IMAGE REPRESENTATION FOR IMAGE UNDERSTANDING

Narendra Ahuja

Department of Electrical and Computer Engineering, Coordinated Science Laboratory, and
Beckman Institute,
University of Illinois at Urbana-Champaign
Urbana, Illinois, USA

We consider one of the most basic questions in computer vision, that of finding a low-level image representation that could be used to seed diverse, subsequent computations of image understanding. Can we define a relatively general purpose image representation which would serve as the basis for diverse needs of image understanding? What makes good image representation? How do we evaluate it?

The representation we develop is called *Connected Segmentation Tree* (CST), defined in terms of intra- as well as inter-inter-region properties of image segments. It consists of three parts: (1) The nodes in the tree correspond to the extracted multiscale image segments, and capture their geometric and photometric properties. (2) The parent child relationships capture recursive embedding (containment) of regions. (3) Additional links among tree nodes specify regions' neighbors in all directions, thus capturing spatial layout of regions. We describe the derivation of CST from images, and its invariance to changes in imaging conditions (e.g., lighting, scale, orientation).

For concreteness, we perform this exercise in the specific context of the following problem. We wish to discover *a priori* unknown themes that may characterize a given, random or strategically chosen, set of images. If objects from a certain categories occur frequently in the set, we say that the categories constitute the theme. No specific categories are specified by the user; indeed, they are not even known to the user *a priori*. Whether, how many, or where instances of any categories appear in a specific image is also not known. To this end, we develop answers to the following basic questions. What is an object category? If, and to what extent, is human supervision necessary to communicate the nature of categories to a computer vision system? What properties should be used to define a good category representation? We define an object category as consisting of (2D) subimages that have similar photometric, geometric and topological properties. We pose the following subproblems: (1) Discovering whether any categories occur in the image set. (2) Building a compact model that captures the intrinsic nature of the categories. (3) Learning the relationships among the different categories, thus building a taxonomy of all discovered categories. (4) Using the learned taxonomy to recognize all occurrences of all categories in previously unseen images. (5) Segmenting each occurrence. (6) Articulating the reasons for recognition. We present solutions to (1-6) that are almost completely unsupervised.

A low-level image representation should be useful to diverse tasks. The general nature of (1-6) helps extend their solutions to detecting themes of other kinds. As the second problem, we present one such extension, that of identifying and extracting stochastically repeating parts of visual textures, commonly called texture elements. We evaluate the performance of CST here through the quality of detected elements in real-world textures.