Guest Editors' Introduction to the Special Section on Probabilistic Graphical Models

Qiang Ji, *Senior Member*, *IEEE*, Jiebo Luo, *Fellow*, *IEEE*, Dimitris Metaxas, *Member*, *IEEE*, Antonio Torralba, *Member*, *IEEE*, Thomas S. Huang, *Fellow*, *IEEE*, and Erik B. Sudderth, *Member*, *IEEE*

1 Introduction

An exciting development over the last decade has been the gradually widespread adoption of probabilistic graphical models (PGMs) in many areas of computer vision and pattern recognition. Representing an integration of graph and probability theories, a number of families of graphical models have been studied in the statistics and machine learning literatures. Examples of directed graphical models include Bayesian networks (BNs) and hidden Markov models (HMMs). Alternatively, undirected graphical models such as Markov random fields (MRFs) and conditional random fields (CRFs) are widely used to model spatial dependencies. Graphical models lead to powerful computational methods for representation, learning, and inference in computer vision. In particular, PGMs provide a unified framework for representing observations and domain-specific contextual knowledge, and for performing recognition and classification through rigorous probabilistic inference.

The history of PGMs in computer vision closely follows that of graphical models in general. Research in the late 1980s by Judea Pearl and Steffen Lauritzen, among others, played a seminal role in introducing this formalism to areas of AI and statistics. Not long after, the formalism spread to other fields, including systems engineering, information theory, pattern recognition, and computer vision. One of the earliest applications of graphical models in computer vision was work by Binford et al. [1], in which Bayesian inference in a hierarchical probability model was used to match 3D object models to groupings of curves in a single image. The following year marked the publication of Pearl's influential book *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* [2].

Since then, hundreds of technical papers have been published that address different aspects and continuing applications of PGMs in computer vision. Many classical tools from the vision literature, including MRFs, HMMs, and the Kalman Filter, are unified and generalized by PGMs. For example, MRFs have become a widely used, standard framework for both image segmentation and stereo reconstruction. More recently, discriminatively trained CRFs have increasingly been used to incorporate rich, nonlocal features in image labeling tasks. For motion analysis, event modeling, and activity recognition, HMMs have become the standard tool. Markov networks have similarly led to progress in multitarget tracking, where the relationships of multiple objects can be intuitively captured via probabilistic graphs.

Variants of HMMs (such as hierarchical HMMs and coupled HMMs) have recently been introduced to model nonlocal dependencies and complex dynamic events. In particular, dynamic Bayesian networks (DBNs) are increasingly used to capture complex spatio-temporal patterns for object tracking and for modeling and recognizing complex dynamic activities such as human behaviors and facial expressions. More generally, PGMs excel as a natural framework for integrating top-down and bottom-up visual processes. Image segmentation, object detection, and object recognition have been shown to not only benefit from one another, but also be computationally feasible due to efficient PGM inference and learning algorithms.

We believe the powerful capability of graphical models to encode and integrate information at different levels, from complex high-level domain knowledge to low-level image properties, will continue to produce extremely effective models of visual data. In spite of recent successes, however, PGMs' use in computer vision still has tremendous room to expand in scope, depth, and rigor. The main objective of this special section is to review the latest developments in applications of PGMs to various computer vision tasks, identify important modeling and computational issues, and point out future research directions.

- *Q. Ji is with the Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180. E-mail: qji@ecse.rpi.edu.*
- *J. Luo is with Kodak Research Laboratories, 1999 Lake Avenue, Rochester, NY 14650. E-mail: jiebo.luo@kodak.com.*
- *D. Metaxas is with the Department of Computer Science, Rutgers University, Piscataway, NJ 08854. E-mail: dnm@cs.rutgers.edu.*
- *A. Torralba is with the Computer Science and Artificial Intelligence Laboratory, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139. E-mail: torralba@csail.mit.edu.*
- *T.S. Huang is with the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, 405 N. Mathews Ave., Urbana, IL 61801. E-mail: huang@ifp.uiuc.edu.*
- *E.B. Sudderth is with the Department of Computer Science, 115 Waterman Street, Brown University, Box 1910, Providence, RI 02912. E-mail: sudderth@cs.brown.edu.*

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org.

2 The Review Process

For this special section, we solicited original contributions describing applications of PGMs in all areas of computer vision, including (but not limited to):

1. image and video modeling,
2. image and video segmentation,
3. object detection,
4. object and scene recognition,
5. high-level event and activity understanding,
6. motion estimation and tracking,
7. new inference and learning (both structure and parameters) theories for graphical models arising in vision applications,
8. generative and discriminative models, and
9. models incorporating contextual, domain, or common sense knowledge

We were, in particular, interested in contributions that demonstrate innovative applications to computer vision tasks and that result in significant performance improvements over alternative methods. We also sought to identify new theoretical developments in PGM learning and inference algorithms, as driven by computer vision applications. To reach a wide audience, we published the special section's call for papers through a number of channels, including mailing lists, posters at major computer vision conferences, and advertisement in journals. The response was overwhelming: We received a total of 68 submissions from all over the world, covering almost every aspect of computer vision from shape matching, feature extraction, superresolution, and tracking to recognition, 3D reconstruction, and activity recognition. The theoretical contributions are also diverse, ranging from innovative new models tailored to particular applications, to new algorithms for inference and learning.

To ensure a careful and fair review process, we recruited more than 200 expert reviewers from the computer vision and machine learning communities. Each paper received at least three reviews, with additional reviews and comments by the guest editors. Some papers underwent two or three rounds of reviews before a final decision was made. After much discussion and deliberation, and accounting for the special section's tight publication schedule, we accepted the 10 outstanding papers contained in this section. We are very impressed by the diversity and quality of these contributions and hope this special section will have a long lasting positive effect on the community.

3 Accepted Papers

The 10 accepted papers can be coarsely categorized based on their specific computer vision application domains: four explore frameworks for object or activity recognition, two describe methods for detecting objects or visual patterns, and four consider visual tracking or motion analysis. In the paragraphs below, we briefly summarize the papers in each category.

3.1 Object and Activity Recognition

Scene text recognition deals with the problem of recognizing text in a natural environment, such as on traffic signs and store fronts. In contrast to standard document analysis, this problem is complicated by extreme font variability, uncontrolled viewing conditions, and minimal language context. In their paper titled "Scene Text Recognition Using Similarity and a Lexicon with Sparse Belief Propagation," Weinman, Learned-Miller, and Hanson propose a solution to this problem, where a discriminative model is developed that combines character appearance, character bigrams, character similarity, and a lexicon. Using an efficient sparse belief propagation algorithm, inference is simultaneously performed on factor graphs built from various cues, avoiding the error accumulation that often occurs with sequential processing.

The paper titled "Unsupervised Learning of Probabilistic Object Models (POMs) for Object Classification, Segmentation, and Recognition Using Knowledge," by Chen, Zhu, Yuille, and Zhang, addresses the difficult problem of learning probabilistic object models with minimal supervision for object classification, segmentation, and recognition. The authors formulate the task of learning POMs which model object categories as a structure induction problem. Then, they use the strategy of *knowledge propagation* to enable POMs to provide information to other POMs. Extensive experimental results, including analysis of robustness and comparisons to state-of-the-art methods, demonstrate the power of this formulation.

Collections of discriminative local features, when combined with models originally used to learn topics underlying document collections, have recently shown promise for modeling visual object categories. In their paper "Human Action Recognition by Semilatent Topic Models," Yang and Mori show that related approaches also provide effective models of human actions in video. Mapping quantized motion descriptors to "visual words" and topics to action classes, they use variational inference methods to accurately cluster test video frames into semantic activities.

The paper titled "Observing Human-Object Interactions: Using Spatial and Functional Compatibility for Recognition" by Gupta, Kembhavi, and Davis, introduces an integrated system for recognizing actions and objects. By solving these two recognition tasks in an integrated system, the authors show that the system can recognize objects and actions even when appearances are not discriminative enough. For instance, when only analyzing hand trajectories, differentiating among actions might be hard, but discrimination can become easier by integrating information about the objects involved in the action. Likewise, recognition of similar objects might be easier if they are involved on easy to discriminate actions.

3.2 Object Detection

In the paper titled "A Probabilistic Framework for 3D Visual Object Representation," authors Detry, Pugeault, and Piater tackle the difficult task of detection and localization of objects within highly cluttered scenes by developing a generative 3D visual representation. They encode the 3D geometry and visual appearance of an object into a part-based model, and develop mechanisms for

autonomous learning and probabilistic inference in that model. A Markov network is used to combine local appearance and 3D spatial relationships through a hierarchy of increasingly expressive features. The authors demonstrate efficiency and robustness to input noise, viewpoint changes, and occlusions.

Park, Brocklehurst, Collins, and Liu, in their paper "Deformed Lattice Detection in Real-World Images Using Mean-Shift Belief Propagation," address the issue of detecting repetitive patterns within deformed 2D lattices. The lattice detection problem is formulated as a spatial, multitarget tracking problem, solved within the MRF framework using a novel mean-shift belief propagation method. The model is evaluated in a wide range of images that contain near-regular textures (such as windows in buildings, fruits, wiry structures, etc.)

3.3 Tracking and Motion Analysis

An elastic motion is a nonrigid motion constrained by some degrees of smoothness and continuity. In their paper titled "A Mixture of Transformed Hidden Markov Models for Elastic Motion Estimation," authors Di, Tao, and Xu address the problem of tracking deformable objects, such as faces and human figures, by assuming an elastic deformation. Their method extracts edge points from each frame and formulates motion estimation as a shape registration task that also incorporates temporal consistency constraints. A mixture of transformed hidden Markov models is used to account for both spatial smoothness and temporal continuity. The end result is a more globally coherent interpretation of elastic motion from local edge features, which are often subject to ambiguities, incompleteness, and clutter.

Tracking multiple people from a mobile platform is challenging due to camera motion, occlusion, and dynamic backgrounds. In their paper titled "Robust Multiperson Tracking from a Mobile Platform," authors Ess, Leibe, Schindler, and van Gool propose an integrated approach which uses a graphical model to systematically capture dependencies between different components (camera position, depth, object detection, tracking) of a multiperson tracking system. With such a model, they can perform joint camera calibration, 3D reconstruction, object detection, and tracking in a principled manner. The evaluation of the system on busy inner city streets demonstrates its robust performance under realistic scenarios.

Linear dynamical systems are widely used in the analysis of video sequences and motion data. In the paper "Discriminative Learning for Dynamic State Prediction," Kim and Pavlovic propose a pair of novel, conditional models for continuous time series. These discriminative objective functions lead to efficient algorithms for selecting and incorporating sophisticated, potentially nonlocal features. In experiments, these approaches lead to more accurate predictions of articulated motion than standard generative models.

So-called dynamic texture models describe repetitive motion patterns via the evolution of a latent dynamical system. In their paper, Chan and Vasconcelos have proposed a family of "Layered Dynamic Textures" which allow videos to be segmented into multiple, independently evolving motions. Spatial continuity of segments is encouraged via a Markov random field prior and inference performed via either variational or Monte Carlo methods. Their approach is validated on a range of synthetic and real-world video sequences.

Acknowledgments

We would like to thank many people for their contributions to this special section, which would not have been possible without their efforts. First and foremost, we thank the authors for their enthusiastic response and innovative contributions. Second, we would like to express our great appreciation to the more than 200 reviewers for their timely, detailed, and thoughtful reviews. Their work was essential to ensure the quality of this special section. Third, we would are thankful for the support we have received from the TPAMI editorial office, including past Editor-in-Chief Professor David Kriegman, current Editor-in-Chief Professor Ramin Zabih, and editorial assistant Elaine Stephenson. In particular, we thank Elaine for her timely answers to our inquiries, and for actively keeping the special section on schedule.

Qiang Ji
Jiebo Luo
Dimitris Metaxas
Antonio Torralba
Thomas S. Huang
Erik B. Sudderth
Guest Editors

References

[1] T. Binford, T. Levitt, and W. Mann, "Bayesian Inference in Model Based Machine Vision," *Proc. Third Ann. Conf. Uncertainty in Artificial Intelligence*, pp. 73-96, 1987.
[2] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

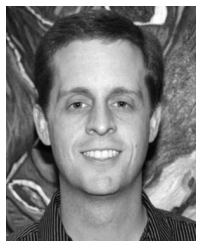
Qiang Ji received the PhD degree in electrical engineering from the University of Washington. He is currently a professor in the Department of Electrical, Computer, and Systems Engineering at Rensselaer Polytechnic Institute (RPI). He is also a program director at the US National Science Foundation (NSF), managing part of NSF's computer vision and machine learning programs. He has held teaching and research positions with the Beckman Institute at the University of Illinois at Urbana-Champaign, the Robotics Institute at Carnegie Mellon University, the Department of Computer Science at the University of Nevada at Reno, and the US Air Force Research Laboratory. He currently serves as the director of the Intelligent Systems Laboratory (ISL) at RPI. His research interests are in computer vision, pattern recognition, and probabilistic graphical models. He has published more than 150 papers in peer-reviewed journals and conferences. His research has been supported by major governmental agencies including NSF, NIH, DARPA, ONR, ARO, and AFOSR, as well as by major companies including Honda and Boeing. He is an editor on several computer vision and pattern recognition related journals and he has served as a program committee member, area chair, and program chair for numerous international conferences/workshops. Professor Ji is a senior member of the IEEE.

Jiebo Luo received the PhD degree from the University of Rochester in 1995. He is a senior principal scientist with the Kodak Research Laboratories, Rochester, New York. His research interests include image processing, machine learning, computer vision, multimedia data mining, and computational photography. He has authored more than 150 technical papers and holds 50 US patents. Dr. Luo has been involved in organizing numerous leading technical conferences sponsored by the IEEE, ACM, and SPIE, most recently being the general chair of the 2008 ACM International Conference on Content-based Image and Video Retrieval (CIVR), area chair of the 2008 IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), and program cochair of the 2007 SPIE International Symposium on Visual Communication and Image Processing (VCIP). He serves or has served on the editorial boards of the *IEEE Transactions on Pattern Analysis and Machine Intelligence* (*TPAMI*), the *IEEE Transactions on Multimedia* (*TMM*), *Pattern Recognition* (*PR*), and the *Journal of Electronic Imaging*. He has been a guest editor for a number of special issues, including "Real-World Image Annotation and Retrieval" (*TPAMI*, 2008) and "Integration of Content and Context for Multimedia Management" (*TMM*, 2009). He is a fellow of the IEEE and of SPIE.

Dimitris Metaxas received the PhD degree in 1992 from the University of Toronto and was a tenured faculty member at the University of Pennsylvania from 1992 to 2001. He is a Professor II (Distinguished) in the Computer Science Department at Rutgers University. He is currently directing the Center for Computational Biomedicine, Imaging, and Modeling (CBIM). Dr. Metaxas has been conducting research toward the development of formal methods upon which both computer vision, computer graphics, and medical imaging can advance synergistically, as well as on massive data analytics problems. In computer vision, he works on the simultaneous segmentation and fitting of complex objects, shape representation, deterministic and statistical object tracking, learning, ASL, and human activity recognition. In medical image analysis, he works on segmentation, registration, and classification methods for cardiac and cancer applications. In computer graphics, he is working on physics-based special effects methods for animation. He has pioneered the use of Navier-Stokes methods for fluid animations that were used in the movie *Antz* in 1998. He has published more than 300 research articles in these areas and has graduated 27 PhD students. His research has been funded by the NSF, NIH, ONR, AFOSR and ARO. He has published a book on his research activities titled *Physics-Based Deformable Models: Applications to Computer Vision, Graphics and Medical Imaging* (Kluwer Academic). He is on the editorial board of *Medical Imaging* and is an associate editor of *GMOD* and *CAD*. Dr. Metaxas has received several best paper awards for his work on in the above areas. He was awarded a Fulbright Fellowship in 1986, is a recipient of NSF Research Initiation and Career awards, an ONR YIP, and is a fellow of the American Institute of Medical and Biological Engineers and a member of the ACM and of the IEEE. He was also the program chair of ICCV '07 and the general chair of MICCAI '08.

Antonio Torralba received the degree in telecommunications engineering from the Universidad Politécnica de Cataluña, Spain, and was awarded the PhD degree in signal, image, and speech processing by the Institut National Polytechnique de Grenoble, France. He is an associate professor of electrical engineering and computer science in the Computer Science and Artificial Intelligence Laboratory (CSAIL) at the Massachusetts Institute of Technology (MIT). He spent postdoctoral training at the Brain and Cognitive Science Department and the Computer Science and Artificial Intelligence Laboratory at MIT. He is a member of the IEEE.

Thomas S. Huang received the ScD degree from the Massachusetts Institute of Technology (MIT) in electrical engineering and was on the faculty of MIT and Purdue University. He joined the University of Illinois at Urbana-Champaign in 1980 and is currently the William L. Everitt Distinguished Professor of Electrical and Computer Engineering, Research Professor of the Coordinated Science Laboratory, Professor of the Center for Advanced Study, and cochair of the Human Computer Intelligent Interaction major research theme of the Beckman Institute for Advanced Science and Technology. Professor Huang is a member of the National Academy of Engineering and has received numerous honors and awards, including the IEEE Jack S. Kilby Signal Processing Medal and the King-Sun Fu Prize of the International Association of Pattern Recognition. He has published 21 books and more than 600 technical papers in network theory, digital holograpy, image and video compression, multimodal human computer interfaces, and multimedia databases. He is a fellow of the IEEE.



Erik B. Sudderth received the bachelor's degree (summa cum laude) in electrical engineering from the University of California, San Diego, and the master's and PhD degrees in electrical engineering and computer science from the Massachusetts Institute of Technology. He is an assistant professor in the Brown University Department of Computer Science. From 2006 through 2009, he was a postdoctoral scholar at the University of California, Berkeley. His research interests include probabilistic graphical models, nonparametric Bayesian methods, and applications of statistical machine learning in object recognition, tracking, visual scene analysis, and image processing. He was awarded a National Defense Science and Engineering Graduate Fellowship (1999), an Intel Foundation Doctoral Fellowship (2004), and in 2008 was named one of "AI's 10 to Watch" by *IEEE Intelligent Systems*. He is a member of the IEEE.