OPPORTUNITIES FOR THE MATHEMATICAL SCIENCES

JUNE 26-27, 2000 WORKSHOP

Complex Stochastic Models for Perception and Inference

D. MUMFORD

The problem of duplicating the human intellectual skills of perception and inference in a computer has a long and checkered history. Promises have been made roughly every ten years since 1950 that a "breakthrough" had been found or was just around the corner. A new wave of optimism has arisen in the last five years or so, which, participants hope, is based on a deeper and sounder appreciation of the nature of these skills and the computations needed to sustain them.

The basis of the new set of ideas is to recast everything, every percept, every thought, as a probability. Although not new, what characterizes the present attack is that more complex stochastic models are being developed, more sophisticated algorithms for statistical inference with these models are being invented, and larger datasets are being crunched by learning algorithms that feed these models. In particular, the fields of neural nets, computer vision and natural language understanding are seeing a wave of new results based on these ideas.

Stochastic methods started in control theory, e.g., with the Kalman filter, and spread to speech recognition, e.g., with the hidden Markov models. They spread to AI as a method for making expert systems more robust, e.g., with the theory of Bayesian belief networks. However, in vision, stochastic methods were slow to show their power because of the huge diversity of images and the computational resources of both speed and memory required to deal statistically with them. I will focus on the field of vision, which illustrates well what is new in the present attack on this whole array of problems.

We are accustomed to the fact that our brains present our conscious self seemingly instantaneously not merely with a retinal image of the world but with a full explanation of this image as a 3D scene with multiple recognized and labeled objects. The effortlessness of this process masks the fact that this "parsing" of an image is extremely hard to compute and involves massive inference based on past experience. How and why are we making progress understanding this process now?

i.   Memory storage and speed are becoming adequate to deal with images. All the images seen by a kitten learning the geometry and visual appearance of the world can be stored roughly in a terabyte; but databases of world scenes (and sequences of scenes) of 10 gigabytes and more are now available.

ii.   The study of the universal statistics of images has shown that images must be modeled by very non-Gaussian statistics, and this has helped break the bias that Gaussian models are always reasonably good.

iii. Complex stochastic models which are both nonlinear and which incorporate mixed discrete and continuous variables have been crafted. One family of examples are new models for shape and for warping of shapes/images (e.g. warping a template medical image onto a patient's image). Another family are stochastic hierarchical models, such as probabilistic context free and context sensitive grammars which have been adapted to vision.

iv.   Ten years ago, reasoning with stochastic models was restricted to a) linear discriminants and classification trees in unstructured situations, b) dynamic programming in the simplest 1D situations, c) very slow simulated annealing Monte Carlo, and d) mean field-type crude guessing. Now all these approaches have been extended and improved. Some examples are

Last Modified: Oct 17, 2001

Vapnik's support vector machines, belief propagation on graphs with cycles, multiple sample Monte Carlo algorithms such as "particle filtering," and multi-scale "pyramid"-based adaptations of these algorithms.

The progress just outlined is the result of an extended interaction of mathematicians, statisticians, computer scientists, and engineers. The topic itself is not the intellectual property of any one of these fields, and the area of specialization of the individual scientists who have contributed appears to be a random variable with any of these four values. We have not touched on the extremely exciting interactions with people studying the real mind and brain-- cognitive psychologists and neuroscientists.