Guest Editors' Introduction to the Special Section on Graphical Models in Computer Vision

James M. Rehg, Vladimir Pavlovic, Thomas S. Huang, *Fellow*, *IEEE*, and William T. Freeman

◆

THE last 10 years have witnessed rapid growth in the popularity of graphical models, most notably Bayesian networks, as a tool for representing, learning, and computing complex probability distributions. Graphical models provide an explicit representation of the statistical dependencies between the components of a complex probability model, effectively marrying probability theory and graph theory. As Jordan puts it in [2], graphical models are "a natural tool for dealing with two problems that occur throughout applied mathematics and engineering—uncertainty and complexity—and, in particular, they are playing an increasingly important role in the design and analysis of machine learning algorithms."

Graphical models provide powerful computational support for the Bayesian approach to computer vision, which has become a standard framework for addressing vision problems. Many familiar tools from the vision literature, such as Markov random fields, hidden Markov models, and the Kalman filter, are instances of graphical models. More importantly, the graphical models formalism makes it possible to generalize these tools and develop novel statistical representations and associated algorithms for inference and learning.

The history of graphical models in computer vision follows closely that of graphical models in general. Research by Pearl [3] and Lauritzen [4] in the late 1980s played a seminal role in introducing this formalism to areas of AI and statistical learning. Not long after, the formalism spread to fields such as statistics, systems engineering, information theory, pattern recognition, and, among others, computer vision. One of the earliest occurrences of graphical models in the vision literature was a paper by Binford et al. [1]. The paper described the use of Bayesian inference in a hierarchical probability model to match 3D object models to groupings of curves in a single image. The following year marked the publication of Pearl's influential book [3] on graphical models. Since then, many technical papers have been published in IEEE journals and conference proceedings that address different aspects and applications of graphical models in computer vision.

Our goal in organizing this special section was to demonstrate the breadth of applicability of the graphical models formalism to vision problems. Our call for papers in February 2002 produced 16 submissions. After a careful review process, we selected six papers for publication, including five regular papers, and one short paper. These papers reflect the state-of-the-art in the use of graphical models in vision problems that range from low-level image understanding to high-level scene interpretation. We believe these papers will appeal both to vision researchers who are actively engaged in the use of graphical models and machine learning researchers looking for a challenging application domain.

The first paper in this section is "Stereo Matching Using Belief Propagation" by J. Sun, N.-N. Zheng, and H.-Y. Shum. The authors describe a new stereo algorithm based on loopy belief propagation, a powerful inference technique for complex graphical models in which exact inference is intractable. They formulate the dense stereo matching problem as MAP estimation on coupled Markov random fields and obtain promising results on standard test data sets. One of the benefits of this formulation, as the authors demonstrate, is the ease with which it can be extended to handle multiview stereo matching.

In their paper "Statistical Cue Integration of DAG Deformable Models" S.K. Goldenstein, C. Vogler, and D. Metaxas describe a scheme for combining different sources of information into estimates of the parameters of a deformable model. They use a DAG representation of the interdependencies between the nodes in a deformable model. This framework supports the efficient integration of information from edges and other cues using the machinery of affine arithmetic and the propagation of uncertainties. They present experimental results for a face tracking application.

Y. Song, L. Goncalves, and P. Perona describe, in their paper "Unsupervised Learning of Human Motion," a method for learning probabilistic models of human motion from video sequences in cluttered scenes. Two key advantages of their method are its unsupervised nature, which can mitigate the need for tedious hand labeling of data, and the utilization of graphical model constraints to reduce the search space when fitting a human figure model.

- *J.M. Rehg is with the Georgia Institute of Technology, 801 Atlantic Drive, Atlanta, GA 30332-0280. E-mail: rehg@cc.gatech.edu.*
- *V. Pavlovic is with the Department of Computer Science, Rutgers University, 110 Frelinghuysen Road, Piscataway, NJ 08854-8019. E-mail: vladimir@cs.rutgers.edu.*
- *T.S. Huang is with the Department Information Department of Electrical & Computer Engineering, 2039 Beckman Institute for Advanced Science and Technology, 405 North Mathews, Urbana, IL 61801. E-mail: huang@ifp.uiuc.edu.*
- *W.T. Freeman is with the Massachusetts Institute of Technology (MIT) Artificial Intelligence Laboratory, 200 (545) Technology Square, MIT Building NE43, Cambridge, MA 02139. E-mail: billf@ai.mit.edu*

Published by the IEEE Computer Society

M.J. Beal, N. Jojic, and H. Attias present a graphical model approach to the multimodal problem of audiovisual tracking in "A Graphical Model for Audiovisual Object Tracking." Their framework exploits the composability of graphical models in order to combine audio and video cues at the signal level. One benefit of their approach is automatic calibration of the sensors through parameter learning. Their paper reflects a growing trend in the vision community of fusing information from multiple modalities in order to tackle challenging sensing problems such as robust object tracking.

The last regular paper is an invited submission by T.O. Binford and T.S. Levitt entitled "Evidential Reasoning for Object Recognition." This paper summarizes the authors' early pioneering work in applying graphical models to 3D object recognition from a single image. It contains a great deal of material of historical interest from the late 1980s that has never before appeared in archival form.

The final paper of this special section is by M. Marengoni, A. Hanson, S. Zilberstein, and E. Riseman. "Decision Making and Uncertainty Management in a 3D Reconstruction System" describes an application of Bayesian networks coupled with the utility theory to the problem of aerial image interpretation. This framework supports optimal decision making, exemplified here as the selection of the most important pieces of visual evidence during classification.

The papers in this special section demonstrate the wide applicability of graphical modeling techniques to problems in computer vision. The graphical models formalism has the potential to unify many seemingly disparate approaches within the vision literature by placing them in a common framework. It also provides a connection to a rich literature on learning, inference, and modeling techniques. We look forward to continued rapprochement between the computer vision and machine learning communities and the growth of graphical modeling techniques within the vision literature.

We would like to extend our thanks to the authors and the reviewers for their efforts on behalf of the special section. We also thank the staff of *TPAMI* and, especially, Hilda Hosillos for her help and guidance throughout the publication process.

James M. Rehg
Vladimir Pavlovic
Thomas S. Huang
William T. Freeman
Guest Editors

REFERENCES

[1] T.O. Binford, T.S. Levitt, and W.B. Mann, "Bayesian Inference in Model-Based Machine Vision," *Proc. Int'l Conf. Uncertainty in Artificial Intelligence,* pp. 73-96, 1987.
[2] M. Jordan, *Learning in Graphical Models.* MIT Press, 1998.
[3] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, 1988.
[4] S.L.L. Lauritzen, *Graphical Models.* Oxford: Clarendon Press, 1996.

James M. Rehg received the PhD degree from Carnegie Mellon University in 1995. From 1996 to 2001, he led the computer vision research group at the Cambridge Research Laboratory of the Digital Equipment Corporation, which was acquired by Compaq Computer Corporation in 1998. In 2001, he joined the faculty of the College of Computing at the Georgia Institute of Technology, where he is currently an associate professor. His research interests include computer vision, machine learning, human-computer interaction, computer graphics, and distributed computing.

Vladimir Pavlovic received the PhD degree in electrical engineering from the University of Illinois at Urbana-Champaign in 1999. He is an assistant professor in the Computer Science Department at Rutgers University and an adjunct assistant professor in the Bioinformatics Program at Boston University. From 1999 to 2001, he was a member of the research staff at the Cambridge Research Laboratory, Cambridge, Massachusetts. His research interests include statistical modeling of time-series, statistical computer vision, machine learning, and bioinformatics.

Thomas S. Huang received the BS degree in electrical engineering from the National Taiwan University, Taipei, Taiwan, China, the MS, and ScD degrees in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge. From 1963 to 1973, he was on the faculty of the Department of Electrical Engineering at MIT. From 1973 to 1980, he was on the faculty of the School of Electrical Engineering and director of its Laboratory for Information and Signal Processing at Purdue University. In 1980, he joined the University of Illinois at Urbana-Champaign, where he is now William L. Everitt Distinguished Professor of Electrical and Computer Engineering, a research professor at the Coordinated Science Laboratory, the head of the Image Formation and Processing Group at the Beckman Institute for Advanced Science and Technology, and a cochair of the Institute's major research theme-Human Computer Intelligent Interaction. He has published 14 books and more than 500 papers in network theory, digital filtering, image processing, and computer vision. He is a member of the National Academy of Engineering, a foreign member of the Chinese Academies of Engineering and Sciences, and a fellow of the International Association of Pattern Recognition, IEEE, and the Optical Society of American. He has received a Guggenheim Fellowship. He was awarded the IEEE Third Millennium Medal in 2000. Also in 2000, he received the Honda Lifetime Achievement Award for "contributions to motion analysis." In 2001, he received the IEEE Jack S. Kilby Medal. In 2002, he received the King-Sun Fu Prize, International Association of Pattern Recognition; and the Pan Wen-Yuan Outstanding Research Award.

William T. Freeman studied computer vision for the PhD degree in 1992 from the Massachussetts Institute of Technology (MIT). He is an associate professor of electrical engineering and computer science at the Artificial Intelligence Laboratory at MIT, joining the faculty in 2001. From 1992 to 2001, he worked at Mitsubishi Electric Research Labs (MERL), in Cambridge, Massachusetts, most recently, as a senior research scientist and associate director. His current research interests include machine learning applied to computer vision, Bayesian models of visual perception, and interactive applications of computer vision. In 1997, he received the Outstanding Paper prize at the Conference on Computer Vision and Pattern Recognition for work on applying bilinear models to "separating style and content." Previous research topics include steerable filters and pyramids, the generic viewpoint assumption, color constancy, and computer vision for computer games.