On Detection and Representation of Multiscale Low-Level Image Structure

NARENDRA AHUJA

Beckman Institute and Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign ⟨ahuja@vision.ai.uiuc edu⟩

The objective of computer vision is interpretation of visual images. Any data-interpretation task of such magnitude requires models of the data. For example, in speech the audio signal is parsed into phonemes, which are successively merged into increasingly complex units and eventually into an interpretation, often with feedback from higher levels. Another example is hierarchical interpretation of computer programs in a given language through the use of grammars. In image data, analogues of phonemes and characters correspond to structural primitives that compress the data to a manageable size without eliminating any possible final interpretations.

Because images are significantly larger and more complex than speech signals, a capability for initial, bottom-up data reduction is even more critical. The low-level structure would serve as a lossless image abstraction and help initiate hierarchical, closed-loop image interpretation, for example, for recognition by enforcing a priori semantic constraints involving part–whole relationships. This note is not concerned with interpretation processes; it describes some desirable characteristics of strategies for the detection and representation of low-level perceptual structure or multiscale segmentation, which remains an open problem.

Homogeneous image regions may be used as structural primitives. A region can be characterized as possessing a certain degree of interior homogeneity and a contrast with the surround that is large compared to the interior variation. This is a satisfactory characterization from both the perceptual and quantitative viewpoints. Past work on image segmentation has not yielded acceptable algorithms. This is due to the following main challenges. First, the type of region homogeneity and the magnitude of the contrast may vary and the regions may have arbitrary size and shape. Although a region can be detected by identifying either its interior or its border, the latter method has been more thoroughly investigated. These methods use different models of border geometry (e.g., straightness), and brightness variation along borders (e.g., linearity), across borders, and within regions. Most methods are linear. Although such models and methods simplify processing, they lead to fundamental limitations in the detection accuracy and the sensitivity of the resulting segmentation.

The second reason involves the multiscale nature of image structure, that is, geometric and photometric sensitivity to detail. A pixel may belong simultaneously to different regions, each having a different contrast value (photometric scale) and size (geometric scale). Analogously, an edge contour that separates two regions of a given contrast scale may not be detected at a higher scale associated with a larger contrast.

If the sensitivity to contrast and homogeneity is increased, the resulting regions are more homogeneous and have lower contrast. Although the general notion of multiscale operators has been examined for a long time, there has been limited work on the definition, analysis, and automatic detection of multiscale image structure.

Thus the major issues in successful multiscale image segmentation include:

Shape and topology invariance: The regions should be correctly detected regardless of their shapes and relative placement. For example, a border point must be detected at only one and the correct location, regardless of whether the edge in the vicinity of the point is straight, curved, or even, or contains a corner or a vertex where multiple regions meet.

Photometric scaling: It should be possible to detect all regions that are in contrast to their surround, regardless of the actual degree of within-region homogeneity and the value of the contrast.

Spatial scaling: It should be possible to detect all regions regardless of their sizes.

Stability and automatic scale selection: Because the contrast and sizes of regions contained in an arbitrary image are *a priori* unknown, they should be identified automatically along with the associated structures.

Consequently, the ultimate objective should be to derive a multiscale segmentation of the image and represent it through a hierarchical (usually tree) structure in which the different image segments, their parameters, and their spatial interrelationships are made explicit. The bottom (leaf) nodes of such a hierarchy correspond to regions consisting of individual image points or connected components of constant gray level, and the path from a leaf to the root node specifies how the leaf regions recursively merge with adjacent regions to form larger regions, each of which is homogeneous relative to its surround and is characterized by its own contrast. The resulting structural information is associated with each image pixel to form an annotated pixel array. Alternate representations of the same image structure and contrast information are also possible, for example, by ordering regions according to contrast.

To achieve such segmentation performance it is necessary that minimal *a priori* restrictions be placed on the detection process. One way of achieving this is to perform detection by identifying groupings of image points, rather than by testing for the existence of specific local structures formed by the points, as commonly done in the past. This would allow the structure to "emerge" bottom-up from "interactions" among the points, instead of imposing *a priori* chosen models of region shape. As one consequence of this, the emergent region geometry is not restricted, because pixels can group together to form any connected set. Such an approach is analogous to physical processes in which microscopic homogeneity of physical properties leads to islands of, say, similar particles or molecules. An island shape is congruent with the space occupied by a set of contiguous, similar particles, regardless of how complex the island shape is. The particles group together and coalesce into regions based on the similarity of their intrinsic properties only, regardless of their relative locations. The common property of particles then characterizes the region they form. As an alternate analogy, the groupings process is like the alignment of microscopic domains over an area of ferromagnetic material. The key process is that of interaction among particles which leads to bindings among similar particles.

The scientific significance as well as practical impact of a solution to the problem of multiscale-structure detection with the characteristics stated previously will be major, ranging from the design of human-friendly medical aids to weather data interpretation to automatic video access and retrieval. For example, the structural details of brain anatomy captured in magnetic resonance images could be seen in three dimensions, and human interactions with video databases could be carried out naturally, using the familiar image-space descriptors rather than the less transparent, but currently standard, transform-space representations.