Int J Comput Vis (2011) 93: 115–116
DOI 10.1007/s11263-011-0420-8

Editorial

Sinisa Todorovic · Rama Chellappa

Received: 3 January 2011 / Accepted: 6 January 2011 / Published online: 26 January 2011

Syntactic description of images took off in the mid sixties due to seminal works done by Dr. R. Narasimhan (Labeling schemata and syntactic descriptions of pictures, Information and Control, Vol. 7, pp. 151–179, June 1964), while at University of Illinois, Prof. K.S. Fu at Purdue University, and Prof. A. Rosenfeld at University of Maryland. Realizing that deterministic production rules characterizing the image patterns are too rigid, Prof. Fu initiated the development of stochastic image grammars in the late seventies, and designed various stochastic inference rules for syntactic pattern recognition.

When Prof. Fu passed away in 1985, the field of syntactic pattern recognition almost died with him. After lying nearly dormant for a quarter of a century, stochastic image grammars are resurging as a common framework for studying diverse vision problems. This is, in part, thanks to recent advances in grammar modeling, learning, and inference techniques.

The notion of stochastic image grammars has come a long way from assigning probabilities to production rules as suggested by Prof. Fu to more general notion that encompasses hierarchical representations of objects and events, semantic and spatiotemporal contexts, taxonomy of visual categories, and their associated learning and inference algorithms. Recent work shows that the virtue of image grammars lies in their expressive power to represent an exponentially large number of object and event configurations by using a relatively much smaller vocabulary, and a few compositional rules. In addition to objects and events, various other semantic contents can be associated with different levels of hierarchical descriptions in grammars, facilitating rich image interpretations.

This special issue on stochastic image grammars is aimed at presenting recent progress in grammar-based formulations in vision. It consists of seven papers. In particular,

- Lo-Bin Chang et al. in “Context, Computation, and Optimal ROC Performance in Hierarchical Models” propose to model the compositional nature of an observed world as a probabilistic hierarchy of spatial arrangements of object parts. While optimal object recognition using such a model is NP-complete, they provide mathematical evidence that part-based inference can achieve nearly optimal recognition within a feasible number of operations.
- Vinay Shet et al. in “Predicate Logic based Image Grammars for Complex Pattern Recognition” extend first-order predicate logic with a bilattice-based formalism to encode grammars of image patterns that can handle uncertain positive and negative information from different detectors. They propose a rule weight optimization method for inference that converges to a set of optimal rules (i.e., explanations) within the bilattice.
- Liang Wang et al. in “Mining Layered Grammar Rules for Action Recognition” propose a generative hierarchical model to represent human actions whose bottom layer represents action primitives, and each higher layer encodes compositional rules of high-order actions. An action is recognized based on the discriminative configurations generated by the production rules of the parse tree.
- Ryoo and Aggarwal in “Stochastic Representation and Recognition of High-level Group Activities” specify a probabilistic representation of a group activity, capturing how individual activities of its group members are orga-

S. Todorovic (✉)
Oregon State University, Corvallis, OR, USA
e-mail: sinisa@eecs.oregonstate.edu

R. Chellappa
University of Maryland, College Park, MD, USA

nized temporally, spatially, and logically. For recognition they use a Markov chain Monte Carlo sampling.

- Kokkinos and Yuille in "Inference and Learning with Hierarchical Compositional Models" introduce a hierarchical representation for object detection. They represent an object in terms of shape-based parts, which are composed of contours corresponding to the object boundaries and symmetry axes; these are in turn related to edge and ridge features that are extracted from the image. They then propose a coarse-to-fine algorithm for efficient detection which exploits the hierarchical nature of the model. The models are learnt from training images where only the bounding box of the object is provided.
- Wu and Zhu in "Numerical Study of the Bottom-up and Top-down Inference Processes in And-Or Graphs" present a numerical study of the bottom-up and top-down inference processes in hierarchical models using the And-Or graph as an example. Two hierarchical case studies, one on junctions and rectangles in low-to-middle-level vision and the other related to human faces in high-level vision are presented.
- Finally, Simon, Teboul, Koutsourakis and Paragios in "Random Exploration of the Procedural Space for Single-View 3D Modeling of Buildings", address the problem of 3D modeling for urban environment using a modular and flexible approach driven from procedural generation. Types of architectures are modeled through shape grammars that consist of a set of derivation rules and a set of shape/dictionary elements. Appearance of the dictionary elements is then learned using a set of training images. Then, given a new image and the corresponding footprint, the modeling problem is formulated as an exploration of the space of shapes that can be generated on-the-fly by deriving the grammar on the put axiom.

The papers included in this special issue present both theoretical and empirical evidence of advantages of using stochastic image grammars in computer vision. We hope that this first issue in IJCV will lead to new mathematical theories and algorithms toward a unifying theory of image grammars. We would like to thank the reviewers involved, and the Editorial Office of the IJCV for supporting this special issue.

Guest Editors