# Program Chairs' Introduction to the First International Workshop on Stochastic Image Grammars (SIG-09) in conjunction with IEEE CVPR 2009

Sinisa Todorovic
School of EECS
Oregon State University

Song-Chun Zhu
Departments of Statistics and Computer Science
University of California Los Angeles

## 1. Motivation

Hierarchical models, semantic contexts, compositionality, taxonomy of visual categories, visual event ontology, stochastic graph matching, and bottom-up/top-down inference are popular research topics in computer vision and pattern recognition. They can be viewed as different aspects of stochastic image grammars. 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 semantic contexts can also be associated with all levels of hierarchical descriptions in grammars, facilitating rich image interpretations.

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 modeling, learning, and inference techniques. It seems, however, that the renewed interest in image grammars often ignores the problem formulations and solutions of earlier work. Also, old solutions to fundamental questions (e.g., "what is a visual word," "what are object/event parts," and "what is a visual context") are too often taken for granted by the new generation of researchers. In addition, this renaissance of image grammars seems to be a result of individual team efforts, rather than a collaborative study toward a unifying theory.

Therefore, we believe that the timing is right to organize SIG-09, aimed at:

- Reducing the historical disconnect from early work,

- Illuminating directions for future research, and

- Bringing together researchers from different subcommunities and thus increasing the interdisciplinary awareness and collaboration.

## 2. Research Topics

The major theme of SIG-09 is to identify challenges facing the work toward a unifying theoretical foundation of stochastic image grammars. This unification is twofold, and concerns formulating a grammar that would:

- Jointly address vision problems that are traditionally viewed as distinct and thus solved separately (e.g., object segmentation and activity recognition), and

- Encompass mathematical theories and techniques traditionally viewed as distinct and used by different subcommunities (e.g., harmonic analysis, Bayesian inference, sparse coding, Markov random fields, graphical models, etc.)

In particular, the workshop explores the following interrelated research topics:

1. Learning vocabularies of visual primitives and parts

2. Stochastic hierarchical models of events, activities, scenes, and 2D/3D objects

3. Hierarchy of classifiers for object/scene/activity categorization

4. Taxonomies and ontologies of a large number of activities, and objects

5. Statistical and graph-theoretic learning/inference algorithms for image grammars

6. The use of grammars for image parsing, and generating sentences of image interpretation at different levels of abstraction

7. Image datasets and benchmarks for learning and evaluating stochastic image grammars

## 3. Program

We invited six distinguished researchers to give keynote talks at SIG-09. The speakers have a long-track record of research on image grammars. Also, they have diverse backgrounds and scientific interests. This helps SIG-09 achieve a high degree of depth and breadth in the treatment of the topics of interest. Our call for papers produced a number of submissions that were reviewed by members of the Program Committee. Based on rigorous recommendations of the Program Committee, we selected five papers for presentation at SIG-09. This work represents the state of the art of the field.

## 4. Program Committee

Narendra Ahuja (Univ. of Illinois at Urbana-Champaign)
Kobus Barnard (University of Arizona)
Rama Chellappa (University of Maryland, College Park)
Jason Corso (University at Buffalo, SUNY)
Pedro Felzenszwalb (University of Chicago)
Alan Fern (Oregon State University)
Edwin Hancock (The University of York)
Iasonas Kokkinos (Ecole Centrale Paris)
Mun Wai Lee (ObjectVideo Inc.)
Ales Leonardis (University of Ljubljana)
Greg Mori (Simon Fraser University)
Ram Nevatia (University of Southern California)
Vladimir Pavlovic (Rutgers University)
Ilya Pollak (Purdue University)
Deva Ramanan (University of California, Irvine)
Mubarak Shah (University of Central Florida)
Erik Sudderth (Brown University)
Antonio Torralba (Massachusetts Institute of Technology)
Andrea Torsello (Ca' Foscari University of Venice)
Zhuowen Tu (University of California, Los Angeles)
Nuno Vasconcelos (University of California, San Diego)
Ying Nian Wu (University of California, Los Angeles)
Alan Yuille (University of California, Los Angeles)
Shimon Ullman (Weizmann Institute of Science)

## 5. Acknowledgments

# SIG-09 Program

| | |
|---|---|
| 8:30 – 8:35 | Song-Chun Zhu (University of California, Los Angeles) <br><br> "Welcome and Opening Remarks" |
| 8:35 – 8:50 | Sinisa Todorovic (Oregon State University) <br><br> "Brief Review of Grammars" |
| 8:50 – 9:30 | Stuart Geman (Brown University) <br><br> Keynote talk: "Generative Hierarchical Models for Image Analysis" |
| 9:30 – 10:10 | Ales Leonardis (University of Ljubljana) <br><br> Keynote talk: "Learning a Hierarchical Compositional Representation of Multiple Object Classes" <br><br> S. Fidler, M. Boben, and A. Leonardis (University of Ljubljana) <br><br> "A bottom-up and top-down optimization framework for learning a compositional hierarchy of object classes" |
| 10:10 – 10:30 | Coffee break |
| 10:30 – 11:05 | David Mumford (Brown University) <br><br> Keynote talk: "Is There a General Structure for Grammars?" |
| 11:05 – 12:00 | Alan Yuille (University of California, Los Angeles) <br><br> Keynote talk: "Recursive Compositional Models: Representation, Learning, and Inference" <br><br> I. Kokkinos and A. Yuille (Ecole Centrale de Paris and University of California, Los Angeles) <br><br> "Learning and inference using hierarchical compositional models" |
| 12:00 – 1:30 | Lunch break |
| 1:30 – 1:50 | V. Shet, M. Singh, C. Bahlmann, and V. Ramesh (Siemens Corporate Research, Princeton, USA) <br><br> "Predicate Logic based Image Grammars for Complex Pattern Recognition" |
| 1:50 – 2:20 | B. Yao, X. Yang, and T. Wu (Lotus Hill Institute and University of California, Los Angeles) <br><br> "Image Parsing with Stochastic Grammar: The Lotus Hill Dataset and Inference Scheme" |
| 2:20 – 3:05 | Narendra Ahuja (University of Illinois at Urbana-Champaign) <br><br> Keynote talk: "A Syntax for Image Understanding" |
| 3:05 – 3:20 | Coffee break |
| 3:20 – 3:40 | M. S. Ryoo and J. K. Aggarwal (University of Texas at Austin) <br><br> "Stochastic Representation and Recognition of High-level Group Activities" |
| 3:40 – 4:25 | Sven Dickinson (University of Toronto) <br><br> Keynote talk: "Beyond One-to-One Feature Correspondence: The Need for Many-to-Many Matching and Image Abstraction" |
| 4:25 – 5:00 | Panel discussion |