficult to arrive at a simple judgment of which image has better image quality. One of the images might have more egregious color quantization artifacts, but the other might look very blurred. Which mechanisms are involved in the judgments we are asking our observers to make? Are they low-level mechanisms with tuned spatial, temporal, and chromatic responses? Or perhaps they are mechanisms in higher centers of the visual cortex tuned to higher-level visual features? If so, what are these mechanisms, and how are the responses of these mechanisms organized? How do the responses of these mechanisms change depending on experience?

These questions do not have simple psychophysical answers. The complex stimuli we create, and the complex judgments we want human observers to make, challenge our knowledge of human perception. Every year, our electronic technology allows us to create more complex, more interactive, more intelligent visual representations, causing us to expand our questions about how humans detect, judge, and interact with electronic imagery. We create montages of images on our display screens, 3-D shaded representations of molecules and stock prices, interactive displays that allow us to color and mark regions on the screen and have other regions update interactively on our screen, or in a virtual reality cave in another country. We create electronic systems that allow people to see interesting patterns in their data, navigate through seas of images, videos and books, and become better problem solvers because they can interact directly with the visual representations we create electronically. At each step, we are challenged to improve the technology by understanding more about human perception and cognition; and as technology develops further, we may be challenged to answer questions about human judgment, aesthetics and emotion as we try to increase image quality.

Image Output

Color output devices can broadly be classified into three types: additive, subtractive, and hybrid. Additive color systems produce color through the superposition of different spectral primaries, e.g., a CRT. Color in subtractive systems is produced through a process of removing (subtracting) unwanted spectral components from "white" light. Dye sublimation printers, color photographic prints, and color slides represent the subtractive process. Hybrid systems use a combination of additive and subtractive processes to produce color. Their main use is for color halftoning in printing, including desktop color printers.

Any practical output system is capable of producing only a limited range of colors. The range of producible colors of a device is referred to as its *gamut*. Since most subtractive and hybrid systems are nonlinear, their gamuts have irregular and complex shapes. The problem of producing the same image with two different output devices is called *gamut mapping*, and is an active research area. Stone et al. [46] laid down some principles of gamut mapping that were culled from psychophysics and accepted procedures in graphic arts. For printing images displayed on CRT monitors, they described an interactive gamut mapping strategy involving translation, scaling, and rotation of colors in CIE XYZ space. For an identical scenario, simulations of a number of clipping and compression-based gamut mapping schemes using CIELUV [47] and CIELAB [48] color spaces have also been reported.

In order to achieve colorimetric reproduction, color output devices must be *calibrated*. This means obtaining a mapping from a colorimetric, device-independent color space to the *control values* used to drive a device. The first step in calibration is to estimate an m-D mapping from control values to colors specified in the device-independent color space. This mapping is referred to as the (device) characterization. Since specified colors in a device-independent color space need to be mapped to device control values to obtain colorimetric output, it is necessary to determine the inverse of the m-D device-characterization function. If the device's operation can be accurately represented by a parametric model, the characterization is readily done by determining the model parameters from a few measurements. If no useful model exists, a purely empirical approach is necessary, in which the characterization function is directly measured over a grid of device control values. This process uses the basic signal-processing areas of estimation and interpolation. Descriptions of work on CRTs is found in [49, 50] and for halftone printers in [46, 51]. Efficient implementation of these mappings has been addressed in [52].

Image Display

Although the CRT remains the dominant medium for color image display, flat panel technologies that overcome the CRT's problems of bulk, weight, and high power consumption are developing rapidly. These new approaches, which require processing of the input signal to get the best image output, have created new research areas. Among the new displays are active and passive color liquid-crystal displays (LCDs), color light-emitting diodes (LEDs), electro-luminescent displays, and plasma displays. Most of them are additive color systems and use a mosaic of red, green, and blue "dots," though there are also some LCD devices based on the subtractive principle or on spectrally selective reflection.

Often color-display devices reduce memory requirements by restricting the number of colors that can be displayed simultaneously. Usually *8*, *12*, or *16* bits of video memory are allocated to each pixel allowing simultaneous display of 2^8, 2^{12}, or 2^{16} colors, respectively. The user has the capability to choose a palette of simultaneously displayable colors from a much larger set of colors that the device is capable of rendering. A *palletized image*, which has only the colors contained in the palette, can be stored in the video memory and rapidly displayed using look-up tables implemented in hardware.

Selecting a palette is actually a vector quantization problem. If the true color image has N distinct colors and the palette is to have K entries, then palette selection may be viewed as the process of dividing N colors into K clusters in 3-D color space, and selecting a representative color for each cluster. In many applications, computational efficiency is an important consideration, ruling out the use of iterative algorithms such as the Linde-Buzo-Gray or K-means algorithm. A number of fast tree-based quantizers have been developed for pallettization, including two K-D tree-based strategies—median cut [53] and variance-based quantization [54], binary splitting [55], and sequential scalar quantization (SSQ) [56]. SSQ is especially efficient and yields 8-bit palletized images that are virtually indistinguishable from 24-bit originals. In some applications, a fixed universal palette is required. For example, in a window-based display environment, the windows must share a common palette if they are to simultaneously have the correct color appearance. In this case, halftoning must be combined with pallettization to yield satisfactory image quality [57].

Hardcopy

The predominant technologies for desktop printers are ink jet and electrophotography. With ink jet systems, colorant is forced from a nozzle directly onto the paper. Scanning in one direction is provided by the paper-advance mechanism, and in the other direction by a system for moving the cartridge, which contains many nozzles, across the paper. Complex interlacing schemes are used to control the bleeding of ink deposited at adjacent pixels, and to reduce artifacts caused by misfiring or misaligned nozzles.

Electrophotographic printers also use the paper-advance mechanism for scanning in one direction. In the other direction, a rotating polygon mirror scans a laser beam across a photoconductive drum. The laser is modulated pixel-by-pixel to write a latent image on the drum. The drum is then brought into contact with a bed of charged toner, which is attracted to those areas of the drum that were discharged by the laser beam. Finally, the toner is transferred from the drum to paper, and then fused to the paper by a hot roller. The predominant image artifact in electrophotographic desktop printers is banding caused by fluctuations in the paper-advance velocity.

Both ink jet and electrophotographic printers employ hybrid color systems with four colorants: cyan, magenta, yellow, and black. Since composite black can also be produced by printing cyan, magenta, and yellow together, an *undercolor removal* (UCR) scheme is employed to determine the optimal combination of colorants to reproduce the desired color. Recently, ink jet printers with more than four colorants have appeared on the market. The use of additional colorants results in an expanded gamut of reproducible colors. In the past, both ink jet and electrophotographic systems functioned in a binary mode, either placing a certain amount of each colorant or none of that colorant at each printer-addressable pixel location. Recently, multiple dye load systems for ink jet and subpixel

Digital Cameras

Majid Rabbani and Ken Parulski, Eastman Kodak Company

The market for digital cameras is rapidly growing in consumer, professional, and scientific applications ranging from desktop publishing to remote sensing. Digital cameras allow users to instantly view, process, and transmit images. They use a variety of storage media, including solid-state memory cards and magnetic hard drives. These images can be viewed on home TVs or high-resolution computer monitors. Hard-copy prints can be produced using inkjet, thermal, electrophotographic, or silver halide technologies. Instead of film, most digital cameras use a single-color solid-state sensor with a two-dimensional array of discrete photosites to capture images. The photosites convert the incident photons of light into electron charge packets. Each photosite is overlaid with a color filter. For example, the Bayer color filter array (CFA) contains 50% green pixels in a checkerboard mosaic, with the remaining pixels alternating between red and blue rows. The number of columns and rows ranges from 640 x 480 for low-end "VGA" cameras to 3076 x 2048 for some professional cameras. Most cameras use "CCD" (charge coupled device) sensors, where rows of charge packets are sequentially clocked to a serial readout register that terminates in a charge-to-voltage converter. CCD sensors provide low random and fixed pattern noise, a large dynamic range, and excellent sensitivity. However, recent advances have aroused interest in active CMOS sensors that offer reduced cost, low power consumption, and random-access readout.

Prior to digital processing, the output of each pixel is amplified, analog processed to reduce the sensor's output amplifier noise, and analog to digital (A/D) converted. Professional cameras use linear A/Ds with 12 or more bits, while low-cost cameras use nonlinear 8-bit A/Ds. Most processing is performed in the digital domain, either using custom ICs or a programmable processor in the camera, or using host computer processing as the images are downloaded.

Typical processing includes an interpolation algorithm to reconstruct the missing color pixel values from the CFA pixels. White balance, which corrects for the scene illuminant, is performed by amplifying the red and blue signals to equal green for neutral (white or gray) objects. Color correction may use a 3 x 3 matrix to correct the camera's spectral sensitivities, and tone correction uses a set of lookup tables. Image sharpening, achieved by simple or adaptive spatial filters, compensates for the lens and other sources of blur and provides a subjectively sharper image. Finally, standard JPEG compression may be used to reduce the file size to less than 2 bits per pixel to save storage space.

The continued advancements in camera electronics coupled with innovations in image-processing algorithms could soon lead to intelligent cameras capable of object-based scene capture, manipulation, and storage (e.g., a camera that recognizes objects of interest in a scene and optimizes the capture parameters accordingly, followed by object-scalable compression and editing for archival in a digital library). For additional information the reader is referred to [63, 64, 65].

laser modulation techniques for electrophotography have made it possible to generate a limited number of different levels of each colorant.

With all these systems, halftoning algorithms are fundamental to creating the impression of continuous-tone. These techniques rely on the viewer's limited resolution of the spatial structure of the binary or multilevel texture printed on the page. Halftoning methods can be broadly divided into three groups according to the level of computation required.

The most efficient method is *screening*, which requires at each pixel only a single comparison between the continuous-tone image sample and a threshold from a matrix. At the next level, *error diffusion* involves a threshold operation followed by diffusion of the resulting error to neighboring pixels that have not yet been binarized. Finally, more complex search-based or iterative halftoning schemes have also been developed. Current areas of active research in digital halftoning include the use of search-based or iterative methods for design of large screens with minimally visible textures [58], modification of error diffusion to reduce artifacts [59], incorporation of printer models within the halftoning algorithm [60], [61], and development of more efficient search-based strategies [62].

Image Transforms

Jelena Kovačević, Lucent Technologies, and Martin Vetterli, Swiss Federal Institute of Technology and University of California, Berkeley

Transforms for images come in many forms, depending on the particular application [66]. However, the goal is always the same: find an alternative domain where the processing or task at hand is easier to perform. To name just one example, if convolution needs to be performed, the Fourier domain is often preferred because of the complexity reduction so achieved [67]. Image transforms are useful for fast computation of convolution and correlation, image compression, noise reduction, edge detection and segmentation, image registration and image fusion, template matching and object recognition, texture analysis and synthesis, motion estimation, object tracking, and watermarking.

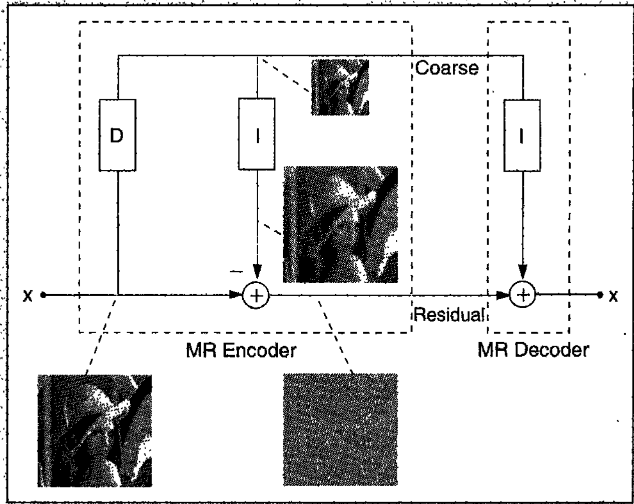
We will mostly concentrate on linear transforms since they are used in the bulk of applications in image processing. Probably the most important transform, both for theoretical and practical reasons, is the Fourier transform [68]. The DFT is a fundamental analytical tool in 1-D and m-D signal processing. The DFT is used for computational reasons since it is discrete in time and frequency and can thus be implemented exactly. Moreover, the FFT algorithm allows its efficient computation.

Many approximations have been proposed to solve the problems associated with the Karhunen-Loeve transform (KLT) [69]. The one most used is the DCT. Although the DCT [70] was originally developed as an approximation to the KLT of a first-order Gauss-Markov random process with a large correlation coefficient [69], it has proven to be a robust approximation. Many of today's image-compression standards such as JPEG [71] use the Type II DCT [70] (There are four types of DCTs, with the most popular one being of Type II). Despite its qualities, the DCT has a serious disadvantage: in order to perform the DCT, we have to block the input data and the boundaries between blocks are arbitrary. In low bit-rate compression, for example, when an image is reconstructed, the block boundaries are quite visible and create annoying "blocking artifacts."

To solve the blocking problem, a natural extension is to try to soften the block boundaries by letting the basis functions overlap; this idea led to lapped orthogonal transforms (LOT) [72]. More general linear expansions are obtained from filter banks.

The problem is now posed as finding the best linear transform (usually orthogonal) according to some given criterion, which is what the researchers in filter banks have been doing for the past decade [13, 73]. Interesting results were obtained for two-channel filter banks and tree structures derived from it.

A particular tree-structured filter bank is obtained when the lowpass branch is iterated (that is, the lower half of the spectrum is iteratively split into two parts using a two-channel filter bank). This structure leads to the discrete wavelet transform. Although discrete-time wavelets in their original form are nothing but octave-band filter banks, this connection [74] introduced both communities to a wealth of new and interesting transforms. One of the motivating factors for using wavelet-like transforms is their ability to distinguish phenomena local in both space (such as a sharp edge) and frequency (such as a pattern at a particular set of frequencies). Other interesting problems involve regularity (smoothness of the basis functions when filter banks are iterated) and search for arbitrary tilings of the time-frequency plane. This last problem led to wavelet packets (arbitrary filter bank tree structures adapted to the signal) as well as time-varying wavelet packets [75]. While many of these filter banks are applied in a separable fashion (to rows and then columns of the image), work on nonseparable filter banks and nonseparable lattices has produced interesting results as well [75]. Another class of expansions consists of overcomplete representations, where the transform domain is in some sense redundant. The classic example is an image pyramid given in Fig. 1. In this case, an $N \times N$ image is transformed into a coarse, $N/2 \times N/2$ version. An $N \times N$ approximation is interpolated as a best guess to the original, leading to an $N \times N$ error image. Perfect reconstruction is easily achieved by adding back the prediction to the error image. The representation is redundant since N^2 samples are now represented by $\frac{5N^2}{4}$ transform coefficients. This decomposition can be iterated.



▲ 1. Pyramid decomposition of an image where encoding is shown on the left and decoding is shown on the right. The operators D and I correspond to decimation and interpolation, respectively.

An even more redundant representation derives different smoothed versions with successively broader smoothing kernels, but without subsampling, as well as differences between successive smoothed versions. Such scale/space representations are useful for vision tasks such as edge detection or object recognition.

Image Models

Rama Chellappa and Azriel Rosenfeld, University of Maryland, and Rangaswamy L. Kashyap, Purdue University

Image modeling [76, 77] refers to the development of analytical representations for explaining the intensity distribution in an image. There are different types of models, each being appropriate for one or more families of applications. Thus, for applications where one is interested in compressing the amount of information in images, one needs a model that can synthesize an image, leading to generative or information-preserving models. Such models can be of various types depending on whether interaction between pixels is modeled or interaction between regions is modeled. On the other hand, if one is interested in classifying images into one of several classes, generative models, although useful, may not always be necessary. For example, one can use features derived from the joint probability density of gray levels, often characterized in the form of co-occurrence matrices, for classification of textures. Each such model captures only some aspects of the intensity distribution, and no single model can capture all the important aspects of the data. Traditionally, pixel-based models have found wide applications in 2-D spectral estimation, image and texture synthesis, image compression, image restoration, edge detection, texture classification, and image and texture segmentation. Region-based models have been found to be useful for image synthesis, image compression, and texture discrimination.

Why do we need image models? The most important reason is the abstraction they provide of the large amounts of data contained in images. Using analytical representations for images, one can develop systematic algorithms for accomplishing a particular image-related task. As an example, model-based optimal estimation-theoretic principles can be applied to find edges in textured images or remove blur and noise from degraded images. Another advantage in using image models is that one can develop a number of techniques to validate a given model for the given image. On the basis of such a validation, the performance of algorithms can be compared. It may be pointed out that for image analysis tasks such as edge detection and segmentation, algorithms are usually derived on an ad-hoc basis. Even in such cases one can identify an implicit model for the images [78].

Most statistical models for image processing and analysis treat images as 2-D data, i.e., no attempt is made to relate the (3-D) world around us and the 2-D projection on the retina. There exists a class of models, known as image-formation models, which explicitly relate the 3-D information to the 2-D brightness array through a nonlinear reflectance map by making appropriate assumptions about the surface being imaged. Such models have been the basis for computer graphics applications, can be customized for the particular sensor at hand, and have been useful for inferring shape from shading [79] and other related applications [80]. More recently, accurate prediction of object and clutter models (trees, foliage, urban scenes) has been recognized as a key component in model-based object recognition as well as change detection. Another class of models known as fractals, originally proposed in Mandelbrot [81] (see also Barnsley [82]) is good for representing images of natural scenes such as mountains, terrain, etc. Successful applications of fractals have been in the areas of image synthesis, compression and analysis.

Over the last 20 years, image models have gone through several phases of improvement. One of the earliest model-based approaches was the multivariate Gaussian model [83] used in restoration. Haralick [84] used a regression model in the form of facets for developing directional edge detectors. Due to the models used one is able to derive hypothesis tests for edge detection as well as deriving probability of detection. Deterministic 2-D sinusoidal models [85] and polynomial models for object recognition [86, 87] have also been effective.

Given that the pixels in a local neighborhood are correlated, researchers have proposed 2-D extensions of time series models for images in particular and spatial data in general since the early 50s. Generalizations have included 2-D causal [88], nonsymmetric half-plane (NSHP) [89], and noncausal models [90-98].

One of the desirable features in modeling is the ability to model nonstationary images. A two-stage approach, regarding the given image as composed of stationary patches [99], has been one of the earliest attempts to deal with nonstationarities in the image. A significant contribution to modeling nonstationary images was made in [100] using the concept of a dual lattice process, in which the intensity array is modeled as a multilevel Markov random field (MRF) and the discontinuities are modeled as line processes at the dual lattice sites interposed between the regular lattice sites. This work led to several novel image processing and analysis algorithms [101-106].

Over the last five years, image modeling has taken on a more physics-based flavor. This has become necessary due to sensors such as infrared [107], laser radar [108], SAR [109] and foliage-penetrating SAR [110], becoming more common in applications such as target recognition and image exploitation. Signatures predicted using electromagnetic scattering theory are used for model-based recognition and eliminating normal changes in the image due to weather and other predictable changes. We feel that such physics-based approaches will be the major focus in the future, not only in the nonvisible spectrum but also in the visible region. One of the key challenges in this area is validating the quality of the image model, the predicted signature of the object and the clutter surrounding the object. If effective validation can be done, then one can potentially use the synthetic images in addition to real sensor data for training target recognition algorithms. This deserves serious attention.

Image Filtering and Restoration

Jan Biemond, Delft University of Technology; Ed Coyle, Purdue University; and Aggelos K. Katsaggelos, Northwestern University

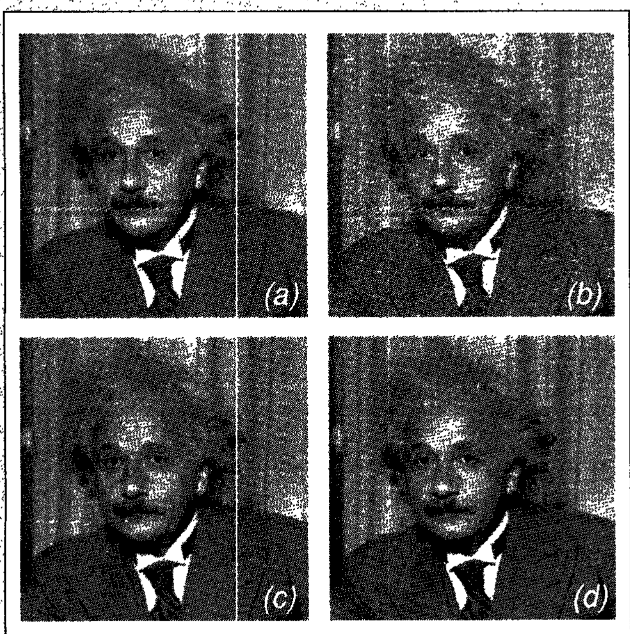
Image filtering and restoration has been one of the major research themes in image processing. Image filtering or enhancement broadly addresses the problem of improving the quality of an image. Many of the earlier image enhancement schemes were derived from common-sense or ad-hoc principles. Typical examples are averaging, transform domain filters, and contrast enhancement schemes [111, Chapter 6]. Many of these operations are linear and often yield blurred images. In the interest of keeping the image features sharper, a class of techniques known as edge-preserving smoothing filters are preferred. These filters perform less smoothing in regions around edges. Another drawback of the commonly used averaging filters is sensitivity to impulse noise.

Rank-order based image and video enhancement filtering algorithms [112, 113], such as median filters [114, 115], order statistic filters [116], weighted median filters [114], and stack filters [117], can suppress many types of non-Gaussian noise while preserving important details such as edges, lines, and scene changes.

In the past there have been two approaches to designing these filters: the structural approach and the estimation approach. The first, of which [118] is an example, requires structural/syntactical descriptions of the image and the process that has altered it. The second, of which [117] is an example, requires statistical descriptions. Recent efforts are based on transform methods or on combinations of the estimation and structural approaches [119]. We briefly review this latter approach in the context of stack filters.

Each stack filter is defined by a positive Boolean function and possesses a weak superposition property called the threshold decomposition. This superposition property is crucial to the development of fast algorithms for designing stack filters under the minimum mean absolute error criterion—the estimation approach. A perceptual error criterion, such as the visible differences predictor [120], can be used to select the weights in a weighted mean absolute error criterion in an effort to minimize distortion of the desired image structures—the structural approach. Together, these approaches produce a better result than would be possible with each one individually.

Figure 2(a) shows the original image of Einstein. Figure 2(b) shows the noisy image that will be the target of the filtering algorithm. Figure 2(c) shows the effects of applying the original algorithm, which finds a stack filter that minimizes the unweighted mean absolute error criterion (MMAE) [117]. Note that many impulsive and strip-type errors remain in the output image. Figure 2(d) shows the result of applying the new algorithm with the perceptually optimal weighted mean absolute error (WMAE) criterion [121]. Essentially all of the noise is gone, despite the high probability of its occurrence. The cost is some loss of image detail when compared with the output of the MMAE stack-filtering technique. The detail that is lost, though, is very tolerable when compared with the visual effects of impulses remaining in the image.



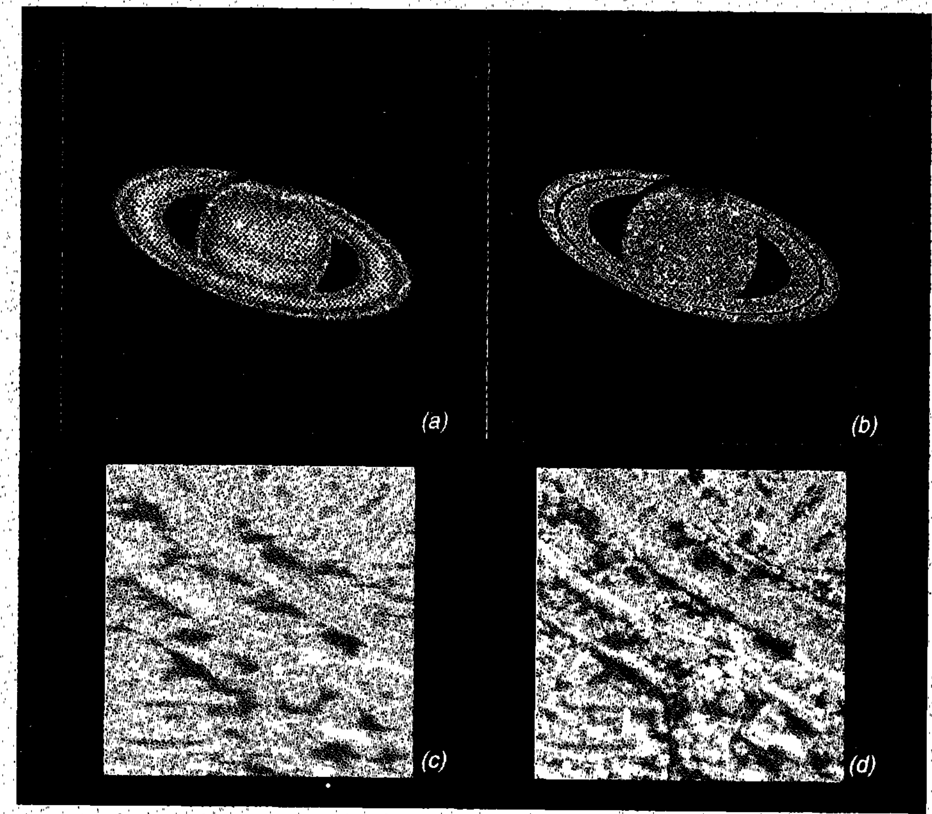
▲ 2. An example of image filtering: (a) Original image; (b) Noisy image; (c) Filtered image obtained using a stack filter that minimizes a mean absolute error criterion; (d) Filtered image obtained by minimizing a perceptually weighted mean absolute error.

Digital image restoration is the process used to recover an original scene from degraded observations. Many of the algorithms used in this area have their roots in well-developed areas of mathematics, such as estimation theory, the solution of ill-posed inverse problems, linear algebra, and numerical analysis [122, 123]. Techniques used for image restoration are oriented toward modeling the degradations, usually blur and noise, and applying an inverse procedure to obtain an approximation of the original scene.

Image restoration is distinct from image-enhancement techniques, which are designed to manipulate an image to produce results more pleasing to the observer, without making use of any particular degradation models. Image-reconstruction techniques are also generally treated separately from restoration techniques, since they operate on a set of image projections and not on a full image. Restoration and reconstruction techniques do share the same objective, however, of recovering the original image, and they end up solving the same mathematical problem, which is one of finding a solution to a set of linear or nonlinear equations.

The first encounters with digital image restoration in the engineering community were in the area of astronomical imaging. Ground-based imaging systems were subject to blurring due to the rapidly changing index of refraction of the atmosphere. Extraterrestrial observations of the Earth and the planets were degraded by motion blur as a result of slow camera shutter speeds relative to rapid spacecraft motion. Astronomical imaging is still one of the primary applications of digital image restoration today. Not only it is still necessary to restore various pictures obtained from spacecraft such as the space shuttle, but the well-publicized problems with the initial Hubble Space Telescope main mirror imperfections have provided an inordinate amount of material for the restoration community. As an example, an image of Saturn acquired by the telescope before the repair mission and a restored version of the image are shown in Fig. 3 [124].

In the area of medical imaging, image restoration has certainly played a very important part. It has also been used in law enforcement and forensic science for a number of years. Image restoration has received some notoriety in the media, and particularly in the movies of the last decade. Another current application of this field is the use of digital techniques to restore aging and deteriorating films and videos of different formats, each with its own share of particular degradations, before copying them onto new (digital) media.



3. Examples of image restoration: (a) Noisy and blurred Saturn image obtained by the Hubble Space Telescope; (b) Restored image; (c) Noisy and motion blurred image; (d) restored image using an iterative algorithm.

Perhaps the most exciting and expanding area of application for digital image restoration is that in the field of image and video coding. Removing artifacts, noise in particular [125], before compression can lead to significant gains in image quality after decoding at identical bitrates, or conversely, to a much lower bitrate at identical quality. As an example of a posteriori restoration, one can also think of reducing decoding artifacts, such as blocking and mosquito noise at very low bitrates [126, 127]. In addition channel errors exacerbate the problem and call for their concealment.

A possible classification of image restoration techniques is shown in [122]. Based on the image model assumed, existing restoration techniques are grouped into deterministic or stochastic, and further into (a) linear or nonlinear; (b) direct, recursive, or iterative; (c) spatial or frequency domain; and (d) adaptive (nonstationary) or nonadaptive (stationary) in the spatial, frequency or iteration domain.

Independent of the class a particular restoration algorithm belongs to, a basic underlying principle is that of regularization. Its purpose is to provide an analysis of an ill posed problem, such as the restoration problem, through the analysis of an associated well posed problem, whose solution will yield meaningful answers and approximations to the ill posed problems. The well posed problem is typically formulated with the use of prior knowledge about the solution. In a stochastic regularization framework, the use of a model for the original image achieves this purpose. In a deterministic regularization framework, the use of a general property or properties of the original image, such as smoothness, is typically used to achieve the same purpose. A regularized solution is then obtained by trading fidelity to the observed data with compliance to the known properties of the solution. If a constrained least-squares approach [128] is adopted, then a parameter (regularization parameter) or a set of parameters controls this trade-off. A set theoretic approach expresses the same idea: convex sets are used to describe images with certain properties, such as the smoothness property. The restored image is then defined by alternating projections onto these convex sets and converging at a point in the intersection of all the sets [129].

Fourier domain techniques were among the first ones to be used [123]. Techniques developed with the use of linear algebra concepts, such as the singular value decomposition, were later developed [123]. Due to the large size of the matrices involved, powerful computers were needed along with efficient computational methods. Two dimensional Kalman restoration filters were developed next [89], along with iterative restoration algorithms [130-132] and the projection onto convex sets [129] family of algorithms. Algorithms for the joint identification of the blur and the restoration of the degraded image were also developed [133]. The interested reader is encouraged to look at the recent review papers [122, chapter 1, 124, 130, 131, 133] and the references therein. An example of a synthetic degradation is shown in Fig. 3. The degradation is due to horizontal motion between the camera and the scene. A restored image by an iterative algorithm is shown in the same figure [134].

The various restoration techniques developed over the years have provided satisfactory solutions to many practical problems. After the more basic approaches were well developed and understood, ways to improve their performance were investigated by removing some of the global assumptions about the behavior of an image, whether from the stochastic viewpoint of stationarity or a deterministic view point of smoothness.

A key question is what will drive future research and development in the area. Will it be mathematical advances, more realistic models for the image and the degradation process, or new applications? The answer most probably will be "all of the above." Relatively recent mathematical tools such as wavelets and diffusion theory have driven the development of new restoration algorithms, which can adapt to the image characteristics. More realistic degradation models with unknown or partially known point-spread function of the degradation system, nonlinear degradations, and signal-dependent noise have been used. From the viewpoint of applications, the future of image restoration depends upon the needs of a variety of video technologies that will require processing of digital images for a number of reasons. Some of the most interesting of these applications lie in the area of consumer products. Consider, for example, hand-held video cameras that provide options such as digital focusing and adjustment for camera jerkiness. Items such as personal communication systems (PCS) that utilize video may provide a great deal of opportunity for the application of restoration ideas to coded images. It may indeed by true that in the future it will be impossible to obtain "bad pictures," but we are still confronted with our past. Unique records of history, artistic, and cultural developments of every aspect of the 20th century are stored in huge stocks of archived moving pictures. Many of the historically significant items are in a fragile state and are in desperate need of conservation and restoration. Restoration improves not only the subjective quality but also the coding efficiency when archived on new digital media with, for instance, the MPEG compression standard.

Image Analysis and Recognition

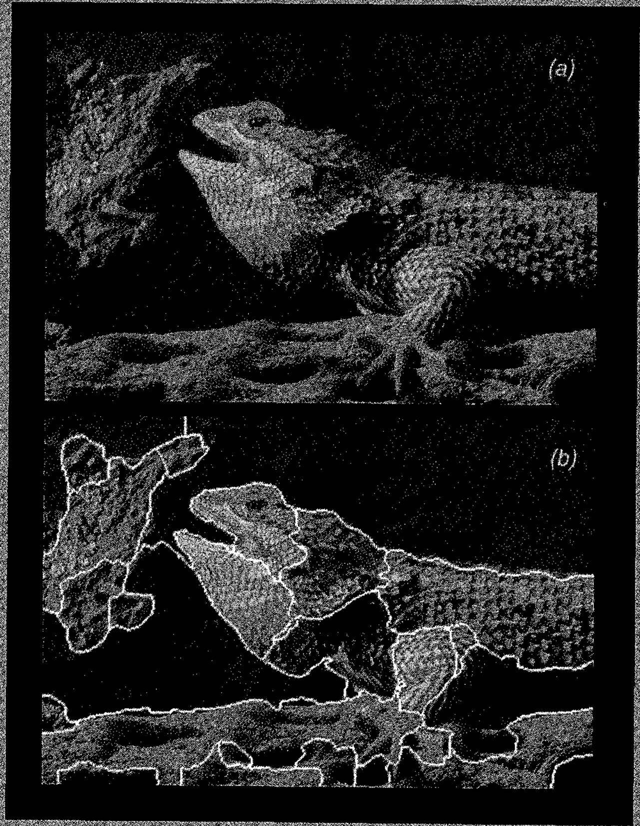
Alan C. Bovik, University of Texas at Austin, and Rama Chellappa, University of Maryland

Recent years have seen increasing interest in the development of algorithms for extracting information content from images. The images being considered may derive from a variety of sources (optical, thermal, SAR, and so on) but the goals of image analysis and recognition are often generically similar across different applications. We may define image analysis as a process of computing specific measurements or features from images, such as textural properties (which relate to the properties of materials that are being imaged), edge, line, or corner descriptors (which relate to the boundaries or other high-information contours of objects that are being imaged), region color, motion, specific shapes, etc. These measurements or features are then typically used as fundamental quantities in higher-level algorithms that attempt to detect or even recognize objects, targets, materials, or patterns. Image recognition is a higher-level process that goes beyond the simple extraction or computation of primitives: a recognition algorithm attempts to recognize objects in images, such as printed characters, military targets (tanks, planes, etc.), fingerprints, faces, and so on—the list is endless.

Before describing some of the contributions made in these areas, it is worth outlining image analysis and recognition from the standpoint of computer vision. Two-dimensional images are projections of a 3-D world in space (and 3-D video of a 4-D space-time). We define computer vision as the field that attempts to reverse this process, for example, to compute a rigid or nonrigid world given images of it. Naturally, there is much overlap between these fields, and many image-processing experts have made contributions to "vision" and vice-versa. Broadly, computer vision involves a significant element of geometric reasoning and of surface modeling, whereas image analysis and recognition typically involve image filtering and pixel-based (image-plane) processing. Of course, this division of disciplines is somewhat artificial and certainly is quite blurred.

Feature detection, which we may broadly define as a process of finding edges, lines, curves, or similar highly localized low-level structures (thus, we use the term less broadly than is used in the discipline of pattern recognition, where a feature may be any image measurement), is a basic image-analysis task that has received significant attention for over 30 years [134]. Some of the more significant signal-processing advances in recent years have been based on the observation that features of all types occur over a range of spatial scales, and so are well-analyzed and detected using multiscale techniques such as the wavelet transform [135]. The multiscale approaches to edge detection that are used today actually have their original roots in the biological and computer vision communities [136, 137], but considerable work has been accomplished that has yielded a deeper and more rigorous mathematical analysis of discontinuities using the tools of linear multiscale signal processing [138]. Complementary to this work, nonlinear multiscale techniques based on the signal-processing discipline of mathematical morphology have contributed to a deeper understanding of the properties of shape features in images [139, 140]. Simultaneously, recent recognition of dynamical systems interpretation of images, and the role of partial differential equations in analyzing images, has led to the development of multiscale feature detection strategies that are better localized, afford greater noise immunity, and most of all, take into account the processes that lead to the formation of image structure [140-143].

Texture analysis is another field that has deep roots, dating to the early days of image processing, computer vision, and pattern analysis [144]. Texture analysis encompasses several problems, including region segmentation, orientation analysis, surface identification, and others. Recent frameworks that have emerged primarily from the signal and image-processing community have emphasized two broad and complementary paradigms: methods that start from statistical models that seek to capture



4. An example of texture segmentation: (a) Original image; (b) segmentation result.

spatial relationships between image samples, and methods that use wavelet-like filters to extract spatio-spectrally localized image properties that characterize texture. The statistical approaches are typified by the use of MRFs to model image textures [145, 146], or equivalently, Gibbs distributions [100], which allow the use of Bayesian formulations to compute texture feature vectors for segmentation, classification, or recognition [147]. In these approaches also, there has been an increasing recognition of the significance of incorporating multiscale modeling and processing to improve performance [148]. Multiscale filtering approaches were originally motivated by Gabor-filter models of neural architecture in the visual cortex [149], but they have been shown to possess many optimal properties for image segmentation [150, 151]. More recently, complementary amplitude- and frequency-modulation (AM-FM) image models have been found to meld seamlessly with the filtering approach and to motivate powerful processing paradigms for analyzing nonstationary images [152-154]]. We expect that a valuable future development will be a *marriage* of the statistical and filtering frameworks, leading to a deeper unified theory of image structure. Encouraging research results in this direction may be found in [155, 156]. An example of texture segmentation [157] with applications to image retrieval is shown in Fig. 4.

Image recognition broadly refers to the problem of either recognizing objects in an image (e.g. faces) or the whole image itself (e.g. fingerprints). Applications of image-recognition techniques include optical character recognition (OCR), fingerprint and face recognition, automatic target recognition (ATR), texture recognition, etc. The main stages in the development of an image-recognition algorithm or system are preprocessing, feature extraction, and recognition. The preprocessing stage typically can be a filtering algorithm to reduce noise or a contour extraction algorithm if the goal is object recognition using shape information. For the latter problem, the feature-extraction step consists of using some mathematical representation of the boundary of the object and using the coefficients of the expansion as features.

A popular contour representation during the 1970s was based on some form of Fourier expansion [158, 159]. Some of these expansions yield representations that are insensitive to scale, translation, and rotation [158], requiring simple recognizers. Other expansions [159] that produce coefficients that are sensitive to one or more of these transformations require searching over these transforms for matching of features in the recognition stage. For problems such as face recognition, the first few coefficients of the Karhunen-Loeve expansion (the so-called eigenfaces) have been effective [160, 161]. Other features that have been effective for image recognition are moments [162, 163], coefficients of autoregressive models [164, 165], and multiscale features [166, 167].

The recognition part of an image-recognition algorithm has been typically based on statistical pattern-recognition techniques [168-170], artificial neural networks [171, 172], or matching of features [173]. Typical techniques have been based on the nearest neighbor rule [174], Bayes networks [175], multilayer perceptrons [176, 177], and radial basis functions [178, 179]. Recently, techniques that have combined segmentation and recognition stages have been shown to be effective in recognizing handwritten script [180]. A recent approach based on "Twenty Questions" [181] is promising.

One of the major applications of image recognition has been ATR [182]. Over the last 30 years, many techniques (too numerous to list here) have been proposed for target detection, recognition, and identification. Earlier efforts focussed on template matching schemes. Given the enormous number of templates that may be required even for a reasonable number of targets, due to changes in aspect, occlusion, concealment, etc., recent efforts have relied on extracting features and matching them to those stored in a database. These methods, when combined with signature prediction capabilities for the different sensors (infrared, laser radar, SAR, and foliage-penetrating SAR), make model-based approaches to the ATR problem feasible. More recently the use of hidden Markov models (HMMs) has become very popular for ATR applications. Although many years of research have gone into these efforts, many challenges remain due to the unpredictability of target and scene signatures.

Motion Estimation

Eric Dubois, INRS-Telecommunications, University of Quebec, and Thomas S. Huang, University of Illinois, Urbana-Champaign

Motion is the major information-bearing feature of time-varying imagery, so it is not surprising that most systems for processing image sequences involve motion detection or motion estimation. Motion can be classified into three major categories: rigid, articulated, and nonrigid. An articulated object is defined as an object consisting of rigid parts linked together by joints (a robot arm is an obvious example). Also, for many applications, the human body and the human hand can be approximately modeled as articulated objects. Nonrigid objects have many forms, the major ones being elastic (heart, face), sheets (clothing, paper), gas (clouds, smoke), and fluid (turbulent water flow). Two major problems we address in this section are motion detection and motion estimation. Motion detection is useful in such applications as surveillance, where the goal is in detecting motion when there should be none. Motion detection has also been used in many signal processing tasks where different actions are taken in moving and in fixed areas, e.g. in temporal filtering for noise reduction, although recently such approaches have been superseded by motion-compensated filtering. Motion estimation, on the other hand, involves the estimation of the actual velocities of projected scene points in the image plane or displacements between different frame times or 3-D motion of an object. In this section the first two problems are treated in a subsection on 2-D motion estimation. 3-D motion and structure estimation is treated in a separate subsection. In general, motion estimation has applications in motion-compensated noise reduction, sampling structure conversion, video coding, video database indexing and retrieval, human-computer interface, monitoring and surveillance, biomedical problems (analysis of heart wall motion), navigation of unmanned ground vehicles, and sensor-based manipulation.

Two-Dimensional Motion Detection and Estimation

The 3-D motion of scene surfaces relative to the camera induces a 2-D motion of the projections of scene points on the camera's image plane. There has been considerable research activity on the determination and analysis of 2-D motion over the last 20 years and several surveys have been published; two recent ones are [183] and [184]. Motion detection involves a decision as to whether temporal change in a given area of the image is due to motion or to other factors such as noise or lighting variation. The result is either a map identifying which pixels in the image are declared to be in motion, or simply a decision as to whether motion is present or not in a given image frame. A recent contribution in this area is [185]. The goal of motion estimation is to assign a motion vector to each pixel in an image frame; the collection of such motion vectors is called a motion field. There are three key components to this problem: representation of the motion field, formulation of an objective function whose minimization yields the motion estimate, and minimization of the objective function.

The most general motion-field representation assigns a motion vector independently to each pixel in the image frame, often called dense motion or optical flow [186]. This can in principle give the best approximation of the true motion and thus the best performance in certain motion-compensated filtering tasks. However, the complexity may be excessive and the cost to transmit this information in a coding context would be too high. For this reason, many other more compact representations of motion fields have been proposed. The first to be proposed and the most widely used is the block-based representation [187]. The motion field is assumed to be constant over rectangular blocks, say of size 16 by 16, and represented by a single motion vector in each block. This is currently used in most motion-compensated coders. A simple refinement of this idea is to allow variable block size; if a single motion vector does not represent the motion well over a given block, the block is divided into four. This leads to a quad-tree structure for the motion field. Other approaches do not assume that the motion field is piecewise constant. Two main categories are subsampled and parametric representations. In the first, the motion vectors are specified at critically chosen control points, and the remaining motion vectors are determined by interpolation. In the second, the motion field over a specific region is specified by a set of parameters, usually the coefficients of expansion with respect to a set of basis functions (low-order polynomials are the most popular). Moulin et al. [188] and Dufaux and Moscheni [183] discuss most of these representations.

An objective function for motion estimation is a nonnegative function of the free parameters of the motion representation whose minimization yields the desired motion estimate. The role of the objective function is to measure how well a given motion field explains the temporal change observed in the image sequence while satisfying a priori properties of motion fields such as smoothness [189]. Criteria that are widely used are the mean-absolute or mean-squared motion-compensated frame difference (MCFD). Since these criteria are sensitive to outliers, other more robust approaches, such as using the median of the absolute MCFD, have been proposed.

When the motion field has a simple representation, a direct search strategy can be used to minimize the objective function. Either an exhaustive search of all candidate motion vectors or a directed search involving progressive refinement can be carried out,. These are the techniques usually used for block-based motion estimation. For more complex representations, gradient-based techniques may be more appropriate. Finally, for dense mo-

5. Results of mosaic construction. This mosaic is composed of 90 frames from a panning, zooming sequence, which initially zooms out and pans from right to left. The zoom ends after approximately 10 frames. The top image shows the mosaic after 90 frames, and the bottom row shows frames 90, 45, and 1. Observe the scale change between frame 1 and 45 and how they appear in the mosaic.

tion fields, techniques such as stochastic or deterministic relaxation techniques are required [189]. For robust estimation, multiscale motion representations have been found to be essential.

One of the interesting applications of motion estimation is image stabilization, defined as the process by which unwanted motion in a sequence due to camera jitter or a moving sensor (unmanned air vehicle) is removed. Depending on a range of motion models that one can assume (translation, similarity, affine, perspective), different algorithms with different throughput rates have been designed [190-193]. Special platforms such as VFE-100 and Datacube have been used. Once the images are stabilized by properly positioning them, one can generate a panoramic view of the scene. An example of mosaicking is shown in Fig. 5. Commercial products using simple stabilization techniques are already included in camcorders. Mosaicking algorithms have also been commercialized recently.

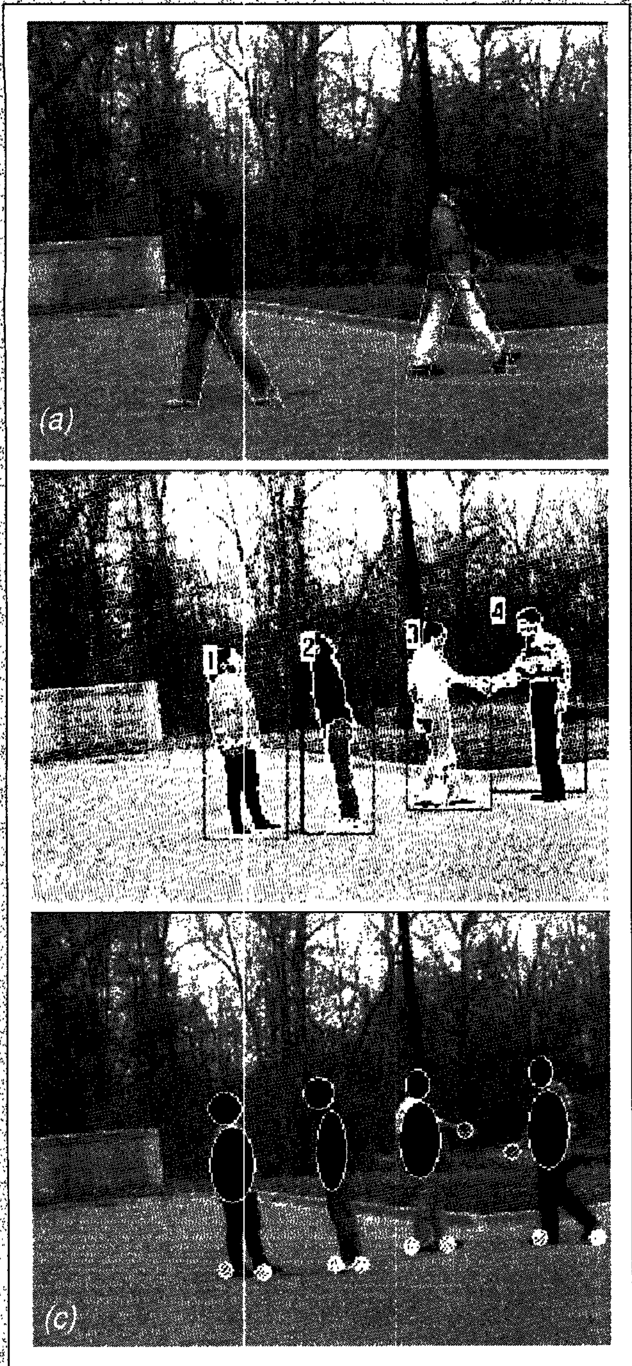
The key problem in motion estimation for the future is finding good representations that can compactly encode motion with the accuracy required for a given application. For example, in coding we need representations that can achieve the best balance between the cost of coding the motion field and the cost of coding the motion-compensated residual. Recent promising work on motion representations can be found in Moulin et al. [188].

Three-Dimensional Motion Estimation

The goal is to estimate the 3-D motion of an object based on image sequence(s) taken of the object. The discussion has been divided into three sections corresponding to motion estimation of rigid, articulated, and nonrigid objects.

Rigid Motion

The most basic scenario is that we have only a single video camera looking at the 3-D moving object. First, we extract from the image sequence the apparent 2-D (in the image plane) movements of features on the object (usually points, but sometimes also lines). Then the 3-D motion parameters (a total of 6 parameters: 3 for translation and 3 for rotation) are derived from the 2-D movements. Two major methods for the first step are using optic flow and tracking discrete features.



▲ 6. Human body tracking: (a) Segmentation of two walking human bodies into heads and legs; (b) Detection of four independently moving humans in an outdoor scene; (c) Display of the heads, bodies, hand, and feet of the humans in (b).

The optic flow method is effective when the motion from one image frame to the next is small. For larger motion, tracking discrete features is more appropriate. Tracking features requires finding point and line correspondences over time.

In most real-world applications, the scene will contain multiple moving objects (e.g., a moving car in a stationary background; the background is an object moving with zero velocity). Thus, it is necessary to segment the scene into the individual objects. Although we are not concerned with the segmentation problem here (motion-detection methods described earlier can be used for this problem), it is worthwhile to point out that many of the motion-estimation techniques can in fact be used to help segmentation.

Our main concern here is the second step of the 3-D motion estimation problem: how to derive the 3-D rigid motion parameters from the apparent 2-D movements of point (and line) features on the object.

When we write equations relating the 3-D motion parameters to the 2-D apparent movements, it becomes evident that the ranges of the feature points appear in the equations as unknown variables. Thus, in estimating the 3-D motion parameters, we are simultaneously estimating the 3-D "structure" of the object, i.e., the 3-D coordinates of the object feature points [194].

It also becomes evident that with a single camera, the translation vector and the feature-point ranges can be estimated only to within a global scale factor. When more than one video camera is used, we can use stereo techniques to find point or line correspondences in 3-D. The 3-D motion estimation problem then becomes much easier.

Earlier work concentrated on the two-view problem: i.e., only two consecutive image frames are processed at a time, and the goal is to estimate the 3-D motion parameters of the rigid object from the first to the second frame. A number of deep theoretical results on the uniqueness and number of solutions have been obtained and several elegant linear algorithms for 3-D motion estimation have been developed [195-199].

Later, it was realized that two-view 3-D motion estimation algorithms are quite sensitive to noise [200]. Attention was turned to the use of long image sequences and to motion modeling. Kalman filtering approaches [201-203] to long-sequence 3-D motion estimation problems turned out to be effective in combating problems due to noise and for estimating the motion parameters. More recently, a factorization approach [204] for long sequence 3-D motion estimation has been developed; first for the orthographic projection case, and later generalized to the perspective transformation case.

Articulated Motion

Since an articulated object consists of joined rigid parts, many of the 3-D rigid motion estimation techniques can be applied [205, 206]. In many problems, the articulated object may contain a large number of rigid parts. Estimating all the motion parameters at once may result in the need to solve a very large set of nonlinear equations. A decomposition approach to deal with this difficulty has been proposed in [207].

Motivated by applications in surveillance and in human-computer interfaces, there has been an increasing amount of research on the analysis of human body motion from video [208]. An example is shown in Fig. 6.

Nonrigid Motion

To paraphrase Tolstoy's first sentence of *Anna Karenina*: all rigid objects move the same way, but each nonrigid object is nonrigid in its own manner. Thus, early attempts to study nonrigid motion in a general framework did not lead to fruitful results. Researchers then realized that nonrigid motion problems have to be application driven, and each type of nonrigid motion needs its own methodology.

For elastic motion, important work has been done for heart-wall motion and for human facial movement. Heart-wall motion analysis is of tremendous interest to the biomedical community. Here, as in all nonrigid object cases, 3-D shape and motion modeling is the key. Superquadrics and implicit polynomials have been used to model the 3-D shape of the heart. The methodology proposed in [209] for tracking the shape and motion of nonrigid articulated objects looks very promising. Modeling human facial movements is a topic of central interest in MPEG-4's synthetic natural hybrid coding subgroup. The 3-D geometry of the face is modeled by a wireframe; facial movements are effected by moving the vertices of the wireframe. A facial animation parameter set is determined to describe the motion efficiently. Each parameter accounts for the movement of a subset of the vertices [210].

Although much work has been done in clothing animation in the computer graphics community, not much research has been done on the analysis side. More recently, algorithms for estimating the mechanical parameters in cloth models by observing the draping of the cloth have been developed in [211]. Interesting work on visualization of turbulent flow may be seen in [212].

Although much of our discussion has been on quantitative recovery of 3-D motion and structure, there are many situations where extraction of qualitative information related to motion and structure may be sufficient and/or effective. We refer the interested reader to [213]. Alternatives to recovery of motion and structure are also of interest in situations where only a restricted set of parameters constitute sufficient information to carry out a desired object or task; for example, in mobile robotics, temporal information can be visually extracted and used to perform trajectory control; time-to-contact (TTC) and time-to-synchronization (TTS) are examples of this type of information. Several interesting results on estimation of TTC and TTS can be found in [214-217].

Given the ubiquity of video data that is becoming available, motion-related research has gained momentum in recent years. Tasks such as scene modeling using video, monitoring of activities of humans and vehicles from groundbased and airborne video, human-computer interfaces (face, gesture, sign language recognition), and medical applications are becoming increasingly viable.

Video Signal Processing

A. Murat Tekalp, University of Rochester

Digital video enjoys many advantages over conventional analog video, including bandwidth compression, robustness against channel noise, interactivity, and ease of manipulation. This section focuses on the latter two aspects of digital video.

Filtering

Video filters can be classified as intraframe (field) spatial, motion-adaptive, and motion-compensated filters [218]. Spatial filters are easiest to implement; however, they do not make use of the high temporal correlation in the video signals. Motion-adaptive filters do not rely on motion estimation; rather, they adapt their response according to a motion-detection function. Finally, motion-compensated filters require highly accurate correspondence estimation between successive views. Some applications of video filtering, which is a relatively mature subject, are discussed below.

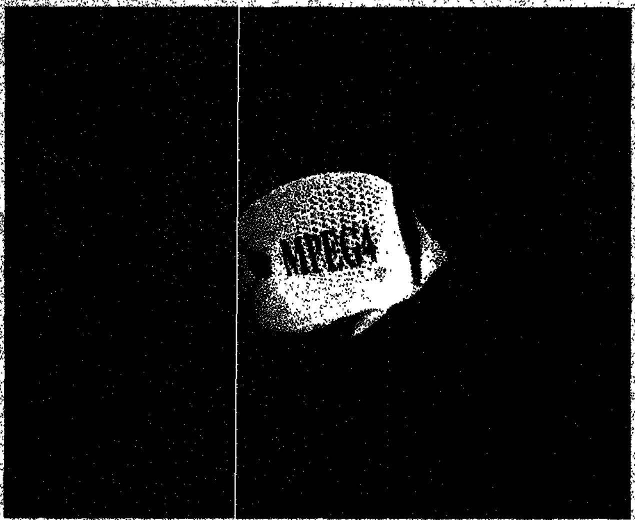
Format Conversion

Digital video signals come in many formats, e.g., broadcast TV signals are digitized with the ITU-R 601 format, which is 30 frames/sec, 720 pixels by 488 lines per frame, 2:1 interlaced, 4:3 aspect ratio, and 4:2:2 chroma sampled. MPEG-1 compression requires an input signal in the standard input format (SIF), which is 360 pixels by 244 lines, progressive, and 4:2:0 chroma sampled; most PC monitors employ the VGA format, which is 60-72 frames/sec, 640 pixels by 480 lines, progressive, and 4:4:4 chroma sampled. With the advent of high-definition digital video, recent standardization efforts between the TV and PC industries have resulted in the approval of 18 different digital video formats by the FCC in the U.S.A. Clearly, seamless exchange of video signals between TV and PC's then requires effective format conversion on the fly. Some commonly used simple intraframe field filters for format conversion, e.g., ITU-R 601 to SIF and vice versa, and 3:2 pull down to display 24 Hz motion pictures in 60 Hz format, have been reviewed in [219]. Other more sophisticated format conversion methods include motion-adaptive field rate doubling and de-interlacing [220], and motion-compensated frame rate conversion [218].

Preprocessing

Video signals suffer from several degradations and artifacts. These include sensor noise and lens aberrations (e.g., barrel distortion and optical blur) in signals digitized from inexpensive commercial camcorders, color bleeding in digitized VHS recorded composite signals (e.g., NTSC), scratches and yellow spots in signals digitized from old motion pictures, blocking artifacts and mosquito noise due to high compression, and so on. Some of these degradations may be acceptable under certain viewing conditions; however, they become objectionable for freeze-frame or printing from video applications. Denoising filters attempt to represent undesired variations in image intensity (noise) by statistical models (e.g., Gaussian, Poisson, and impulsive) and then smooth them accordingly. Some filters are adaptive to scene content in that they aim to preserve spatial and temporal edges while removing the noise. Examples of edge-preserving filters include median, weighted median, adaptive linear mean square error, and adaptive weighted-averaging filtering [218]. Deblurring filters can be classified as those that do require a model of the degradation process (e.g., inverse, constrained least squares, and Wiener filtering) and those that do not (e.g., contrast adjustment by histogram specification, and unsharp masking) [221]. Deblocking filters smooth intensity variations across block borders.

Video contains a high amount of temporal redundancy; i.e., successive frames generally have large overlaps with each other. Assuming that frames are shifted by subpixel amounts with respect to each other (which is almost always the case), it is possible to exploit this redundancy to obtain a high-resolution reference image (mosaic) [222] of the regions covered in multiple views. High resolution reconstruction methods employ least-squares estimation, backprojection, or projection-onto-convex sets methods based on a simple instantaneous camera model [223, 224] or a more sophisticated camera model including motion blur [225].



▲ *7. An example of object-based video editing using 2-D dynamic mesh representation. First, a 2-D dynamic mesh representation of the video object "Bream" (the fish) is computed. Next, the text overlay "MPEG-4" is registered with this mesh model thus enabling animation of the text overlay with the motion of the underlying video object. The animated text overlay can then be alpha blended with the video object for display.*

Spatio-Temporal Segmentation and Object Tracking

One of the challenges in digital video processing is to decompose a video sequence into its elementary parts (i.e., shots and objects). A video sequence is a collection of shots, a shot is a group of frames, and each frame is composed of synthetic or natural visual objects. Thus, temporal segmentation generally refers to finding shot boundaries, spatial segmentation corresponds to extraction of visual objects in each frame, and object tracking means establishing correspondences between the boundaries of objects in successive frames.

Temporal segmentation methods aim to detect shot boundaries including such special edit effects as cuts, dissolves, fades, and wipes. Thresholding and clustering using histogram-based similarity metrics have been found effective for detection of cuts [226-228]. Detection of special effects with high accuracy, on the other hand, requires customized methods in most cases, and is a current research topic. Segmentation of objects by means of chroma-keying is relatively easy and is commonly employed. Automatic methods based on color, texture, and motion similarity, however, often fail to capture semantically meaningful objects [229, 230]. Semiautomatic methods, which aim to help a human operator perform interactive segmentation by tracking boundaries of a manual initial segmentation, are usually required for object-based video editing applications. Object tracking algorithms, which can be classified as boundary, region or model-based tracking methods, can be based on 2-D or 3-D object representations with appropriate parametric motion representations [231]. Effective motion analysis is an essential part of digital video processing and remains an active research topic.

Nonlinear Editing

Editing of analog video required an assembly of tape decks and studio equipment and is generally called linear editing because it can only be performed sequentially on a rolling tape recorder. The advent of digital video removed many of the restrictions of analog editing. Digital video allows random access for editing; hence, nonlinear editing. However, video editing remained confined to studios with expensive production-quality equipment until not long ago. Recent commercialization of digital video and availability of inexpensive PCs with digital video capability brought nonlinear video editing to the desktop. Editing tasks may include object deletion and transformation, virtual object overlay (augmented reality), alpha blending, morphing, and synthetic transfiguration [231]. An example of object-based video editing using a 2-D dynamic mesh representation is shown in Fig. 7.

Content-Based Indexing for Search and Browsing

Storage and archiving of digital video in shared disks and servers in large volumes, browsing of such databases in real time, and retrieval over switched and packet networks pose many new challenges, one of which is efficient and effective description of content. The simplest method to index content is by means of a thesaurus of keywords, which can be assigned manually or semiautomatically to programs, shots, or visual objects [232]. It is also desirable to supplement these keywords with visual features describing appearance (color, texture, and shape) and action (object and camera motion), as well as sound (audio and speech) and textual (script and close-caption) features [227, 233]. Furthermore, it is of interest to browse and search for content using compressed data since almost all video data will likely be stored in compressed format [234].

Video indexing systems may employ a frame-based, scene-based, or object-based video representation [235]. The basic components of a video indexing system are temporal segmentation, analysis of indexing features, and visual summarization. The temporal segmentation step extracts shots, scenes, and/or video objects. The analysis step computes content-based indexing features for the extracted shots, scenes, or objects. Content-based features may be generic or domain-dependent. Commonly used generic indexing features include color histograms, type of camera motion [236], direction and magnitude of dominant object motion, entry and exit instances of objects of interest [237], and shape features for objects. Domain-dependent feature extraction requires a priori knowledge about the video source, such as news programs, particular sitcoms, sportscasts, and particular movies. Content-based browsing can be facilitated by a visual summary of the contents of a program, much like a visual table of contents. Among the proposed visual summarization methods are story boards, visual posters, and mosaic-based summaries. With the upcoming MPEG-7 standardization effort on content-based video description, this subject is sure to remain an active research topic in the near future.

Image and Video Coding

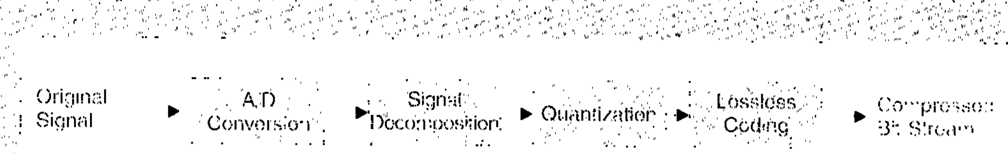
Bernd Girod, University of Erlangen-Nuremberg; Robert Gray, Stanford University; Jelena Kovacevic, Lucent Technologies; and Martin Vetterli, Swiss Federal Institute of Technology and University of California, Berkeley

Due to the vast amount of data associated with images and video, compression is a key technology for their digital transmission and storage. The availability and demand for images and video continue to outpace increases in network capacity. Hence, the importance of compression is not likely to diminish, in spite of the promises of unlimited bandwidth.

Compression takes advantage of the structure of images and video, especially of the statistical redundancy inherent in such signals. It can also exploit the limitations of human visual perception to omit components of the signal that will not be noticed.

A typical compression system has several stages as depicted in Fig. 8. The analog-to-digital converter samples and finely quantizes an image, producing a digital representation. A signal decomposition uses linear transforms or filter banks to break the signal into parallel channels of separate images (bands or groups of coefficients). Such decompositions serve to compact the energy into a few coefficients, to decorrelate separate subbands or groups of coefficients, and to convert the images into a form where perceptual processing is naturally incorporated. Most of the compression occurs in the quantization stage, which operates in the transform domain on individual pixels (scalar quantization) or on groups of pixels (vector quantization). Lossless compression (entropy coding) typically involves run-length coding combined with Huffman or arithmetic codes to further save bits in an invertible fashion. Occasionally specific components will be combined or omitted, but most current image-coding standards possess all of the components in some form.

The enormous commercial potential of image and video coding has stimulated rapid growth of the research efforts to find improved or entirely new techniques. In this short survey we do not attempt a comprehensive overview; rather, we selectively summarize some of the most common themes and principles of this exciting field. We start by reviewing source coding principles, which are not specific to image coding, but are fundamentally important nevertheless. We then discuss subband and transform coding, move on to predictive coding, motion compensation and rate distortion methods in compression systems, and close with a discussion of image communication systems issues.



8. A typical compression system. The analog-to-digital converter samples and finely quantizes an image, producing a digital representation. A signal decomposition uses linear transforms or filter banks to break this digital representation into parallel channels of separate images. Most of the compression occurs in the quantization stage. Lossless compression (entropy coding) typically involves run-length coding combined with Huffman or arithmetic codes to further save bits in an invertible fashion.

Source Coding Principles

In an ideal image coder (or any source coder) one would like high fidelity of the reconstructed image at the receiver in combination with low bit-rate and low complexity of the coder and the decoder. Alas, these requirements are at odds with each other, and good design is a question of finding optimal trade-offs. Source-coding theory focuses on the basic trade-off between average rate, the bit-rate in terms of bits per pixel, and average distortion, measured commonly by mean-squared error. Although often maligned, weighted versions of squared error (especially when applied to transformed signals and allowed to depend on the input) have proved quite useful as indications of perceived quality. Complexity considerations may enter through structural constraints on the specific types of codes considered.

The optimal trade-off between rate and distortion can be precisely defined in several equivalent ways: by minimizing the (average) distortion, D, subject to a constraint on the (average) rate, R, by minimizing the rate subject to a constraint on the distortion, or by an unconstrained minimization of the Lagrangian cost function, $D + \lambda R$, where the Lagrange multiplier, λ, provides a weighting of the relative importance of distortion and bit-rate.

The theory of data compression has two principal branches, both of which are celebrating their 50th birthday: Shannon's rate-distortion theory, a branch of information theory sketched in his 1948 paper [238] and developed in remarkably complete form in his 1959 paper [239], and high-rate or high-resolution quantization theory, an approach involving approximations for low distortion and high bit-rate that began with the classical work on PCM (pulse coded modulation) of Oliver, Pierce, and Shannon [240] and the work on quantization error spectra by Benett [241]. Rate-distortion theory [242-245] provides unbeatable lower bounds to the obtainable average distortion for a fixed average rate, or vice versa. It also promises that codes exist that approach these bounds when the code dimension and delay become large. High-rate quantization theory [246, 247] provides approximations for distortion and rate that can be optimized for fixed and finite dimension and delay. These results imply many useful comparisons of the gains achievable by transform coding and other structured codes in comparison with the Shannon optima.

Unfortunately, the theory does not provide us with an explicit method for constructing a practical optimum coder and decoder. It can nevertheless give very important hints about the properties of an optimum coder/decoder. For example, for wide-sense stationary Gaussian sources with memory and mean-squared error distortion, the mathematical form of the rate-distortion function suggests that an optimum coder splits the original signal into spectral components of infinitesimal bandwidth and encodes these spectral components in an independent manner [248]. This is consistent with the high-rate quantization theory, which demonstrates that the decorrelating KLT is optimal for transform coding of Gaussian sources. The corresponding bit allocation for the separate subband quantizers should make the rate proportional to the logarithm of energy in each subband. For high bit-rate, uniform scalar quantization coupled with high-order or adaptive entropy coding can at best achieve a performance 0.255 bits or 1.5 dB away from the Shannon bound. The gap can be closed further if vector quantization or trellis-coded quantization is used [249, 250]. These theoretical insights have motivated the widespread use of transform and subband coders, even when the rates are not high and the images are certainly not Gaussian.

Transform and Subband Coding

Transform coding and subband coding (SBC) refer to compression systems where the signal decomposition (Fig. 8) is implemented using an analysis filter bank. At the receiver, the signal is reassembled by a synthesis filter bank. By transform coding, we usually mean that the linear transform is block-based (such as a block-wise DCT in JPEG). When transform coding is interpreted as an SBC technique, the impulse responses of the analysis and synthesis filters are at most as long as the subsampling factor employed in the subbands; thus, the image can be subdivided into blocks that are processed in an independent manner. General SBC, on the other hand, allows the impulse responses to overlap and thus includes transform coding as a special case.

As pointed out earlier in this article, one of the most important tasks of the transform is to pack the energy of the signal into as few transform coefficients as possible. The DCT yields nearly optimal energy concentration for images, while being a lot easier to implement than the KLT, which is the theoretically best orthonormal energy-packing transform. As a result, almost all image transform coders today employ the block-wise DCT, usually with a block size of 8 x 8 pixels. The transform is followed by quantization (most often scalar uniform quantization) and entropy coding. Typically, the transform coefficients are run-level encoded; that is, successive zeros along a zig-zag path are grouped with the first nonzero amplitude into a joint symbol which is then Huffman coded. For example, the widely used JPEG standard works in this fashion, and so do the prediction error encoders for MPEG, H.261, and H.263. The LOT could be substituted for the DCT in the above process to avoid some of the typical blocking artifacts that become visible with coarse quantization. Figures 9(b)-(c) show the DCT decomposition of the *Barbara* image in Fig. 9(a) as well as the JPEG coding result at 0.5 bits/pixel.

The full potential of SBC is unleashed when nonuniform bandsplitting is used to build multiresolution representations of an image. Beside excellent compression, multiresolution coders provide the successive approximation feature;

Visual Coding Standards: A New World of Visual Communication

Gary J. Sullivan, PictureTel Corporation

In the past few years, the advent of international standards for digital image and video coding has given the world a vast new array of capabilities for visual communication. Widespread communication is impossible without the use of a common language, and a visual-coding "standard" is a specialized language that defines how to interpret the digital representations of pictures.

The compressed coding of pictures into digital form was primarily the domain of cloistered research laboratories just a few years ago. Today, millions of people worldwide watch television pictures that have been digitally transmitted; view and create coded pictures and video on the Web, view coded video on their personal computers, and even use digital videotelephony for interactive remote conversations. The key to making this transition from the research lab to the casual user has been the creation of standards for visual communication.

These standardization activities have provided great opportunities for researchers to directly influence the world of the future. The standardization groups meet to closely examine the capabilities, performance, and practicality of the various concepts, features, and designs that are brought forth from the research community. They then collaborate to merge the best of these ideas into a coherent and fully defined specification that all can use. Repeatedly, the final written standard has become a better overall technical solution than any of the individual proposals that were brought into the collaborative process.

The standards have themselves become touchstones for new creative research, since they provide a well-known reference for comparison. The creation and promulgation of a standard organizes the collective thoughts of the technical community, creating a breadth of understanding and experience that could not have been achieved by research alone.

The biggest names in the realm of standardization of visual information coding are the ITU-T and the ISO/IEC JTC1 organizations. The ITU-T (formerly called the CCITT) approved the first digital video-coding standard (Rec. H.120) in 1984 and has been updating its methods periodically since then by revising H.120 in 1988, then moving on to increasingly successful standards in 1990 (Rec. H.261) and 1995 (Rec. H.263), and enhancing its latest standard this year (Rec. H.263+). In 1993, the ISO/IEC JTC1 completed the MPEG-1 video-coding standard (IS 11172-2) and joined with the ITU-T to develop the JPEG standard for still pictures in 1994 (IS 10918-1) and the MPEG-2 standard for video in 1996 (IS 13818-2). Each of these standards has led to increasing growth in the use and variety of applications for digital coded visual information, and both groups are working on new efforts for the future (such as the MPEG-4 project in ISO/IEC JTC1 and the H.263++ and H.26L projects in the ITU-T).

Today's standards development process has become very responsive to the progress of research, and the research world has been helped by the progress of standardization. This symbiotic relationship will continue into the future, providing a fertile field for new technology development.

as higher-frequency components are added, higher-resolution, better-quality images are obtained. Moreover, multiresolution techniques fit naturally into joint source-channel coding schemes. Figures 9(d)-(e) show the uniform subband decomposition of the *Barbara* image as well as the SBC coding result at 0.5 bits/pixel, while Figures 9(f)-(g) show the octave-band subband decomposition and the coding result at 0.5 bits/pixel. Subband coders with octave-band decomposition such as illustrated in Figures 9(f)-(g) are also often referred to as discrete wavelet transform (DWT) coders, or wavelet coders.

The multiresolution image representation in Figures 9(f)-(g) is a critically sampled subband pyramid. However, overcomplete representations, first introduced as the Laplacian pyramid by Burt and Adelson [251], are also very powerful. An input image is fed into a lowpass filter followed by a downsampler to produce a coarse approximation that is then used to interpolate the original (by upsampling and filtering) and calculate the difference as the interpolation error. This process can be recursively applied to the coarse version. Thus, instead of compressing the original image one compresses the coarse version and the interpolation errors at various resolutions. The interpolation can be based on lower-resolution images with or without quantization error (referred to as open-loop and closed-loop pyramid coders). The overcomplete pyramid provides energy concentration and possesses the successive approximation property, since one can start with the coarsest version and then add detail (interpolation errors) to reconstruct higher-resolution versions. Moreover, the pyramid coding scheme allows for nonlinear operations for producing the coarse version and the details. Its only disadvantage is that it produces a redundant representation.

Today, many state-of-the-art multiresolution image coders draw on the ideas introduced by Shapiro in his embedded zero-tree wavelet algorithm (EZW) [252]. The algorithm employs a data structure called zero-tree, where one assumes that if a coefficient at a low frequency is zero, it is highly likely that all the coefficients at the same spatial location at all higher frequencies will also be zero; thus, when encountering a zero-tree root, one can discard the whole tree of coefficients in higher-frequency bands. Moreover, the algorithm uses successive approximation quantization, which allows termination of encoding or decoding at any point. These initial ideas have produced a new class of algorithms aimed at exploiting both frequency and spatial phenomena [253].

While research has shown that wavelet coders can produce superior results, transform coders employing a

Visual Coding Standards: A Research Community's Midlife Crisis?

Michael Orchard, Princeton University

Yes, standards have turned JPEG and MPEG into household terms and brought digital images and video into millions of homes worldwide. But what have they done for us visual compression researchers, the community responsible for developing the algorithms? Now that our best ideas have been perfected, packaged, and polished for public dissemination, what remains for researchers in this field to do? Should research continue on algorithms whose chances of ever becoming a standard might be questionable?

The visual-compression research community has wrestled with these kinds of questions over the past five years, and the answers that have been offered have reshaped the field. Widespread acceptance of visual coding standards have forced us to reassess directions and priorities, but overall, it is clear that standards open more doors than they close, and standards cannot alter the nature or diminish the importance of truly fundamental research advances in the field.

By accelerating the development of visual applications, standards have helped uncover challenging new problems offering exciting opportunities for the research community. Robust transmission of images and video over packet networks and video transmission over wireless channels have become hot research topics. Digital video libraries, content-based retrieval, and digital watermarking are examples of active new research areas spawned by the widespread application of coding standards and involving problems of visual representation that are closely related to the coding problem.

Balanced against their positive effects, standards have also had the unfortunate effect of diverting attention from important fundamental questions in image and video compression. The success of standards has suggested that they are based on sound technical approaches to the coding problem and has focused the community's attention on the refinement of those approaches for improved performance. In fact, today's standards are built on ad-hoc frameworks that reflect our very limited understanding of the fundamental structure of image and video sources. There is very little reason to believe either that today's standards come close to the ideal performance possible for these sources (that is, it is unlikely that they are near the fundamental entropy of these sources), or that there cannot exist simple, practical coding algorithms performing much better than today's standards. In particular, the standard hybrid framework for motion-compensated video coding is based on a naive understanding of the relationship between motion and intensity uncertainty models for video sequences.

The gaps in our understanding are wide, and progress in bridging those gaps requires continued strong research efforts by the community. Unfortunately, the fundamental advances that are needed are not likely to produce immediate practical algorithms to challenge today's standards, and this has discouraged research in these directions. It is particularly important that young researchers entering the field be encouraged to apply their creativity and healthy skepticism toward challenging accepted frameworks, engaging basic issues, and proposing sound alternative approaches, no matter how far-fetched they may appear. In the long term, these efforts promise progress on important fundamental questions, a more vibrant research community, AND superior standards.

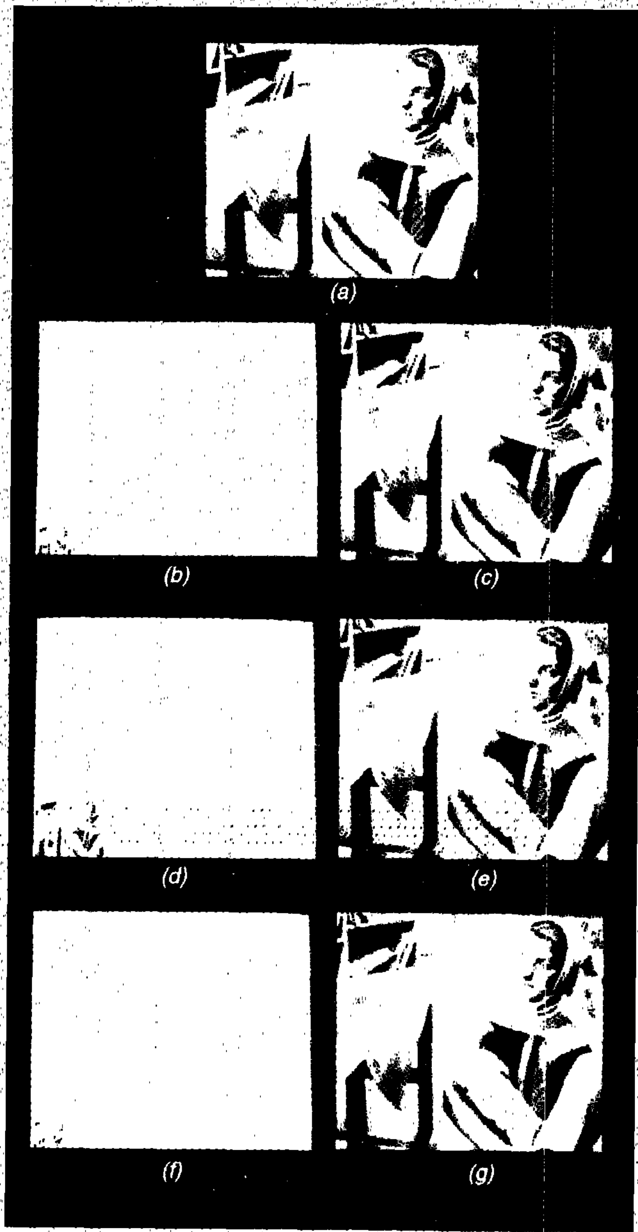
block-wise DCT are still dominant today. After years of use, DCT coders are very well understood and many improvements have been made, for example in the area of fast algorithms or by imposing perceptual criteria. The next still-image coding standard, JPEG 2000, as well as the next in the line of MPEG standards, MPEG-4, might very well include wavelet coding, in addition to or in place of the DCT.

Predictive Coding

Except when used with subband or transform coding, predictive coders do not decompose the image into independent components. Instead, both the coder and the decoder calculate a prediction value for the current signal sample. Then, the prediction error, rather than the signal sample itself, is transmitted. This principle can be used for both lossy and lossless image coding. Most commonly, the predictors calculate linear combinations of previous image samples, since general nonlinear predictors, addressed by combinations of, say, 8-bit pixels, would often require enormous look-up tables for the same performance.

For lossy predictive coding, differential code pulse modulation (DPCM), invented by Cutler in 1952, has been used since the early days of image coding. Intraframe DPCM exploits dependencies within a frame of a video sequence. Typically, pixels are encoded in line-scan order and the previous sample in the current line and samples from the previous line are combined for prediction. Today, this simple scheme has been displaced by vastly superior transform SBC schemes, without doubt a result of the unnatural causal half-plane constraint for the region of support of the predictor. In fact, lossy predictive intraframe coding is alive and well in the form of predictive closed-loop pyramid coders that feed back the quantization error before encoding the next higher-resolution layer (see the section on Image Transforms). It has been shown recently that closed-loop pyramid coders even outperform the equivalent open-loop overcomplete pyramid representations when combined with scalar quantizers [254].

For interframe coding where statistical dependencies between successive frames of a video sequence are exploited, DPCM is the dominating scheme today and for the foreseeable future. Other than, for example, spatio-temporal SBC, interframe DPCM avoids the undesirable delay due to buffering of one or several frames. Moreover,



9. Subband image coding results: (a) Barbara *image of size 512 x 512 pixels with 8 bits/pixel; (b) 8 x 8 DCT transform of* Barbara *used in JPEG; (c) JPEG coded* Barbara *with 0.5 bits/pixel and 28.26 dB SNR; (d) Uniform subband decomposition of* Barbara*; (e) SBC coded* Barbara *using uniform subband decomposition at 0.5 bits/pixel with 30.38 dB SNR; (f) Octave-band subband/wavelet decomposition of* Barbara*; (g) SBC coded* Barbara *using octave band subband/wavelet decomposition at 0.5 bits/pixel with 29.21 dB SNR*

it is straightforward to incorporate motion adaptation and motion compensation into a temporal prediction loop and combine motion-compensated prediction with other schemes for encoding of the prediction error.

Motion-Compensated Video Coding

All modern video compression coders such as those standardized in the ITU-T Rec. H.261 [255] and H.263 [256], or in the ISO MPEG standards [257], are motion-compensated hybrid coders. Motion-compensated hybrid coders estimate the displacement from frame to frame and transmit the motion vector field as side information in addition to the motion-compensated prediction error image. The prediction error image is encoded with an intraframe source encoder that exploits statistical dependencies between adjacent samples. This intraframe encoder is an 8 x 8 DCT coder in all current video coding standards [255-257], but other schemes, such as subband coders or vector quantizers, can be used as well.

Motion-compensated hybrid coding can theoretically outperform an optimum intraframe coder by at most 0.8 bits/pixel in moving areas of an image, if motion compensation is performed with only integer-pixel accuracy [258]. For half-pixel accuracy, this gain can be up to 1.3 bits/pixel. In addition, in nonmoving areas (or other parts of the image that can be predicted perfectly) no prediction error signal has to be transmitted and these areas can simply be repeated from a frame store, a technique often referred to as conditional replenishment.

Motion compensation works well for low spatial frequency components in the video signal; for high spatial frequency components even a small inaccuracy of the motion compensation will render the prediction ineffective. Hence, it is important to spatially low-pass filter the prediction signal by a loop filter. This loop filter is explicitly needed for integer-pixel accurate motion compensation. For subpixel accurate motion compensation, it can be incorporated into the interpolation kernel required to calculate signal samples between the original sampling positions. The loop filter also improves prediction by acting as a noise reduction filter. Prediction can be further improved by combining multiple independently motion-compensated prediction signals. Examples are the bidirectionally predicted B-frames in MPEG [257] or overlapped block motion compensation [259] that has also been incorporated in the ITU-T Rec. H.263 [256].

Especially at low bit rates, motion compensation is severely constrained by the limited bit rate available to transmit the motion vector field as side information. Rate-constrained estimation [260] and a rate-efficient representation of the motion vector field are therefore very important. For simplicity, most practical video coding schemes today still employ blockwise constant motion compensation. More advanced schemes interpolate between motion vectors, employ arbitrarily shaped regions, or use triangular meshes for representing a smooth motion vector field. Ultimately, we might expect 3-D models to be incorporated into motion compensation one day, we hope with such success that transmission of the prediction error is no longer required. This is a goal of ongoing research into model-based video coding, although the success of such schemes for general types of video material is still uncertain.

Rate-Distortion Methods in Compression Systems

As outlined earlier, the formal framework for compression methods consists of rate distortion theory and high-rate quantization theory. How can this theory be applied to practical compression schemes? Recent work has made progress in this direction by bridging at least in part the gap between theory and practice in source coding. The idea is to use standard optimization procedures such as Lagrangian methods to find local optimal operating points in a rate-distortion sense, under some assumptions about the source. Such techniques were first introduced in the context of PCM by Lloyd in the 1950s, and were later used for vector quantization designs (for example, [261]) and other problems involving transforms and quantization. As an example, consider the problem of finding best orthonormal bases for compression from a large collection of possible transforms. This can be posed as a Lagrangian optimization problem: each set of transform coefficient generates an operational rate-distortion curve, and optimal allocation of bit-rate between transform coefficients is standard. Among all possible transforms, Lagrange optimization allows one to choose the winning transform, and this can be done in an efficient tree-pruning manner if the transforms have some structure [262].

Similar ideas can be used for many of the other problems appearing in practical compression schemes. As examples, we can cite rate control for video coders using dynamic programming [263], allocation of rate between competing units (for example, motion and residual), and optimization of quantization schemes. The important point is that, under certain assumptions such as independence, an optimal or locally optimal solution is sought, as opposed to the somewhat ad-hoc methods that are often used in practical compression schemes.

Image Communication System Issues

Image and video compression is usually not done in isolation, but integrated in to a larger system, typically a communication system. This poses some interesting challenges to the designer of the compression system.

In his groundbreaking 1948 paper [238] that laid the foundations of information theory, Shannon showed that for point-to-point communication over a well-defined channel, a separate optimization of source coding and channel coding can lead to a performance arbitrarily close to the information-theoretic bound for the entire system. Therefore, traditionally, the coding problem has been split into two independent subproblems: source compression and channel coding [264]. This has resulted in largely independent research in the two areas. However, many practical situations do not satisfy the assumption required for the separation principle to hold. For example, most communication is done with a finite-delay constraint, which leads to finite block sizes. Under such delay constraints, error probabilities can be non-negligible, and thus the separation principle has to be revisited.

In such cases, practical schemes using unequal error protection of different parts of the coded stream are used. For example, in video coding for noisy channels, the motion vectors are highly protected, since their loss would be catastrophic. Motion residuals (errors after motion compensation) are either not protected, or much less protected than the motion vectors. In multiresolution source coders, it is natural to protect the coarse resolutions (which are absolutely necessary) more than the fine details; this is another instance of unequal error protection. Note that multiresolution video coding requires multiscale motion compensation [265]. An interesting application of multiresolution coding is the progressive transmission of images, in particular in browsing applications. Instead of coding at a fixed rate, one has to accommodate many rates, depending on the resolution desired by the particular user accessing the image.

Given a source and a channel, how do we allocate resources to source and channel coding (still under the finite-delay constraint)? The answer turns out to be more complicated than expected and is only partly known [266]. The question becomes even more intricate when protocol issues are included (as in channels with feedback [267]).

All methods that allow some interaction of the source and the channel coders go under the name of a joint source/channel coding system. They do not fit the classic separation principle; rather, they solve the practical problem of robust communication when errors do occur. An instance where the separation principle cannot be used is in the case of multiple channels, as in broadcast or multicast scenarios. Then, users with very different channels have to be accommodated, which leads to schemes that adapt to the particular channels they face. For example, embedded modulation together with multiresolution source coding leads to a robust scheme with graceful degradation when the channel degrades [268]. Similar ideas can be adapted to multicast over packet networks [269]. Finally, recent work on multiple description coding addresses the question of transmitting several source descriptions over multiple channels. For example, two descriptions of a source are sent to three receivers, where the first two receive either description, while the third receives both. This interesting theoretical question is relevant to transmission over lossy or delay-constrained packet networks, where random drops may occur. Recently, Vaishampayan derived quantization schemes for this problem [270].

Finally, let us stress the importance of protocol issues. If an error occurs, it can have catastrophic consequences (for example, loss of synchronization in a variable-length lossless code). Therefore, there exists a need for a powerful mechanism to recover from errors, using feedback channels (in point-to-point communications) or resynchronization points (in multicast situations). In a practical image/video communication system, one can do

Exploiting Self-Similarities: Fractal Image Compression

Geoff Davis, Dartmouth College

The prototypical fractal coder, first proposed by Barnsley and Jacquin [271], encodes images via a process resembling vector quantization. The twist is that fractal coders employ a vector codebook derived from the image being coded rather than using a prespecified codebook. The result is a coded image representation very different from that for transform coders. the stored data defines a contraction map of which the original image is an approximate fixed point. Images are decoded by iterating this stored map to convergence.

Transform coders take advantage of spatial redundancy in images. Fractal coders take a somewhat different approach. they exploit redundancy in *scale*. Common image features such as edges and linear gradients are self-similar, in the sense that they are invariant under contractions up to an affine transform. This self-similarity of key image features motivates the particular codebook used by fractal coders—a set of affine transforms of contracted image blocks. One intriguing property of fractal coders is that this self-similarity property can be used to synthesize finer detail, particularly at edges, than was present in the original image.

A major drawback of fractal coding is the high complexity of the encoding process, and considerable effort has been devoted to finding efficient encoding algorithms. Additional important research areas have included bounding reconstruction errors and determining conditions under which the iterative decoding algorithm converges.

Although the mechanics of fractal coding are quite different from transform coders, fractal coders have recently been shown to be closely related to wavelet coders [272]. The link is a natural one, since wavelet bases possess a dyadic self-similar structure similar to that found in fractal coders. This wavelet-fractal synergy provides important insights into the workings of fractal coders. The new understanding has improved the performance of fractal coders considerably, and it has revealed some basic limitations of current coders that will require further research to overcome.

Model-Based Video Coding

Don Pearson, University of Essex, UK

Mention *model-based coding* (MBC) to someone and they tend to think of animated texture mapped wire-frame heads, some of them looking distinctly zombie-like! This is indeed the way MBC began back in the early 1980s; it is also where many young people start their research today. But it is not necessarily the way it will end; coding methods invented for one purpose sometimes find their application elsewhere. An example is run-length coding, which was first investigated for gray scale picture compression before finding its home in facsimile.

What will determine MBC's ultimate fate is its coding efficiency. This depends on the picture material, as it does with all image coding methods: no method works well for all types of objects and all types of movement. We may have many different options within a coder, each suited to a particular type of visual material, and that we (or rather the coder) will choose or switch between. Experiments alone will show where MBC fits in. Those conducted so far tend to indicate that the method works best for large, relatively rigid moving areas in translational or rotational movement. This is not surprising when we think about it, since shape information has to be added to that for motion and texture. The additional overhead must save texture bits to be worth sending.

With increasing levels of sophistication in facial analysis and modeling, it is quite likely that the traditional approach to MBC will eventually yield highly efficient and believable talking heads in low bit-rate applications. But it is also possible that MBC will be found to be useful on a selective basis for coding large nonfacial moving objects in higher resolution, higher bit rate video. MBC is theoretically efficient for such objects, and they are the very objects that cause difficulties in the current generation of MPEG-2 coders [273, 274].

better by jointly designing source coder, channel coder, and transmission protocols. Research that addresses these issues is still in its infancy.

Better Compression Forever?

Students of image coding often ask how many bits at least are required to represent an image or a motion video with reasonable quality. They ask this question not only out of scientific curiosity, but they also want to find out whether research in the field has a future, or whether all the interesting problems have already been solved.

As illustrated by the success of image and video coding standards (see the "Visual Coding Standards: A New World of Communication" and "Visual Coding Standards: A Research Community's Midlife Crisis" side-bars), image and video coding is a mature discipline today. It rests solidly on the foundations of source coding theory. Often, practical schemes perform close to their information-theoretic bounds. Note, however, that most of these bounds are calculated on the basis of crude models about the structure of images. As image models become more refined, compression ratios can improve further. Moreover, many interesting open problems are yet to be solved on how to gracefully integrate image and video codecs into communication systems, where previously neglected requirements, such as robustness, delay, or random access, have to be taken into account. We believe that image and video coding will remain a quick-paced, exciting field well into the next millennium.

Image-Processing Software and Hardware

Ed J. Delp, Purdue University

The hardware and software tools available to acquire and process digital images have changed a great deal in 50 years. In the early 1960s it would have required more than $500,000 to set up a small laboratory that could digitize and process images. An accurate film scanner alone would have cost $100,000 [275]. Many laboratories developed their own software tools written in Fortran and later in C because commercial software packages were not available. The times have certainly changed! Any individual can now, for less than $5000, acquire a system that can easily surpass the performance of any 1960's image-processing research laboratory. Just as in the past when typical users had a digital audio card in their PCs, we are now seeing PC systems that have the capability to acquire and process digital images as part of their display systems, through the use of "media" co-processors (e.g. Chromatic's media processor [276]) and/or built into the instruction set of the microprocessor (e.g. Intel's MMX instructions [277, 278]). What is responsible for this? The absolute explosion in computer hardware performance. One must also consider the availability of inexpensive and large-capacity storage devices. Another driving force in hardware and software is the fact that digital images (and digital video) are now part of many commercial and consumer applications.

Hardware

Digital image processing was born in the late 1950s and was driven by applications in the military and space exploration areas [279-281]. Early efforts at using general-purpose computers to process digital images were not satisfactory for many applications. Computers were not fast enough. For applications requiring more processing speed, special-purpose computer architectures were developed. These included application-specific integrated circuits (ASICs) that could perform specific image-processing operations very quickly, such as FFTs [282], parallel processing systems [283-285], and digital signal processors (DSPs) [286, 287].

While these special systems were very effective, they were very difficult to program and were not widely adopted in research laboratories. One advantage of using a general-purpose computer is the availability of a programming environment. As more general-purpose computers were used for scientific applications, programming tools flourished. In many cases it was far easier to trade off processing speed for programmability. This led to the problem that continues today: special-purpose processors vs. faster general-purpose processors. In applications that require embedded systems (e.g., a digital camera), special-purpose processors are still widely used. This situation may change as microprocessors increase in speed and have instruction sets designed for image/signal-processing functions. In many cases the issues are blurred.

Image processing has not benefited as much as speech and audio processing from the development of DSPs. This has been due to the relatively slow speed of DSPs. In the early 1980s Texas Instruments introduced the TMS320 digital signal processor for applications in speech and audio [287]. Others including AT&T/Lucent, Motorola, and Analog Devices have also developed DSP architectures. Many new DSPs have been introduced that will have an impact on image- and video-processing hardware in the future. These include TI's TMS320C67 [288] and Analog Devices' SHARC processor [289]. These new DSPs have the capability of doing real-time digital video processing. Other developments include Chromatic's Mpact[276] media processor, which is capable of accelerating several multimedia functions simultaneously.

The new microprocessors that are being developed have the capability of real-time image-processing operations. As clock speeds exceed 300 MHz, many real-time image-processing operations can now be performed on these processors without hardware acceleration. The term "native signal processing" (NSP) is sometimes used to describe this situation [287, 290]. The most visible use of NSP is Intel's MMX processors [277, 278]. In MMX, 57 special instructions are added to the Intel processors that allow speed-up in the execution of DSP operations. Sun Microsystems is also pursuing a similar strategy with its VIS instruction set [291]. VIS has been used to implement a real-time MPEG2 decoder on a Sun workstation.

Software

The development of image-processing software has trailed the development of faster processors [292]. Up until quite recently, few tools have been widely available for image-processing software development [293]. There has been an explosion in image-processing software systems in the last five years [294]. These range from very sophisticated software systems, such as Khoros [295, 296], to simple, inexpensive, yet powerful consumer packages such as Adobe's PhotoDeluxe. It is also interesting to note the that one of the goals of the new MPEG-4 standard is software-based codecs [297].

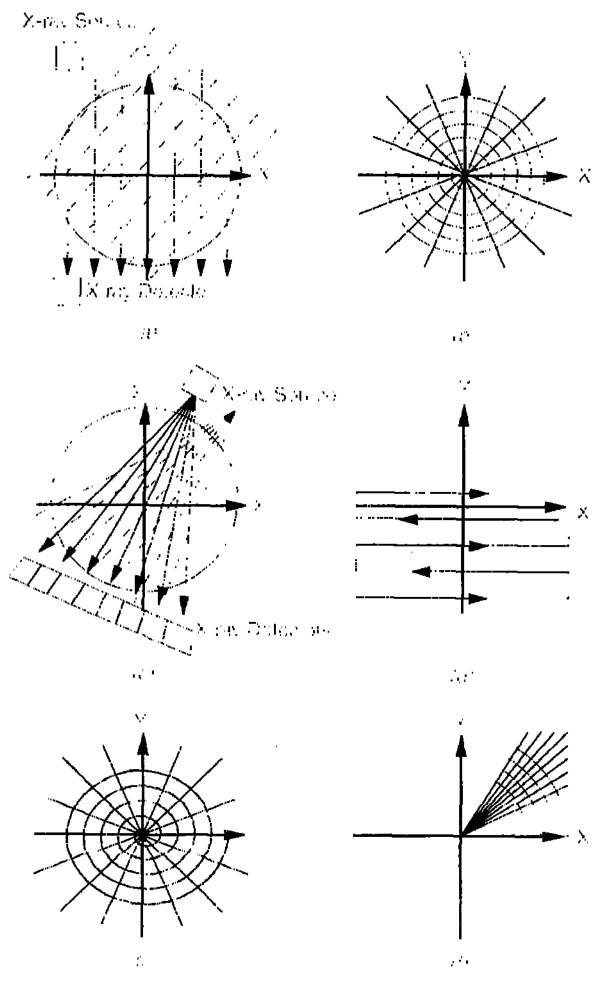
Many of these software packages are designed for a specific imaging applications. For example, Khoros [295, 296] has been developed for image-processing research and has very sophisticated tools and interfaces for the researcher. Khoros has been used for integrating software developed by several distributed research groups. An example is the DARPA sponsored program known as MSTAR. Photoshop [298] was developed for use by graphic artists and designers. Other software tools that have appeared in recent years include MATLAB's Image Processing Toolbox [299], Mathematica [300], LabView [301], the Image Understanding Environment (IUE) [302], and NIH Image [303].

The image- and video-processing field will continue to benefit from the trends in computer technology. We will see faster DSPs and microprocessors that will be able to do real-time image processing. New software tools will be developed that will allow one to use the new multiprocessor PCs that will be arriving soon on everyone's desk. Software environments that have a graphical user interface, such as Khoros and MATLAB, will be the way we all do image processing in the future. The future is indeed bright!

Computed Imaging

Richard Leahy, University of Southern California, and David C. Munson Jr., University of Illinois, Urbana-Champaign

Advances in the design and fabrication of imaging detectors, combined with increasingly fast computers, has fueled the development of a spectacular variety of computed imaging systems. Such systems use numerical algorithms to transform measured data into an image. Imaging data can be collected using a wide range of the electromagnetic spectrum. Quasistatic through radio frequencies, microwave, infrared, optical, x-rays, and gamma rays are now routinely employed in a wide range of applications from microscopy [304] to medical imaging of the whole body [305], and from high-resolution radar imaging [306] to radioastronomy [307, 308]. Acoustic phenomena, again over a wide range of frequencies, are also used as the basis for microscopic, medical, underwater, and geophysical imaging [309-312]. In this very brief review we will restrict attention to imaging systems that collect measurements of line integrals of the unknown image, or samples of its Fourier transform, and applications in medical imaging and remote sensing.



10. Geometries for data collection in computed imaging: (a) Collection of parallel beam projections in CT; (b) Polar Fourier domain grid; (c) Collection of fan beam projections in CT; (d) Rectangular Fourier scan in MR imaging; (e) Spiral Fourier scan in MR imaging; (f) Partial polar Fourier-domain grid in SAR.

In its simplest form, a tomographic imaging system collects measurements of 1-D line integrals or projections along parallel paths through a 2-D object as illustrated in Figure 10(a). By collecting these projections at different angles relative to a fixed coordinate system, we build up a 2-D function from which the image is reconstructed. This function is referred to as the Radon transform of the image after the mathematician Johann Radon, who was the first to investigate the properties of the integrals of n-D functions over (n-1)-D hyperplanes [313]. It is also often referred to as a sinogram since an image consisting of a single point produces a sinusoidal image in Radon transform space.

Consider, for example, the first clinical x-ray CT system designed by Godfrey Hounsfield [314, 315]. A collimated x-ray source and detector are translated on either side of a patient. The detected x-ray measurements provide a projection through the patient of the 2-D distribution of x-ray attenuation coefficients. By rotating the source and detector around the patient other projections are measured (see Fig. 10(a)). Reconstruction of the spatial distribution of attenuation coefficients provides information of diagnostic significance since bone and different soft tissues have distinct attenuation properties.

The projection slice theorem (PST) [2, 315] is the fundamental result that is used to reconstruct images from their projections. This theorem shows that the Fourier transform of a 1-D projection at a particular angle is equal to the 2-D Fourier transform of the image evaluated along the line passing through the origin at the same angle. Using this result it is straightforward to compute a reconstruction of the image by first Fourier transforming each projection to produce samples of the 2-D image Fourier transform, lying on the polar grid in Figure 10(b). These polar data are then interpolated onto a regular Cartesian sampling grid and the image is computed using the inverse 2-D discrete Fourier transform [316, 317]. This form of image reconstruction is called a direct Fourier method.

An alternative approach is based on analytic manipulations using the PST, showing that the image can be recovered directly from the data without using Fourier transforms. Each projection is filtered with a ramp filter and then backprojected along the path over which the line integration was originally performed. Carrying out this procedure for each projection and then summing the results produces the reconstructed image.

[318, 319]. This method, typically referred to as filtered backprojection (FBP), is the basis for reconstruction in almost all commercially available computed tomography systems. Interestingly, it is also an almost direct numerical implementation of the inverse formula first derived in 1917 by Radon. In recent years it has been discovered that the FPB method is a fast means of implementing the direct Fourier method for the case where the polar-to-Cartesian interpolator uses a 2-D sinc kernel [320, 321].

X-ray projection data can be collected far more quickly using the fan-beam employed in the current generation of CT scanners (see Figure 10(c)). It is possible to re-sort this fan beam data into equivalent parallel projections before reconstruction. Fortuitously, however, a simple change of variable in the Radon inversion formula allows reconstruction by directly weighting, filtering, and backprojecting the fan beam data [319, 322, 323].

A limitation to the FBP approach is the assumption that the data are exact line integrals of the image. An entirely different approach to reconstruction from projections is to model the relationship between the sampled data and the unknown pixelated image as a large set of simultaneous linear equations where the matrix elements represent the fractional contribution of each pixel to each projection measurement. The huge dimensionality of these systems, coupled with the special sparse matrix structure and the presence of noise in the data, calls for specially tailored reconstruction algorithms. One such approach is the iterative row-action method or algebraic reconstruction technique (ART) [323, 324]. This method updates all pixels at each iteration using a weighted backprojection of the error in a single projection. By performing this procedure in a cyclic fashion for each projection in turn, a solution to the set of equations is obtained. As with the analytic approach, ART is actually a special case of a method that predates CT, i.e., the method of Kaczmarz [325], which solves sets of equations by successive projections onto the hyperplanes formed by each row of the system. Many modifications of ART have been described that use different cost functions and constraints to resolve ambiguities in the solution and to better account for the presence of noise [323, 324]. It is this ability to deal with inconsistent data, together with the ability to accurately model data formation, that makes ART and other iterative approaches an attractive alternative to their analytic counterparts.

More recently, there has been a great deal of interest in reconstruction of images from photon-limited data that can be modeled as a collection of Poisson random variables [326]. These data arise in nuclear medicine systems in which a patient is injected with a radiolabelled pharmaceutical. These systems detect photons produced either by gamma ray emission (as in single photon emission computed tomography (SPECT)) or by positron-electron annihilation (as in positron emission tomography (PET)). By producing images of the spatial distribution of the radiotracer, physicians are able to study the behavior of the labeled pharmaceutical in the body. In 1982, Shepp and Vardi [327] proposed a maximum likelihood (ML) solution for this Poisson imaging problem based on the expectation-maximization (EM) algorithm. The resulting images are more robust to noise than either FBP or ART. However, due to the large number of unknowns (pixels) to be estimated from a finite data set, the ML images tend to exhibit high variance. Consequently a large number of researchers have studied methods for stabilizing the solutions using penalized-ML or Bayesian estimation techniques with Markov random field priors [328-330]. The Bayesian approach allows inclusion of prior information either in the form of a general smoothness constraint, or more specific structural information that may have been extracted from a co-registered image obtained using a different imaging modality [331].

Recently there has been increasing interest in fully 3-D CT systems in which data are collected either as cone-beam projections (x-ray CT and SPECT) or as 2-D projections of 3-D objects along large numbers of oblique angles (PET). Analytic approaches for reconstruction of images from these data remain a rich area for research [332-334]. Similarly, these data present significant challenges for iterative reconstruction methods [335].

Magnetic resonance (MR) imaging differs from the x-ray and emission CT in the sense that the image Fourier transform or k-space is measured directly. This is achieved by using a magnetic field gradient to produce a spatial frequency encoding of the magnetic resonance signal from hydrogen nuclei in the body. Using combinations of time-varying magnetic field gradients and radio-frequency pulses, it is possible to obtain k-space (Fourier) measurements with a wide range of sampling patterns. Early research in MR sampled k-space using the polar grid shown in Figure 10(b) and reconstructed the images using methods similar to those employed in CT [336]. More commonly, MR data are collected directly on rectangular sample patterns (Figure 10(d)) [337], although fast acquisition can be achieved using more complex patterns such as the spiral scan method [338] (Figure 10(e)). In addition to varying the manner in which the Fourier space is sampled, different pulse sequences can be used to alter contrast in the images through varying the impact of spin-spin and spin-lattice relaxation constants on the resonance signal. MR imaging remains a highly active research field with particular interest in the development of methods for dynamic imaging of the beating heart and functional techniques for studying brain activity, blood flow, and other physiological processes [339, 340].

In addition to the modalities of x-ray CT, PET, SPECT, and MR discussed here, there are an increasing number of experimental medical imaging modalities that use measurement of line integrals, or other mappings between image and data, together with computed inverses, to form images. Examples include ultrasonic tomography in which refraction and diffraction effects produce data that are no longer simple integrals along straight lines [319]. Similarly, electromagnetic imaging methods using frequencies ranging from the quasistatic through optical wavelengths present special challenges in signal processing due to the fact that the data are not simple line integrals and the inverse problems are often highly ill-posed due to combinations of limited data, poor SNR, and the ambiguities inherent in the underlying EM equations.

Outside of the medical realm, computed imaging is employed in many arenas including x-ray crystallography, electron microscopy, seismic imaging, radio astronomy, and SAR. The Fourier transform ties together the theory of imaging in these seemingly unrelated areas. We will only very briefly mention the first three of these types of imaging and then elaborate somewhat on the latter two. In x-ray crystallography the collected data represent samples of the magnitude of the 3-D Fourier transform of the electron density. Given limited Fourier magnitude information (phase is missing), inversion procedures incorporate positivity and a priori information on crystalline structure to successfully form an image [311, 312]. Transmission electron microscopy is a tomographic system where an electron beam is passed through a specimen. The specimen is rotated on a stage to collect line integrals at different angles. The angle cannot be too steep; otherwise the specimen is too thick in the beam direction. This results in a missing cone of data in the Fourier domain. The so-called missing cone problem has received widespread study in both CT and electron microscopy [343]. Seismic imaging techniques exist that use ideas from beamforming [312]. Tomographic imaging is also employed in seismic exploration [311, 344].

We now turn our attention to radio astronomy, where the goal is to make a map of the brightness, in an RF band, of a portion of the "sky." In the 1950s, Ronald Bracewell pioneered projection-based radioastronomical imaging [307, 323], which predated the very similar methods that were later independently developed for medical x-ray CT. In the radioastronomical version of tomographic imaging, the projections are acquired by observing an astronomical target during occlusion by another object. For example, it is possible to image the sun by collecting measurements taken during occlusion by the moon. Suppose that the total energy in the image is measured as a function of time. Then the difference between two successive time measurements will be proportional to the integrated energy in the narrow strip on the sun (a line integral) that is either covered or uncovered during the time period between the two measurements. In practice, it is sometimes possible to collect enough of these measurements to produce an image of very useful resolution. A second form of radio astronomy, which employs interferometry, is more common. An interferometric radio telescope points an array of antennas at the same region in the sky. The individual antennas may be located in the same geographic region, or they may be widely spread out, even on different continents. The Fourier domain imaging principle employed is based on the Van Cittert-Zernike theorem, which basically states that when imaging a spatially uncorrelated source distribution, the correlation of any two antenna outputs provides a sample of the 2-D Fourier transform of the brightness of the sky in the region being imaged [308]. By tracking the same region of the sky for many hours, data is collected from different vantage angles, providing Fourier data lying on an ellipse. Similarly, by performing pairwise correlations for all antennas in the array, Fourier data is obtained lying on many ellipses. The locations of the data ellipses in the Fourier plane depend on the geometry of the array. Typically, the array geometry is selected so that the data ellipses will provide fairly uniform coverage of the Fourier plane. The simplest form of image reconstruction proceeds to interpolate the available Fourier data to a Cartesian grid and then apply an inverse 2-D discrete Fourier transform. When imaging clusters of stars (point sources), more sophisticated imaging algorithms, such as maximum entropy, are employed.

Synthetic aperture radar is used to produce high resolution microwave imagery in all-weather conditions for both surveillance and scientific purposes. The name of this type of radar stems from the fact that instead of using a large antenna array, the SAR concept is to fly a small antenna to the positions that would be occupied by the individual elements in the large array, and then use signal processing to form a high-resolution image. The oldest form of SAR is known as strip-mapping [306], where the antenna remains fixed with respect to the radar platform and thereby sweeps out a strip of terrain. The customary image formation algorithm is correlation-based. A newer algorithm, first applied in geophysics [30] and called the omega-k or wavenumber approach, produces Fourier-domain data as an intermediate step [345, 346]. Jack Walker and his colleagues at ERIM invented a second type of SAR, called spotlight-mode [347]. In spotlight-mode SAR, the radar antenna is steered to illuminate a scene from many different vantage angles. This type of SAR was originally described in range-Doppler terms, but it was later discovered that spotlight-mode SAR can be conveniently explained in terms of the projection-slice theorem from CT [329]. Assuming a stepped-frequency or linear FM transmitted waveform, the signal from the quadrature demodulator in spotlight-mode SAR directly provides polar samples of the Fourier transform of the 2-D scene reflectivity on a small section of a polar grid, offset from the origin, as shown in Fig. 10(f). The angular coverage in the Fourier domain is the same as the range of angles through which the radar views the target. The inner and outer radii of the polar sector are proportional to the lowest and highest frequencies in the transmitted waveform. The polar sector where data is available is generally small and is therefore nearly Cartesian. Thus, simple interpolation suffices prior to computing an inverse Fourier transform for image formation [321]. Although the low-frequency Fourier components are missing in the data, it is still possible to form a high-resolution image because the radar reflectivity is complex with fairly random phase [348]. Spotlight-mode SAR can also be used in an inverse mode, where the radar is stationary and the target moves. This is useful in imaging ships on the ocean surface [349] and has also been used in planetary radar astronomy [350]. Some excellent reference books on SAR include [351] for strip mapping and [352, 353] for spotlight-mode.

We close this section by pointing out that two groups of researchers have driven the development of the field of computed imaging. The first group is characterized by their tie to specific imaging applications and modalities. Researchers from this group have pioneered the development of new imaging modalities in an independent way. Thus, each application area initially has developed its own terminology and its own approach to data modeling and inversion. In recent years, a second group of researchers has come along, drawn primarily from the signal- and image-processing community. Researchers in this second group have an interest in applications, but are experts on Fourier transforms, convolution, spectral analysis, linear algebra, statistical signal processing, and the implementation of the associated numerical operations. Signal processing researchers who have worked on more than one imaging application are in a particularly good position to help provide unification to the field of computed imaging. We are just now at the point where textbooks in this field are becoming available [354- 356].

Conclusions

This article has presented a select overview of progress and advances made in the area of image and multidimensional signal processing over the last three decades. Owing to the availability of inexpensive digital cameras, a multitude of nonvisible sensors and imaging modalities, powerful and affordable computers, and global high-bandwidth networks, this area will continue to witness explosive growth over the next decade and beyond.

It will be interesting to see whether today's major paradigms collected in this "time-capsule" of current contributions will be valid in the distant future. We speculate that, yes, many of them will be useful, but almost certainly they will be augmented by many other new and radically different approaches. While the field originally started out by extending and adapting familiar methods from related disciplines (such as 1-D signal processing), we expect that future progress will come increasingly from methods that are specific to image and multidimensional signals.

Acknowledgments

The editors are thankful to Professor A. Rosenfeld for reading this article several times and improving the presentation. We are also thankful to the IEEE Signal Processing Society Staff for their patience and guidance.

Rama Chellappa is with the University of Maryland, College Park, MD, USA. *Bernd Girod* is with the University of Erlangen-Nuremberg, Germany. *David Munson Jr.* is with the University of Illinois, Urbana-Champaign, IL, USA. *A. Murat Tekalp* is with the University of Rochester, Rochester, NY, USA. *Martin Vetterli* is with the Swiss Federal Institute of Technology, Lausanne, Switzerland and the University of California, Berkeley, CA, USA.

References

1. N.K. Bose, *Applied Multidimensional Systems Theory*, Van Nostrand Rheinhold, New York, NY, 1982.
2. D.E. Dudgeon and R.M. Mersereau, *Multidimensional Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1984.
3. T.S. Huang, ed., *Two-Dimensional Digital Signal Processing, I*, in *Topics in Applied Physics*, 42, Springer-Verlag, Berlin, 1981.
4. T.S. Huang, ed., *Two-Dimensional Digital Signal Processing, II*, in *Topics in Applied Physics*, 43, Springer-Verlag, Berlin, 1981.
5. J.S. Lim, Two-Dimensional Signal and Image Processing, Prentice-Hall, Englewood Cliffs, NJ 1990.
6. IEEE, ASSP Society's MDSP Committee, ed., *Selected Papers in Multidimensional Digital Signal Processing*, IEEE Press, New York, NY, 1986.
7. D.P. Petersen and D. Middleton, "Sampling and reconstruction of wave-number limited functions in N-dimensional Euclidean spaces," *Information and Control*, 5, pp. 279-323, 1962.
8. R.M. Mersereau, "The processing of hexagonally sampled two-dimensional signals," *Proc. IEEE*, 67, pp. 930-49, 1979.
9. T.S. Huang, "Two-dimensional windows," *IEEE Trans. Audio Electroacoust.*, 20, pp. 88-90, 1972.
10. D.B. Harris, J.H. McClellan, D.S.K. Chan, and H.W. Schuessler, "Vector radix fast Fourier transform," *Proc. Int. Conf. Acoust., Speech, Signal Proc.*, pp. 548-551, 1977.
11. G.K. Rivard, "Direct fast Fourier transform of bivariate functions," *IEEE Trans. Acoust. Speech, Signal Proc.*, 25, pp. 250-252, 1977.
12. E. Dubois, "The sampling and reconstruction of time-varying imagery with applications in video systems," *Proc. IEEE*, 73, pp. 502-522, 1985.
13. P.P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice-Hall, Englewood Cliffs, NJ, 1993.
14. R. DeCarlo, J. Murray, and R. Saeks, "Multivariate nyquist theory," *Int. J. Control*, 25, pp. 657-675, 1977.
15. E.I. Jury, *Inners and Stability of Dynamic Systems*, Wiley Interscience, New York 1974.
16. J.L. Shanks, S. Treitel, and J. H. Justice, "Stability and synthesis of two-dimensional recursive filters," *IEEE Trans. Audio Electroacoust*, 20, pp. 115-128, 1972.
17. M.G. Strintzis, "Test of stability of multidimensional filters," *IEEE Trans. Circ. Syst.*, 24, pp. 432-437, 1977.
18. M.P. Ekstrom, R.E. Twogood, and J.W. Woods, "Two-dimensional recursive filter design—a spectral factorization approach," *IEEE Trans. Acoust., Speech, Signal Proc.*, 28, pp. 16-26, 1980.
19. D.B. Harris and R.M. Mersereau, "A comparison of algorithms for minimax design of two-dimensional linear phase FIR digital filters," *IEEE Trans. Acoust., Speech, Signal Proc.*, 25, pp. 492-500, 1977.
20. J.H. Lodge and M.M. Fahmy, "An efficient *lp* optimization technique for the design of two-dimensional linear-phase FIR digital filters," *IEEE Trans. Acoust., Speech, Signal Proc.*, 28, pp. 308-313, 1980.
21. A.E. Cetin, O.N. Gerek, and Y. Yardimci, "Equiripple FIR filter design by the FFT algorithm," *IEEE Signal Processing Magazine*, 14, pp. 60-64, 1997.

22. R.M. Mersereau, W.F.G. Mecklenbräuker, and T.F. Quatieri, Jr., "McClellan transformations for 2-D digital filtering: I. Design," *IEEE [illegible]*, 23, pp. [illegible], 1976.

23. J.H. McClellan and D.S.K. Chan, "A 2-D FIR filter structure derived from the Chebyshev recursion," *IEEE Trans. Circuits Syst.*, 24, pp. [illegible], 1977.

24. J.H. McClellan, "Multidimensional spectral estimation," *Proc. IEEE*, 70, pp. 1029-1039, 1982.

25. G. Sharma and R. Chellappa, "A model-based approach for the estimation of 2-D maximum entropy power spectra," *IEEE Trans. Inform. Theory*, 31, pp. 90-99, 1985.

26. S.W. Lang and J.H. McClellan, "Multidimensional MEM spectral estimation," *IEEE Trans. Acoust., Speech, Signal Proc.*, 30, pp. 880-887, 1982.

27. J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, 57, pp. 1408-1418, 1969.

28. H.J. Scudder, "Introduction to computer aided tomography," *Proc. IEEE*, 66, pp. 628-637, 1978.

29. D.C. Munson, Jr., J.D. O'Brien, and W.K. Jenkins, "A tomographic formulation of spotlight-mode synthetic aperture radar," *Proc. IEEE*, [illegible], pp. 917-925, 1983.

30. R.H. [illegible], "[illegible] Fourier transform," [illegible], pp. [illegible], 19[illegible].

31. M.H. Hayes, "The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform," *IEEE Trans. Acoust., Speech, Signal Proc.*, 30, pp. 140-154, 1982.

32. A. Zakhor and A.V. Oppenheim, "Reconstruction of two-dimensional signals from level crossings," *Proc. IEEE*, [illegible], pp. 31-55, 1990.

33. G. Sharma and H.J. Trussell, "Digital Color Imaging," *IEEE Trans. Image Proc.*, 6, pp. 901-932, 1997.

34. P.L. Vora, M.L. Harville, J.E. Farrell, J.D. Tietz, and D.H. Brainard, "Image capture: synthesis of sensor responses from multispectral images," *Proc. [illegible] IS&T/SPIE Conf. [illegible]*, San Jose, CA, 3018, pp. 2-11, 1997.

35. D.H. Brainard, "Bayesian method for reconstructing color images from trichromatic samples," in *IS&T 47th Annual Conf. [illegible] (ICPS'94): The Physics and Chemistry of Imaging Systems*, 2, pp. [illegible]-380, 1994.

36. D.R. Cok, "Reconstruction of CCD images using template matching," in *IS&T [illegible] Annual Conf. [illegible] (ICPS'94) [illegible]*, 2, pp. 380-385, 1994.

37. P.L. Vora and H.J. Trussell, "Mathematical methods for the design of color scanning filters," *IEEE Trans. Image Proc.*, 6, pp. 312-320, 1997.

38. M.J. Vrhel and H.J. Trussell, "Optimal color filters in the presence of noise," *IEEE Trans. Image Proc.*, 4, pp. 814-823, 1995.

39. M. Wolski, J.P. Allebach, C.A. Bouman, and E. Walowit, "Optimization of sensor response functions for colorimetry of reflective and emissive objects," *IEEE Trans. Image Proc.*, 5, pp. 507-517, 1996.

40. G. Sharma and H.J. Trussell, "Figures of merit for color scanners," *IEEE Trans. Image Proc.*, 6, pp. 990-1001, 1997.

41. G. Sharma, H.J. Trussell, and M.J. Vrhel, "Optimal nonnegative color scanning filters," *IEEE Trans. Image Proc.*, 7, pp. 129-133, 1998.

42. B.A. Wandell, "The synthesis and analysis of color images," *IEEE Trans. Pattern Anal. Mach. Intell.*, 9, pp. 2-13, 1987.

43. D. Marimont and B.A. Wandell, "Linear models of surface and illuminant spectra," *J. Opt. Soc. Am. A*, 9, pp. 1905-1913, 1992.

44. M.J. Vrhel, R. Gershon, and L.S. Iwan, "Measurement and analysis of object reflectance spectra," *Color Res. Appl.*, 19, pp. 4-9, 1994.

45. J. Ho, B.V. Funt, and M.S. Drew, "Separating a color signal into illumination and surface reflectance components: theory and applications," *J. Opt. Soc. Am. A*, 9, pp. 1905-1913, 1992.

46. M.C. Stone, W.B. Cowan, and J.C. Beatty, "Color gamut mapping and the printing of digital color images," *ACM Trans. Graphics*, [illegible], pp. 249-292, 1988.

47. R.S. Gentile, E. Walowit, and J.P. Allebach, "A comparison of techniques for color gamut mismatch compensation," *J. Imaging Technol.*, 16, pp. 176-181, 1990.

48. E.D. Montag and M.D. Fairchild, "[illegible] gamut mapping [illegible] images and a [illegible]," *IEEE Trans. Image Proc.*, 6, pp. 977-989, 1997.

49. R.S. Berns, R.J. Motta, and M.E. Gorzynski, "CRT colorimetry, Part I: Theory and practice," *Color Res. Appl.*, 18, pp. 299-314, 1993.

50. R.S. Berns, M.E. Gorzynski, and R.J. Motta, "CRT colorimetry, Part II: Metrology," *Color Res. Appl.*, 18, pp. 315-325, 1993.

51. C.B. [illegible] and R. Balasubramanian, "[illegible]," [illegible], pp. [illegible], 1993.

52. Z. [illegible], [illegible], "[illegible]," [illegible], pp. [illegible], 1996.

53. F. [illegible], "[illegible]," [illegible], 26, pp. [illegible], 1982.

54. J. [illegible] and M. [illegible], "[illegible]," [illegible], 1990.

55. M.E. [illegible] and C. [illegible], "[illegible]," [illegible], 1990.

56. R. [illegible], "[illegible]," [illegible], 1996.

57. [illegible], "[illegible]," [illegible], 1993.

58. [illegible], "[illegible]," [illegible], 1990.

59. [illegible], "[illegible]," *IEEE Trans. Image Proc.*, [illegible], 1990.

60. [illegible], "[illegible]," [illegible], 1996.

61. [illegible], "[illegible]," [illegible], 1995.

62. [illegible], "[illegible]," [illegible], 1990.

63. [illegible], "[illegible]," [illegible], 1994.

64. [illegible], "[illegible]," [illegible], 1994.

65. [illegible], "[illegible]," [illegible], 1998.

66. K.R. Rao, [illegible], New York, NY, 1985.

67. [illegible], *[illegible]*, Wiley, New York, NY, 1985.

68. R.M. [illegible], *[illegible]*, [illegible], MA, 1992.

69. A.K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, Englewood Cliffs, NJ, 1989.

70. N. [illegible] and K.R. Rao, "[illegible]," *IEEE [illegible]*, 88, pp. [illegible], 19[illegible].

71. W.B. Pennebaker and J.L. Mitchell, *JPEG: Still Image Compression Standard*, Van Nostrand Rheinhold, New York, NY, 1993

72. H.S. Malvar. *Signal Processing with Lapped Transforms*. Artech House, Norwood, MA, 1992.

73. M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, Prentice-Hall, Englewood Cliffs, NJ, 1995.

74. I. Daubechies, *Ten Lectures on Wavelets*. SIAM, Philadelphia, PA, 1992.

75. C. Herley, J. Kovacevic, K. Ramchandran and M. Vetterli, "Tilings of the Time-Frequency Plane: Construction of Arbitrary Orthogonal Bases and Fast Tiling Algorithms," *IEEE Trans. Signal Proc., Special issue on Wavelets and Signal Processing*, 41, pp. 3341-3359, 1993.

76. A. Rosenfeld, *Image Models*, Academic Press, New York, NY, 1980.

77. U. Grenander, *Lectures in Pattern Theory, Vols. I-III*, Springer-Verlag, New York, NY, 1981.

78. A. Rosenfeld and L.S. Davis, "Image segmentation and image models," *Proc. IEEE*, 67, pp.762-772, 1979.

79. B.K.P. Horn, *Shape from Shading*, MIT Press, Boston, MA, 1989.

80. L. Wolff, S. Shafer and G. Healey, *Physics-Based Vision, Principles and Practice, Shape Recovery*, Jones and Barrett, Boston, MA, 1992.

81. B. Mandelbrot, *Fractals and a New Geometry of Nature*, Freeman, San Francisco, CA, 1981.

82. M. Barnsley, *Fractals Everywhere*, Academic Press, New York, NY, 1988.

83. B.R. Hunt, "Bayesian methods in nonlinear image restoration," *IEEE Trans. Computers*, 26, pp.219-229, 1975.

84. R.M. Haralick, "Digital step edges from zero crossings of second directional derivatives," *IEEE Trans. Patt. Anal. Mach. Intell.*, 6, pp. 58-68, 1984.

85. M.S. Longuet-Higgins, 'Statistical properties of an isotropic random surface," Phil. Trans. Royal Soc., London, A, 250, pp. 151-171, 1957.

86. R.M. Bolle and D.B. Cooper, "Bayesian recognition by approximating image intensity functions with quadratic polynomials," *IEEE Trans. Patt. Anal. Mach. Intell.*, 6, pp. 418-429, 1984.

87. B. Cernuschi-Frias, et al., "Toward a model-based Bayesian theory for estimating and recognizing parameterized 3-D objects using two or more images taken from different positions," *IEEE Trans. Patt. Anal. Mach. Intell.*, 11, pp. 1028-1052, 1989.

88. P.A. Maragos, R.W. Shafer and R.M. Mersereau, "Two-dimensional linear prediction and its applications to adaptive predictive coding of images," *IEEE Trans. Acoust. Speech and Signal Processing*, 32, pp.1213-1229, 1984.

89. J.W. Woods and C.H. Radewan, "Kalman-filtering in two dimensions," *IEEE Trans. Inform. Theory*, 23, pp. 473-482, 1977.

90. P. Whittle, "On stationary processes in the plane," Biometrika, 41, pp. 434-449, 1954.

91. Y.A. Rosanov, "On Gaussian fields with given conditional distributions," *Theory of Probability and Its Applications*, 11, pp. 381-391, 1967.

92. J.W. Woods, "Two-dimensional discrete Markovian fields," *IEEE Trans. Inform. Theory*, 18, pp. 232-240, 1972.

93. J. Besag, "Spatial interaction and the statistical analysis of lattice systems", Jl. Royal Stat. Soc., Ser. B, 48, pp. 192-236, 1974.

94. R.L. Kashyap, "Analysis and synthesis of image patterns by spatial interaction models," *Progress in Pattern Recognition*, L.N. Kanal and A. Rosenfeld (eds.), 1, North-Holland, Amsterdam, 1981.

95. R.L. Kashyap and R. Chellappa, "Estimation and choice of neighbors in spatial interaction models of images," *IEEE Trans. Inform. Theory*, 29, pp. 60-72, 1983.

96. G.R. Cross and A.K. Jain, "Markov random field texture models," *IEEE Trans. Patt. Anal. Mach Intell.*, 5, pp. 25-39, 1983.

97. J.W. Modestino, R.W. Fries and A.L. Vickers, "Stochastic models generated by random tessellations of the plane," *Comput. Graph. and Image Proc.*, 12, pp.74-98, 1980.

98. N. Ahuja and A. Rosenfeld, "Mosaic models for textures," *IEEE Trans. Patt. Anal. Mach. Intell.*, 3, pp. 1-11, 1981.

99. J.W. Woods, "Two-dimensional Kalman filters," in *Two-Dimensional Digital Signal Processing* (ed. T.S. Huang), 42, *Topics in Applied Physics*, Springer-Verlag, New York, pp. 155-205, 1981.

100. S. Geman and D. Geman. "Stochastic relaxation, Gibbs distributions and Bayesian restoration of images," *IEEE Trans. Patt. Anal. Mach. Intell.*, 6, pp. 721-741, 1984.

101. F.C. Jeng and J.W. Woods, "Simulated annealing in compound Gaussian random fields," *IEEE Trans. Inform Theory*, 36, pp. 94-107, 1990.

102. A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, Boston, MA, 1987.

103. S. Lakshmanan and H. Derin, "Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing," *IEEE Trans. Patt. Anal. Mach. Intell.*, 11, pp. 799-822, 1989.

104. S. Geman and C. Graffigne, "Boundary detection by constrained optimization," *IEEE Trans. Patt. Anal. Mach. Intell.*, 12, pp. 609-628, 1990.

105. T. Simchony, R. Chellappa and Z. Lichenstein, "Pyramid implementation of optimal step conjugate gradient algorithms for some computer vision problems," *IEEE Trans. Syst., Man, Cybern*, 19, pp. 1408-1425, 1989.

106. M.J. Black and A. Rangarajan, "On the unification of line processes, outlier rejection and robust statistics," *Intl. Jl. Computer Vision*, 19, pp. 57-92, 1996.

107. S.A. Hovanesian, *Introduction to Sensor Systems*, Artech House, Norwood, MA, 1988.

108. J. Shapiro, B. Capron and R. Harney, "Imaging and target detection with a heterodyne-reception optical radar," *Applied Optics*, 20, pp.3292-3313, 1981.

109. D. Andersh, S. Lee, H. Ling and C. Yu, "XPATCH: A high frequency electromagnetic scattering prediction code using shooting and bouncing ray," *Proc. Ground Target Modeling and Validation Conference*, pp. 498-507, 1994.

110. M.L. VanBlaricum. *Introduction to Ultra-Wideband Radar Systems*, CRC Press, 1995.

111. A. Rosenfeld and A.C. Kak, *Digital Picture Processing, 1*, Academic Press, Boston, second edition, 1982.

112. I. Pitas and A. N. Venetsanopoulos, *Nonlinear Digital Filters: Principles and Applications*, Kluwer Academic, Boston, MA, 1990.

113. J. Astola, *Fundamentals of Nonlinear Digital Filtering*, CRC Press, Boca Raton, FL, 1997.

114. J.W. Tukey, *Exploratory Data Analysis*, Addison Wesley, Reading, MA, 1977.

115. B.I. Justusson, "Median filtering: Statistical Properties," in *Two-dimensional Digital Signal Processing*, T.S. Huang (ed.), 42, *Topics in Applied Physics*, pp. 161-196, Springer-Verlag, Berlin, 1981.

116. A.C. Bovik, T. S. Huang, and D.C. Munson, "A generalization of median filtering using linear combinations of order statistics," *IEEE Trans. on Acoust., Speech and Signal Proc.*, 31, pp. 1345-1349, 1983.

117. E.J. Coyle, J.-H. Lin and M. Gabbouj, "Optimal stack filtering and the estimation and structural approaches to image processing," *IEEE Trans. on Acoust., Speech, Signal Proc.*, 37, pp. 2037-2066, 1989.

118. G.R. Arce and R.E. Foster, "Detail preserving ranked-order-based filters for image processing," *IEEE Trans. on Acousti., Speech Signal Proc.*, 37, pp. 83-98, 1989.

119. M. Gabbouj and E.J. Coyle, "Minimum mean absolute error stack filtering with structural constrains and goals," *IEEE Trans. on Acoust., Speech, Signal Proc.*, 38, pp. 955-968, 1990.

[120] S. Daly, "The visible differences predictor: an algorithm for the assessment of image fidelity," Chapter 14 in Digital Images and Human Vision, ed. by A.B. Watson, MIT Press, Boston, MA, 1993.

[121] T.J. Huang and F.J. Cowie, "Prospects for optimal restoration of images [illegible]," [illegible] Machine Intell., 1967.

[122] A.K. Katsaggelos (ed.), "Digital Image Restoration," Springer Series in [illegible] 23, Springer Verlag, 1991.

[123] H.C. Andrews and B.R. Hunt, Digital Image Restoration, Prentice Hall, Englewood Cliffs, NJ, 1977.

[124] M.R. Banham and A.K. Katsaggelos, "Digital Image Restoration," IEEE Signal Processing Magazine, vol. 14, no. 2, [illegible], 1997.

[125] J.C. Brailean, R.P. Kleihorst, S.N. Efstratiadis, A.K. Katsaggelos, R.L. Lagendijk, "Noise reduction filters for dynamic image sequences: a review," Proc. IEEE, vol. 83, pp. 1272-1292, 1995.

[126] R. Rosenholtz and A. Zakhor, "Iterative procedures for reduction of blocking effects in transform image coding," IEEE Trans. Circ. Syst. Video Technol., vol. 2, pp. 91-95, 1992.

[127] Y. Yang, N.P. Galatsanos, and A.K. Katsaggelos, "Set theoretic spatially adaptive reconstruction of block-transform compressed images," IEEE Trans. Image Proc., vol. 4, pp. 896-908, 1995.

[128] R.R. Hemami, [illegible] "[illegible] least squares [illegible] image restoration [illegible]," IEEE Trans. [illegible], pp. 856-875, 1995.

[129] H. Stark (ed.), Image Recovery: Theory and Application, Academic Press, 1987.

[130] J. Biemond, R.L. Lagendijk, and R.M. Mersereau, "Iterative methods for image deblurring," Proc. IEEE, vol. 78, pp. 856-883, 1990.

[131] A.K. Katsaggelos, "Iterative image restoration algorithms," Optical Engineering, [illegible] pp. [illegible], 1989.

[132] R.L. Lagendijk and J. Biemond, Iterative Identification and Restoration of Images, Boston, Kluwer Academic Press, MA, 1990.

[133] A. Kundur and D. Hatzinakos, "Blind image deconvolution," IEEE Signal Processing Magazine, vol. 13, pp. 43-64, May 1996 and vol. 13, pp. 61-63, Nov. 1996.

[134] A. Rosenfeld and A.C. Kak, Digital Picture Processing, 2nd ed., Academic Press, New York, NY, 1982.

[135] S.G. Mallat, "Multifrequency channel decompositions of images and wavelet models," IEEE Trans. Acoust., Speech, Signal Proc., vol. 37, pp. 2091-2110, 1989.

[136] D. Marr and E. Hildreth, "Theory of edge detection," Proc. R. Soc. Lond. B, vol. 207, pp. 187-217, 1980.

[137] J.F. Canny, "A computational approach to edge detection," IEEE Trans. Pattern Anal. Machine Intell., vol. 8, pp. 679-698, 1986.

[138] S. Mallat and W.L. Hwang, "Singularity detection and processing with wavelets," IEEE Trans. Inform. Theory, vol. 38, 1992.

[139] P. Maragos, "A representation theory for morphological image and signal processing," IEEE Trans. Pattern Anal. Machine Intell., vol. 11, pp. 586-599, 1989.

[140] P. Maragos, "Differential morphology and image processing," IEEE Trans. Image Proc., vol. 5, pp. 922-937, 1996.

[141] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," IEEE Trans. Pattern Anal. Mach. Intell., vol. 12, pp. 629-639, 1990.

[142] [illegible], "Image segmentation, smoothing and edge detection by [illegible] diffusion," SIAM J. Numerical Analysis, vol. 29, pp. 845-866, 1992.

[143] S. [illegible], "[illegible]," [illegible], 1998.

[144] R.M. Haralick, "Statistical and structural approaches to texture," Proc. IEEE, vol. 67, pp. 786-804, 1979.

[145] P. Brodatz, Textures, [illegible], 1966.

[146] F.S. Cohen and D.B. Cooper, "Simple parallel hierarchical and relaxation algorithms for segmenting noncausal Markovian random fields," IEEE Trans. Pattern Anal. Machine Intell., vol. 9, pp. 195-219, 1987.

[147] A.C. Bovik, M. Clark, and W.S. Geisler, "Multichannel texture analysis using localized spatial filters," IEEE Trans. Pattern Anal. Machine Intell., vol. 12, pp. 55-73, 1990.

[148] C.A. Bouman and M. Shapiro, "[illegible]," [illegible], 1994.

[149] A.C. Bovik, M. Clark, and W.S. Geisler, "Multichannel texture analysis using localized spatial filters," [illegible], pp. 55-73, 1990.

[150] A.C. Bovik, [illegible], IEEE Trans. Signal Proc., vol. 39, pp. 2025-2043, 1991.

[151] D. Dunn and W.E. Higgins, "Optimal Gabor filters for texture segmentation," IEEE Trans. Image Proc., vol. 4, pp. 947-964, 1995.

[152] A.C. Bovik, [illegible], IEEE Trans. [illegible], vol. 38, pp. 691-[illegible], 1992.

[153] P. Maragos and A.C. Bovik, "Image demodulation using multidimensional energy separation," J. Opt. Soc. Amer. A, vol. 12, pp. 1867-1876, 1995.

[154] J.P. Havlicek, D.S. Harding, and A.C. Bovik, "[illegible] AM-FM image models," [illegible], pp. 1097-1100, 1996.

[155] S. Zhu, Y. Wu, and D. Mumford, "[illegible] (FRAME) [illegible]," [illegible], 1996.

[156] J. Serra and [illegible], [illegible], 1997.

[157] B.S. Manjunath and W.Y. Ma, "Texture features for browsing and retrieval of image data," IEEE Trans. Pattern Anal. Machine Intell., vol. 18, pp. 837-842, 1996.

[158] C.T. Zahn and R.Z. Roskies, "Fourier descriptors for plane closed curves," IEEE Trans. Comp., vol. 21, pp. 269-281, 1972.

[159] E. Persoon and K.S. Fu, "Shape discrimination using Fourier descriptors," IEEE Trans. Sys. Man, Cybern., vol. 7, pp. 170-179, 1977.

[160] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," J. Opt. Soc. Am. A, vol. 4, pp. 519-524, 1987.

[161] M.A. Turk and A.P. Pentland, "Face recognition using eigenfaces," IEEE Conf. on Computer Vision and Pattern Recognition, [illegible], 1991.

[162] M.K. Hu, "Visual pattern recognition by moment invariants," IRE Trans. Inform. Theory, vol. 8, pp. 179-187, 1962.

[163] S.O. Belkasim, M. Shridhar, and M. Ahmadi, "Pattern recognition with moment invariants," Pattern Recog., [illegible], 1991.

[164] R.L. Kashyap and R. Chellappa, "Stochastic models for closed boundary analysis: representation and reconstruction," IEEE Trans. Inform. Theory, vol. 27, pp. 627-637, 1981.

[165] S.R. Dubois and F.H. Glanz, "An autoregressive model approach to two-dimensional shape classification," IEEE Trans. Pattern Anal. Machine Intell., vol. 8, pp. 55-66, 1986.

166. F. Mokhtarian and A.K. Mackworth, "A theory of multiscale, curve-based shape representation for planar curves," *IEEE Trans. Patt. Anal. Mach. Intell.*, 14, pp. 789-805, 1992.

167. G.C.H. Chuang and C.C. Jay Kuo, "Wavelet descriptor of planar curves: theory and applications," *IEEE Trans. Image Processing*, 5, pp. 56-70, 1996.

168. R. Duda and P. Hart, *Pattern Classification and Scene Analysis*, John Wiley and Sons, New York, 1973.

169. R.L. Kashyap and A. Khotanzad, "A model-based method for rotation invariant texture classifcation," *IEEE Trans. Patt. Anal. Mach. Intell.*, 8, pp. 472-481, 1986.

170. K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," *Jl. Opt. Soc. Am., A*, 14, pp. 1724-1733, 1997.

171. G.L. Martin and J.A. Pittman, "Recognizing hand-printed letters and digits," *Neural Networks*, 3, pp. 258-267, 1991.

172. S. Haykin, *Neural Networks - A Comprehensive Foundation*, Macmillan College Publishing Company, New York, NY, 1994.

173. W.E.L. Grimson, *Object Recognition by Computer: The Role of Geometric Constraints*, MIT Press, Boston, MA, 1990.

174. J.L Blue, et al., "Evaluation of pattern classifiers for fingerprint and OCR applications," *Pattern Recognition*, 27, pp. 485-501, 1994.

175. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufman, San Mateo, CA, 1988.

176. M. Garris, J. Blue and C.L. Wilson, "Neural network based approach for optical character recognition" (to appear in *IEEE Trans. Image Processing*, 1998).

177. Y. LeCun et al., "Back propagation applied to handwritten zip code recognition," *Neural Computation*, 1, pp. 541-551, 1989.

178. T. Poggio and F. Girosi, "Networks for approximation and learning," *Proc. IEEE*, 78, pp. 1481-1497, 1990.

179. R. Brunelli and T. Poggio, "Face recognition: features versus templates," *IEEE Trans. Patt. Anal. Mach. Intell.*, 15, pp. 1042-1052, 1993.

180. G.L. Martin, "Centered-object integrated segmentation and recognition of overlapping handprinted characters," *Neural Computation*, 5, pp. 419-429, 1993.

181. Y. Amit and D. German, "Shape quantization and recognition with randomized trees," *Neural Computation*, 9, pp. 1545-1588, 1997.

182. B. Bhanu, D.E. Dudgeon, E.G. Zelnio, A. Rosenfeld, D. Casasent and I.S. Reed, Special Issue on Automatic Target Detection and Recognition, *IEEE Trans. Image Processing*, 6, 1997.

183. F. Dufaux and F. Moscheni, "Motion estimation techniques for digital TV: a review and a new contribution," *Proc. IEEE*, 83, pp. 858-876, 1995.

184. A. Mitiche and P. Bouthemy, "Computation and analysis of image motion: a synopsis of current problems and methods," *International Journal of Computer Vision*, 19, pp. 29-55, 1996

185. T. Aach, A. Kaup and R. Mester "Statistical model-based change detection in moving video," *Signal Processing*, 31, 1993.

186. B.K.P. Horn and B.G. Schunck, "Determining Optical Flow," *Artificial Intelligence*, 17, pp. 185-203, 1981.

187. F. Rocca, "Television bandwidth compression utilizing frame-to-frame correlation and movement compensation," in *Picture Bandwidth Compression* (Record of the Symposium on Picture Bandwidth Compression, Cambridge, MA, T.S. Huang and O.J. Tretiak, eds.), Gordon and Breach, New York, pp. 675-693, 1972.

188. P. Moulin, R. Krishnamurthy and J.W. Woods, "Multiscale modeling and estimation of motion fields for video coding," *IEEE Trans. Image Processing*, 6, pp. 1606-1620, 1997.

189. E. Dubois and J. Konrad, "Estimation of 2-D motion fields from image sequences with application to motion-compensated processing," in *Motion Analysis and Image Sequence Processing*, M.I. Sezan and R.L. Lagendijk, eds., Kluwer Academic Publishers, Boston, MA, pp. 53-87, 1993.

190. P. Burt and P. Anandan, "Image stabilization by registration to a reference mosaic," *Proc. DARPA Image Understanding Workshop*, Monterey, CA, pp. 425-434, 1994.

191. M. Hansen, et al., "Real-time scene stabilization and mosaic construction," *Proc. DARPA Image Understanding Workshop*, Monterey, CA, pp. 457-465, 1994.

192. C. Morimoto, Design and Evaluation of Eletronic Image Stabilization Algorithms, Ph.D. Dissertation, Department of Computer Science, University of Maryland, College Park, 1997.

193. B. Rousso, S. Peleg and I. Finci, "Mosaicking with generalized strips," Proc. DARPA Image Understanding Workshop, New Orleans, LA, pp. 255-260, 1997.

194. J. Roach and J.K. Aggarwal, "Determining the movement of objects from a sequence of images," *IEEE Trans. on Patt. Anal. Mach. Intell.*, 2, pp. 554-562, 1980.

195. H.C. Longuet-Higgins, "The reconstruction of a scene from two projections-configurations that defeat the 8-point algorithm, *Proc. IEEE Conf. on Artificial Intelligence Applications*, Denver, CO, pp. 395-397, 1984.

196. R.Y. Tsai and T.S. Huang, "Estimating three-dimensional motion parameters of a rigid planar patch," *IEEE Trans. On Acoust., Speech, Signal Processing*, 29, pp. 1147-1152, 1981.

197. R.Y. Tsai, T.S. Huang, and W.L. Zhu, "Estimating three-dimensional motion parameters of a rigid planar patch, II: singular value decomposion," *IEEE Trans. on Acoust., Speech, Signal Processing*, 30, pp. 525-534, 1982.

198. R.Y. Tsai and T.S. Huang, "Estimating three-dimensional motion parameters of a rigid Planar Patch, III: finite point correspondences and the three view problem," *IEEE Trans. on Acoust., Speech, Signal Processing*, 32, pp. 213-220, 1984.

199. R.Y. Tsai and T.S. Huang, "Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces," *IEEE Trans. on Patt. Anal. Mach. Intell.*, 6, pp. 13-27, 1984.

200. J.Q. Fang and T.S. Huang, "Some experiments on estimating the 3-D motion parameters of a rigid body from two consecutive image frames," *IEEE Trans. on Patt. Anal. Mach. Intell.*, 6, pp. 545-554, 1984.

201. T.J. Broida and R. Chellappa, "Estimation of object motion parameters from sequence of noisy images," *IEEE Trans. Patt. Anal. Mach. Intell.*, 8, pp. 90-99, 1986.

202. T.J. Broida, S. Chandrashekhar and R. Chellappa, "Recursive techniques for estimation of 3-D translation and rotation parameters from noisy image sequences," IEEE Trans. Aerospace and Electronic Systems, 26, pp. 639-656, 1990.

203. Z. Zhang and O. Faugeras, "3-D Dynamic Scene Analysis," *Springer Series in Information Sciences*, 27, Springer-Verlag, New York, 1992.

204. C. Poelman and T. Kanade, "A paraperspective factorization method for shape and motion recovery," IEEE Trans. Patt. Anal. Mach. Intell. 19, pp. 206-218, 1997.

205. J.A.Webb and J.K. Aggarwal, "Structure from motion of rigid and jointed objects," Artificial Intelligence, 19, pp. 107-130, 1982.

206. D.D. Hoffmann and B.E. Flinchbaugh, "Interpretation of biological motion," *Biol. Cybernetics*, 42, pp. 195-204, 1982.

207. R.J. Holt, T.S. Huang, A.N. Netravali, and R.J. Qian, Determining articulated motion from perspective views: a decomposition approach, *Pattern Recn.*, 30, pp. 1435-1449, 1997.

208. I. Haritaoglu, D. Harwood and L. S. Davis, "W4 - A real time system for detection and tracking of people and their parts," *Proc. Conf. Face and Gesture Recognition*, Nara, Japan, 1998.

209. D. Metaxas and D. Terzopoulos, "Constrained deformable superquadrics and nonrigid motion tracking," *Proc. IEEE Conf. on Computer Vision and Patt. Recn.*, Maui, HI, pp. 337-343.

260. B. Girod, "Rate-constrained motion estimation," *in Visual Communication and Image Proc. VCIP'94*. A.K. Katsaggelos (ed.), *Proc. SPIE 2308*, pp. 1026-1034. 1994.

261. P.A. Chou, T. Lookabaugh, and R.M. Gray, "Entropy constrained vector quantization," *IEEE Trans. Acoust., Speech, and Signal Proc.*, 37, pp. 31-42, 1989.

262. K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Trans. on Image Processing*, 2, pp. 160-175, 1993.

263. A. Ortega, K. Ramchandran and M. Vetterli, "Optimal trellis-based buffered compression and fast approximations," *IEEE Trans. on Image Processing*, 3, pp. 26-40, 1994

264. T.M. Cover and J.A. Thomas, *Elements of Information Theory*. Wiley Interscience, New York, NY, 1991.

265. U. Horn and B. Girod, "Perfomance analysis of multiscale motion compensation techniques in pyramid coders," *Proc. IEEE Int. Conf. on Image Proc.*, Lausanne, Switzerland, pp. 255-258. 1996.

266. B. Hochwald and K. Zeger, "Trade-off between source and channel coding," *IEEE Trans. on Inform. Theory*, 43, pp. 1412-1424, 1997.

267. E. Steinbach, N. Farber and B. Girod, "Standard compatible extension of H.263 for robust video transmission in mobile environments," *IEEE Trans. on Circ. and Syst. for Video Technology*, 7, pp. 872-881, 1997.

268. K. Ramchandran, A. Ortega, K.M. Uz and M. Vetterli, "Multiresolution broadcast for digital HDTV using joint source-channel coding," *IEEE JSAC*, Special issue on High Definition Television and Digital Video Commun., 11, pp. 6-23, 1993.

269. S. McCanne, "Scalable Compression and Transmission of Internet Multicast Video," PhD Diss., Dept. of EECS, UC Berkeley, 1996.

270. V.A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. on Inform. Theory*, 39, pp. 821-824, 1993.

271. M.F. Barnsley and A. Jacquin, "Application of recurrent iterated function systems to images," *Proc. SPIE*, 1001, pp. 122-131, 1988.

272. G.M. Davis, "A wavelet-based analysis of fractal image compression," *IEEE Trans. on Image Processing*, 1998.

273. D. Pearson, "Developments in Model-based Video Coding", *Proc. IEEE*, 83, pp. 892-906, 1995.

274. D. Pearson, "Variability of Performance in Video Coding," *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Proc.*, pp. 5-8, Munich, Germany, 1997.

275. G. Nagy, "Optical scanning digitizers," *IEEE Computer*, 16, pp. 13-24, 1983.

276. The Chromatic media processor: http://www.chromatic.com/

277. A. Peleg, U. Weiser, "MMX technology extension to the Intel architecture," *IEEE Micro*, 16, pp. 42-50, 1996.

278. A. Peleg, S. Wilkie, U. Weiser, "Intel MMX for multimedia PCs," *Communications of the ACM*, 40, pp. 25-38, 1997.

279. T.S. Huang, W.F. Schreiber, and O.J. Tretiak, "Image processing," *Proceedings of the IEEE*, 59, pp. 1335-1346, 1971.

280. H.C. Andrews, A.G. Tescher, and R.P. Kruger, "Image processing by digital computer," *IEEE Spectrum*, 9, pp. 20-32, 1972.

281. M.D. McFarlane, "Digital pictures fifty years ago," *Proc. IEEE*, 60, pp. 768-770, 1972.

282. P. Ruetz and R. Brodersen, "Architectures and design techniques for real-time image processing ICs," *IEEE Journal of Solid-State Circuits*, 22, pp. 233-250, 1987.

283. K. Hwang (editor), *Computer Architectures for Image Processing, Special Issue of IEEE Computer*, 16, 1983.

284. A. Choudhary and S. Ranka (editors), *Parallel Processing for Computer Vision and Image Understanding, Special Issue of IEEE Computer*, 25, 1992.

285. K. Shen, G.W. Cook, L.H. Jamieson, E.J. Delp, "An overview of parallel processing approaches to image and video compression," *Proceedings of the SPIE Conference on Image and Video Compression*, 2186, San Jose, CA, pp. 197-208, 1994.

286. J. Allen, "Conmputer Architectures for Signal Processing," *Proc. of the IEEE*, 63, 5, pp. 624-633, 1975.

287. P. Lapsley, J. Bier. A. Shoham, E.A. Lee, *DSP Processor Fundamentals*, IEEE Press, 1997.

288. Texas Instruments TMS320C67: http://www.ti.com/sc/docs/dsps/products/c67x/index.htm

289. Analog Devices SHARC processor: http://www.analog.com/products/sheets/ADSP21060.html

290. P. Lapsley, "NSP Shows promise on the pentium and PowerPC," *Microprocessor Reports*, 8, pp. 11-15, 1995.

291. "The visual instruction set: on chip support for new-media processing," *Sun Microsystems Whitepaper 95-022*: http://www.sun.com/sparc/vis/

292. R.M. Haralick, "Interactive image processing software," *Proceedings of the NATO Advanced Study Institute on Digital Image Processing and Analysis*, Bonas, France, June 1976.

293. M.S. Landy, Y. Cohen, and G. Sperling, "HIPS: A Unix-based image processing system," *Computer Vision, Graphics and Image Processing*, 25, pp. 331-47, 1984.

294. L.H. Jamieson, S.E. Hambrusch, A.A. Khokhar, and E.J. Delp, "The role of models, software tools, and applications in high performance computing," *in Developing a Computer Science Agenda for High Performance Computing*, edited by Uzi Vishkin. ACM Press, pp. 90-97, 1994.

295. Rasure and Kubica, "The Khoros application development environment," *Experimental Environments for Computer Vision and Image Processing*, editor H.I Christensen and J.L Crowley, World Scientific, 1994.

296. Rasure and Young, "An open environment for image processing software development," *1992 SPIE IS&T Symposium on Electronic Imaging*, SPIE Proceedings 1659, 14, 1992.

297. MPEG-4 Overview, ISO/IEC JTC1/SC29/WG11 N1730, Rob Koenen, ed., 1997

298. G. Bouton, B. Bouton, and G. Kubicek, *Inside Adobe Photoshop 4*, New Riders Publishing, 1997.

299. MATLAB Image Processing Toolbox: http://www.mathworks.com/products/image/

300. S. Wolfram, *The Mathematica Book*, Third Edition Wolfram Media, Inc. and Cambridge University Press, 1996.

301. National Instruments LabView: http://www.natinst.com/

302. R.A. Lerner, "The image understanding environment: progress since IUW'96," *Proceedings of the 1997 DARPA Image Understanding Workshop*, New Orleans, LA, 1997. (also see: http://www.aai.com/AAI/IUE/IUE.html).

303. NIH Image: http://rsb.info.nih.gov/nih-image/

304. R.A. Crowther, D.J. DeRosier, and A. Klug, "The reconstruction of three-dimensional structure from projections and its applications to electron microscopy," *Proc. Roy. Soc. London A*, 317, pp. 319-340, 1970.

305. A. Macovski, *Medical Imaging Systems*, Prentice Hall, Englewood Cliffs, NJ. 1983.

306. D. A. Ausherman, A. Kozma, J.L. Walker, H.M. Jones, and E.C. Poggio, "Developments in radar imaging," *IEEE Trans. Aerospace and Electronic Systems*, 20, pp. 363-400, 1984.

307. R.N. Bracewell, "Strip integration in radio astronomy," *Aust. J. Phys.*, 9, pp. 198-217, 1965.

308. A.R. Thompson, J.M. Moran, and G.W. Swenson, Jr., *Interferometry and Synthesis in Radio Astronomy*, John Wiley and Sons, New York, NY, 1986.

309. R.E. Sabbagha, *Diagnostic Ultrasound Applied to Obstetrics and Gynecology, 3rd ed.*, J.B. Lippincott, Philadelphia, PA, 1994.

310. B. Cornuelle et al., "Tomographic maps of the ocean meoscale. Part 1: Pure acoustics," *J. Phy. Ocean*, 15, pp. 133-152, 1985.

311. A. Dziewonski and J. Woodhouse, "Global images of the earth's interior," Science, 236, pp. 37-48, 1987.
312. E. A. Robinson and S. Treitel, Geophysical Signal Analysis, Prentice Hall, Englewood Cliffs, NJ, 1980.
313. J. Radon, "On the determination of functions from their integral values along certain manifolds," first published 1917 in German, translated and reprinted in IEEE Trans. Med. Imag., 5, pp. 170-176, 1986.
314. G. N. Hounsfield, "A method of and apparatus for examination of a body by radiation such as x or gamma radiation," Patent Specification 1283915, The U.K. Patent Office, England, 1972.
315. H.H. Barrett and W. Swindell, Radiological Imaging, Vol. 1 and 2, Academic Press, New York, NY, 1981.
316. R. Mersereau and A. Oppenheim, "Digital reconstruction of multidimensional signals from their projections," Proc. IEEE, 62, pp. 1319-1338, 1974.
317. H. Stark et al., "An investigation of computerized tomography by direct Fourier inversion and optimum interpolation," IEEE Trans. Biomed. Eng., BME-28, pp. 496-505, 1981.
318. L. A. Shepp and B.F. Logan, "The Fourier reconstruction of a head section," IEEE Trans. Nucl. Sci., NS-21, pp. 21-33, 1974.
319. A. Kak and M. Slaney, Principles of Computerized Tomographic Imaging, IEEE Press, 1988.
320. M. Soumekh, "Band-limited reconstruction from unevenly spaced sampled data," IEEE Trans. Acoust., Speech, Signal Proc., 36, pp. 110-122, 1988.
321. H. Choi and D. C. Munson, Jr., "Direct-Fourier reconstruction in tomography and synthetic aperture radar," Int. J. Imaging Systems and Technology, 9, pp. 1-13, 1998.
322. G.T. Herman, A.V. Lakshminarayanan, and A. Naparstek, "Convolution reconstruction techniques for divergent beams," Comput. Biol. Med., 6, pp. 259-271, 1976.
323. G.T. Herman, Image Reconstruction from Projections: The Fundamentals of Computerized Tomography, Academic Press, New York, NY, 1980.
324. Y. Censor, "Finite series-expansion reconstruction methods," Proc. IEEE, 71, pp. 409-419, 1983.
325. S. Kaczmarz, "Angenäherte Auflösung von Systemen linearer Gleichungen," Bull. Int. Acad. Pol. Sci. Lett. A, 6, pp. 355-357, 1937.
326. S. Webb, ed., The Physics of Medical Imaging, Institute of Physics Publishing, [illegible], 1988.
327. L.A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," IEEE Trans. Med. Imag., 1, pp. 113-122, 1982.
328. S. Geman and D.E. McClure, "Bayesian image analysis: An application to single photon emission tomography," Proc. Statistical Computing Section, Amer. Statistical Assoc., pp. 12-18, 1985.
329. J.A. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," IEEE Trans. Med. Imag., 13, pp. 290-300, 1994.
330. [illegible], K. Sauer, and Z. [illegible], "[illegible] methods for Bayesian reconstruction [illegible]," [illegible], 12, pp. [illegible]-701, 1993.
331. G. Gindi, M. Lee, A. Rangarajan, and I.G. Zubal, "Bayesian reconstruction of functional images using anatomical information as priors," IEEE Trans. Med. Imag., 12, pp. 670-680, 1993.
332. J. Ollinger and J.A. Fessler, "Positron-emission tomography," IEEE Signal Processing Magazine, 14, pp. 43-55, 1997.
333. P.E. Kinahan and J.G. Rogers, "Analytic 3D image reconstruction using all detected events," IEEE Trans. Nucl. Sci., 36, pp. 964-968, 1989.
334. M. Defrise, P. Kinahan, D. Townsend, C. Michel, M. Sibomana, and D. Newport, "Exact and approximate rebinning algorithms for 3D PET data," IEEE Trans. Med. Imag., 16, pp. 145-158, 1997.
335. C. [illegible], R. Clack, [illegible], M. Defrise, [illegible], "[illegible] system for the 3D [illegible] PET [illegible]," IEEE Trans. Nucl. Sci., 42, pp. 1023-[illegible], 1995.
336. P. Lauterbur, "Image formation by induced local interactions: examples employing nuclear magnetic resonance," Nature, 242, pp. 190-191, 1973.
337. A. Kumar, D. Welti, and R. Ernst, "NMR Fourier zeugmatography," J. Magn. Reson., 18, pp. 69-83, 1975.
338. C. Ahn, J. Kim, and Z. Cho, "High speed spiral-scan echo planar NMR imaging," IEEE Trans. Med. Imag., MI-5, pp. 2-7, 1986.
339. [illegible], "[illegible] MR [illegible]," IEEE Trans. [illegible], pp. [illegible], 1995.
340. [illegible], "[illegible]," [illegible], 1991.
341. [illegible] and R. A. [illegible], [illegible], Plenum Press, New York, NY, 1987.
342. J. [illegible], [illegible], John Wiley and Sons, New York, NY, 1989.
343. A. [illegible], "[illegible]," [illegible], 1993.
344. K. [illegible] and R. [illegible], "[illegible]," [illegible], 1993.
345. C. Cafforio, C. Prati, and E. Rocca, "SAR data focusing using seismic migration techniques," IEEE Trans. Aerosp. Electron. Syst., 27, pp. 194-207, 1991.
346. M. Soumekh, "A system model and inversion for synthetic aperture radar imaging," IEEE Trans. Image Processing, 1, pp. 64-76, 1992.
347. J. Walker, "Range-Doppler imaging of rotating objects," IEEE Trans. Aerosp. Electron. Syst., AES-16, pp. 23-52, 1980.
348. D.C. Munson, Jr., [illegible], "[illegible]," Proc. IEEE, 71, pp. [illegible], 1983.
349. D.R. Wehner, High Resolution Radar, Artech House, Norwood, MA, 1994.
350. J.L.H. Webb, D.C. Munson, Jr., and N.J.S. Stacy, "High-resolution planetary imaging via spotlight-mode synthetic aperture radar," IEEE Trans. Image Proc., [illegible], 1998.
351. J.C. Curlander and R.N. McDonough, Synthetic Aperture Radar: Systems and Signal Processing, John Wiley and Sons, New York, NY, 1991.
352. W.C. Carrara, R.S. Goodman, and R.M. Majewski, Spotlight Synthetic Aperture Radar: Signal Processing Algorithms, Artech House, Norwood, MA, 1995.
353. C.V. Jakowatz, Jr., [illegible], Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach, Kluwer Academic Publishers, Boston, MA, 1996.
354. R.N. Bracewell, Two-Dimensional Imaging, Prentice Hall, Englewood Cliffs, NJ, 1995.
355. M. Soumekh, Fourier Array Imaging, Prentice Hall, Englewood Cliffs, NJ, 1994.
356. R.E. Blahut, Theory of Remote Surveillance, ECE 458 Course Notes, University of Illinois, Urbana, IL, 1993.