

Recognizing Objects in Images

Jitendra Malik

UC Berkeley

In recent years, the problem of visual object recognition has been modeled in the framework of statistical pattern classification, resulting in some striking progress. However for realistic versions of the task, such as object detection in natural scenes as in the PASCAL benchmark, performance numbers of the best systems are still in the the 30-40% range.

I believe that if our goal is to model the human visual recognition system, or to design more practically effective computer recognition systems, we need a richer formalism. Just as we should not formulate the child language acquisition problem as one of starting from a set of transcribed sentences with no access to cues such as from phonetics or social communicative context, so also in vision, we need to consider the rich input which children can exploit to acquire their visual vocabulary. In particular, perceptual organization, object tracking, and functional interaction provide very useful scaffolding for the acquisition of visual object categories.

I will present some specific results in this general philosophy. Having access to a notion of corresponding keypoints across different exemplars enables us to derive a notion of part, “poselet” which deals with issues such as varying 3d pose, articulation, and occlusion. Combining this with the use of bottom-up grouping gives us a powerful attack on the grand challenge of visual object recognition and segmentation.

Joint work with Lubomir Bourdev, Subhransu Maji and Thomas Brox.