Special Section on Human Vision and Electronic Imaging

Bernice E. Rogowitz
IBM T. J. Watson Research Center
Yorktown Heights, New York 10598-0218

Thrasyvoulos N. Pappas
Northwestern University
Dept. of Electrical and Computer Engineering
Evanston, Illinois 60208-3118

Jan Allebach
Purdue University
School of Electrical and Computer Engineering
West Lafayette, Indiana 47907-1285

This collection of papers commemorates the tenth anniversary of the IS&T/SPIE Conference on Human Vision and Electronic Imaging. These papers represent major trends in the conference and demonstrate the interplay between real-world imaging applications and vision research.

In the first paper (Rogowitz, Pappas, and Allebach), we present an overview of this field and provide a context for the papers that follow. We use the metaphor of a food chain to draw attention to the many levels of human visual processing and how they have influenced different imaging applications. In this scheme, low level vision is concerned with the detection and recognition of visual patterns by simple, mainly linear mechanisms, mediated by retinal or striate cortex filters. Moving up the food chain, more complex, often non-linear processes are posited to model the perception of more complex images and objects, and explore higher-level cortical functions such as visual attention and pattern recognition. At the top of the food chain are those visual tasks which involve judgments about very rich environments, aesthetic judgments and emotional responses. In this view, the role of the human observer in imaging systems cannot be modeled as a simple function, but instead, is a complex set of visual, perceptual, and cognitive behaviors. Understanding these behaviors, and applying them appropriately,

is the focus of the Conference on Human Vision and Electronic Imaging and of this special section.

One of the important issues in the design of imaging systems is the most efficient use of the available transmission bandwidth and storage capacity, while preserving the best image quality. The next three papers demonstrate the influence of low-level vision models on the quantitative evaluation of image quality and on the development of image compression techniques. These approaches are based on detectability and evaluation of image artifacts. Watson, Hu, and McGowan's paper derives an objective metric for the evaluation of video image quality, especially for video compression. It makes use of spatiotemporal models of human perception. Daly, Matthews, and Ribas-Corbera use models of foveal eccentricity to allocate most of the limited transmission bandwidth to the face region for video conferencing applications. In the next paper, de Ridder examines the influence of judgment strategies, and in particular the composition of the stimulus set and the instructions, on the outcome of subjective experiments for the evaluation of image quality.

With the next three papers, we consider more complex visual images, more complex visual tasks, and higher level approaches to modeling their perception. An important new approach in vision research has been to under-

stand the relationship between the statistics of natural images and the neural mechanisms that have evolved to process them. In their paper, Zetzsche and Krieger argue that the way the human visual system exploits statistical redundancies in natural images is highly non-linear and involves higher-order statistics. The authors model the perception of complex visual images and explore the implications of their findings for imaging technology. The MacLin and Webster paper provides a vivid example of how exquisitely complex human perception really is. They use a standard experimental paradigm in psychophysics, the adaptation experiment, to demonstrate the ability of the human visual system to adapt to very complex spatial perturbations. Using faces, they find that when observers spend time viewing spatial distortions in a face (e.g., widened distance between the eyes), they perceive non-distorted faces as having been distorted in the opposite direction (e.g., narrowing distance between the eyes). This suggests the existence of very high-level image processing mechanisms, or sets of mechanisms, in human vision, whose response(s) can be weakened through this adaptation process. In exploring the role of higher-level, non-linear, and more sophisticated mechanisms in human vision, it is always important to ask whether complex behaviors might not be sufficiently modeled by lower-level, less

elaborated mechanisms. In his paper, McCann considers the interplay between bottom up ("Early Vision") and top down ("High Vision") mechanisms in explaining visual illusions involving black-and-white geometric patterns. He shows families of cases where low-level mechanisms can model phenomena commonly attributed to higher-level operations.

The next three papers focus on the role of top-down and bottom-up processes in visual attention. Predicting how human observers place their visual fixations can be very important for imaging applications, such as compression, analysis, and understanding. Stark *et al.* consider the fundamental mechanisms which direct the way the high resolution fovea moves from fixation to fixation as it scans the visual scene. They review scanpath theory, which describes the way top-down processes control active eye movements, and extend it to dynamic scenes. In this paper, the authors describe a quantitative metric that measures the similarity of strings of fixations, and test it in a computer vision simulation. To predict the sequence of eye-movements, Schill *et al.* develop a system which combines a top-down knowledge-based reasoning system with low-level visual operators. In this system, each next fixation in a sequence is the one which provides the most information gain. Itti and Koch provide a method for modeling how the response of these low-level operators is combined to identify high-information content regions in an image. They develop a saliency map which combines the contributions of various bottom-up processes, such as color and orientation, and evaluate various combination strategies.

In the final three papers, we consider imaging applications which drive the development of higher-level visual models. In digital libraries, the goal is to help users retrieve images from an archive. This is a difficult problem since images can vary along many dimensions, and can be retrieved by a number of different criteria. In Papathomas *et al.*, the authors use psychophysical experiments to test hypotheses about similarity judgments, presentation order, learning, and the role of pictorial vs. semantic image features, and to

quantify the performance of content-based image retrieval algorithms. In 3-D imaging applications, it is important to render objects and scenes so that they appear to be realistic. This requires an understanding of shape perception. Using 3-D graphics, Browse, Rodger, and Adderley examine how different factors such as shading, lighting direction, surface markings, and specular highlights influence the perception of simple convex objects. The needs of artists and product designers also drive the development of visual models. In the final paper of this collection, Bender proposes color design tools that are based on models of color perception and color harmony, and demonstrates experimentally their value for fashion designers.

As we close this brief introduction to the special section, the question is: What's next? The scope of the field of human vision and electronic imaging has been growing with the evolution of electronic imaging technology. Even though the progress of the last decade has been impressive, we believe that we are still at the beginning of this new discipline. As we move into the new millennium, we expect an ever expanding number of new applications driven by the two-way interaction between advances in human perception and electronic media.

Bernice E. Rogowitz is an experimental psychologist specializing in human vision and its applications in imaging systems. She earned her PhD at Columbia University, and completed an NIH postdoctoral fellowship in the Laboratory of Psychophysics at Harvard University. She currently manages the Visual Analysis Group at the IBM T. J. Watson Research Center. The goal of this group is to develop perceptually-based, intelligent, interactive systems for manipulating, synthesizing and understanding data. In 1988, Dr. Rogowitz founded the IS&T/SPIE Conference on Human Vision and Electronic Imaging, which she continues to co-chair. She has served on the board of IS&T since 1997 and was elected an IS&T fellow in 2000.

Thrasyvoulos N. Pappas received the SB, SM, and PhD degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, in 1979, 1982, and 1987, respectively. In 1999 he joined the Department of Electrical and Computer Engineering at Northwestern University as an associate professor. From 1987 until 1999, he was a member of the technical staff at Bell Laboratories, Murray Hill, NJ. His research interests are in image and multidimensional signal processing. His recent work has been on perceptual image coding, joint source/channel coding for lossy channels, model-based halftoning, color printing, image segmentation, and video/audio integration for teleconferencing. Dr. Pappas is co-chair of the SPIE/IS&T annual conference on Human Vision and Electronic Imaging. He is technical program co-chair of the 2001 IEEE International Conference on Image Processing to be held in Thessaloniki, Greece. He is the electronic abstracts editor and an associate editor for the *IEEE Transactions on Image Processing.* He is also vice-chair of the IEEE Signal Processing Society's Image and Multidimensional Signal Processing Technical Committee and a member of the Multimedia Signal Processing Technical Committee.

Jan P. Allebach received his BSEE from the University of Delaware in 1972 and his PhD from Princeton University in 1976. He was on the faculty at the University of Delaware from 1976 to 1983. Since 1983, he has been at Purdue University in the School of Electrical and Computer Engineering. His current research interests include image rendering, image quality, and color imaging. Dr. Allebach is active in both the IEEE Signal Processing Society and IS&T. He is a Fellow of both societies, has served as Distinguished/Visiting Lecturer for both societies, and has served as an officer and on the Board of Directors of both societies. He received the Senior (best paper) Award from the IEEE Signal Processing Society and the Bowman Award from IS&T. He was co-chair of the conference on Human Vision and Electronic Imaging from 1990 through 1996.