Beyond One-to-One Feature Correspondence: The Need for Many-to-Many Matching and Image Abstraction

Sven Dickinson
Department of Computer Science
University of Toronto

Over the last four decades, object recognition systems have typically assumed a one-to-one correspondence between image and model features. Early categorization systems, with their roots in the AI community, attempted to model the prototypical shape of an object using features whose high level of abstraction supported categorization under the one-to-one correspondence assumption. However, such abstract features could not be effectively recovered from real images of real objects, defining a representational gap that was gradually eliminated by making models more specific and less generic. The propensity toward local features combined with the one-to-one feature correspondence assumption shifted the focus of recognition from categories to exemplars.

The renaissance of object categorization is still rooted primarily in local image features and their one-to-one feature correspondence. However, within-class variation of shape, structure, and appearance, part articulation, and noise and segmentation errors all demand that we move beyond this restrictive assumption. More generally, we need recognition frameworks that can match features many-to-many, of which grammar-based frameworks represent a powerful subclass. Until we develop more powerful local feature grouping and abstraction mechanisms, many-to-many feature matching is particularly critical if we are to move from highly restricted categories to more general categories.

In this talk, I will briefly review three formulations of the many-to-many matching problem as applied to model acquisition, model indexing, and object recognition. In the first scenario, I will describe the problem of learning a prototypical shape model from a set of exemplars in which the exemplars may not share a single local feature in common. We formulate the problem as a search through the intractable space of feature combinations, or abstractions, to find the "lowest common abstraction" that is derivable from each input exemplar. This abstraction, in turn, defines a many-to-many feature correspondence among the extracted input features.

In the second scenario, I will describe a low-dimensional encoding of graph structure, yielding a structural abstraction that's invariant to minor perturbations of graph size and structure. This abstraction is derived from spectral properties of the graph's underlying adjacency matrix, and forms the backbone of a unified framework for both graph indexing and graph matching. Like the first scenario, similar abstractions define a many-to-many correspondence between the nodes in their respective graphs. However, unlike the first scenario, the abstraction is defined for a single graph, thereby supporting the important subtask of indexing.

In the third, and final, scenario, we again address the intractable problem of many-to-many graph matching. Drawing on low-distortion graph embedding techniques, we first embed the nodes of two graphs to be matched into a low-dimensional vector space, such that attributed nodes map to weighted points and shortest-path distances between nodes map to Euclidean distances between points. Next, the two weighted point distributions are efficiently matched many-to-many using the Earth Mover's Distance algorithm. The resulting flows, like the first two scenarios, define a many-to-many node correspondence between the original graphs. However, unlike the first two scenarios, explicit abstractions are not generated.

These three scenarios each illustrate the need to move beyond one-to-one feature correspondence and accommodate the structural variability due to within-class variation and imperfect feature extraction. More importantly, they illustrate the need for more powerful image abstractions to help manage this variability.