

## **"In Search of a Unifying Theory for Image Interpretation"**

Don Geman

Johns Hopkins University

Applied Mathematics and Statistics, and Center for Imaging Science

### **Abstract**

Image interpretation, which is effortless and instantaneous for human beings, is the grand challenge of computer vision. The dream is to build a "description machine" which produces a rich semantic description of the underlying scene, including the names and poses of the objects that are present, even "recognizing" other things, such as actions and context. Accuracy entails invariance to transformations which preserve semantic labels. Also, such a machine should learn and run efficiently. Many would argue that we are at best inching towards this goal; certainly algorithms and strategies abound for sub-tasks and some generic tools have been honed.

But what about a unifying theory, such as the one Shannon provided for communication -- one which achieves abstraction and still guides the development of practical algorithms? Might this accelerate progress? Mathematical frameworks are advanced from time to time, but none is yet widely accepted, and none clearly points the way to closing the gap with natural vision. There might be some lessons from information theory, for example thinking about representation and processing at the same time. Finally, I will argue that hierarchical structures should be at the core of such a theory and support this claim by summarizing several theoretical results about the efficiency of hierarchical search and representation.