Bayesian Models for Visual Information Retrieval

by

Nuno Miguel Borges de Pinho Cruz de Vasconcelos

Submitted to the Program in Media Arts and Sciences,
on June 2, 2000 in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis presents a unified solution to visual recognition and learning in the context of visual information retrieval. Realizing that the design of an effective recognition architecture requires careful consideration of the interplay between feature selection, feature representation, and similarity function, we start by searching for a performance criteria that can simultaneously guide the design of all three components. A natural solution is to formulate visual recognition as a decision theoretical problem, where the goal is to minimize the probability of retrieval error. This leads to a Bayesian architecture that is shown to generalize a significant number of previous recognition approaches, solving some of the most challenging problems faced by these: joint modeling of color and texture, objective guidelines for controlling the trade-off between feature transformation and feature representation, and unified support for local and global queries without requiring image segmentation. The new architecture is shown to perform well on color, texture, and generic image databases, providing a good trade-off between retrieval accuracy, invariance, perceptual relevance of similarity judgments, and complexity.

Because all that is needed to perform optimal Bayesian decisions is the ability to evaluate beliefs on the different hypothesis under consideration, a Bayesian architecture is not restricted to visual recognition. On the contrary, it establishes a universal recognition language (the language of probabilities) that provides a computational basis for the integration of information from multiple content sources and modalities. In result, it becomes possible to build retrieval systems that can simultaneously account for text, audio, video, or any

other content modalities.

Since the ability to learn follows from the ability to integrate information over time, this language is also conducive to the design of learning algorithms. We show that learning is, indeed, an important asset for visual information retrieval by designing both short and long-term learning mechanisms. Over short time scales (within a retrieval session), learning is shown to assure faster convergence to the desired target images. Over long time scales (between retrieval sessions), it allows the retrieval system to tailor itself to the preferences of particular users. In both cases, all the necessary computations are carried out through Bayesian belief propagation algorithms that, although optimal, are extremely simple, intuitive, and easy to implement.

Thesis Supervisor: Andrew B. Lippman
Associate Director, MIT Media Laboratory