

A conceptual framework of computations in mid-level vision

Jonas Kubilius, Johan Wagemans, & Hans P. Op de Beeck (KU Leuven, Belgium)

The goal of visual processing is to extract information necessary for a variety of tasks, such as grasping objects, navigating in scenes, and recognizing them. While ultimately these tasks might be carried out by separate processing pathways, they nonetheless share a common root in the early and intermediate visual areas. What representations should these areas develop in order to facilitate all of these higher-level tasks? Several distinct ideas have received empirical support in the literature so far: (i) boundary feature detection, such as edge, corner, and curved segment extraction; (ii) second-order feature detection, such as the difference in orientation or phase; (iii) computation of summary statistics, that is, correlations across features. Here we provide a novel synthesis of these ideas into a single framework. We start by specifying the goal of mid-level processing as the construction of surface-based representations. To support it, we propose three basic computations: (i) computation of feature similarity across local neighborhoods; (ii) pooling of highly similar features, and (iii) inference of new, more complex features. These computations are carried out hierarchically over increasingly larger receptive fields and refined via recurrent processes when necessary.

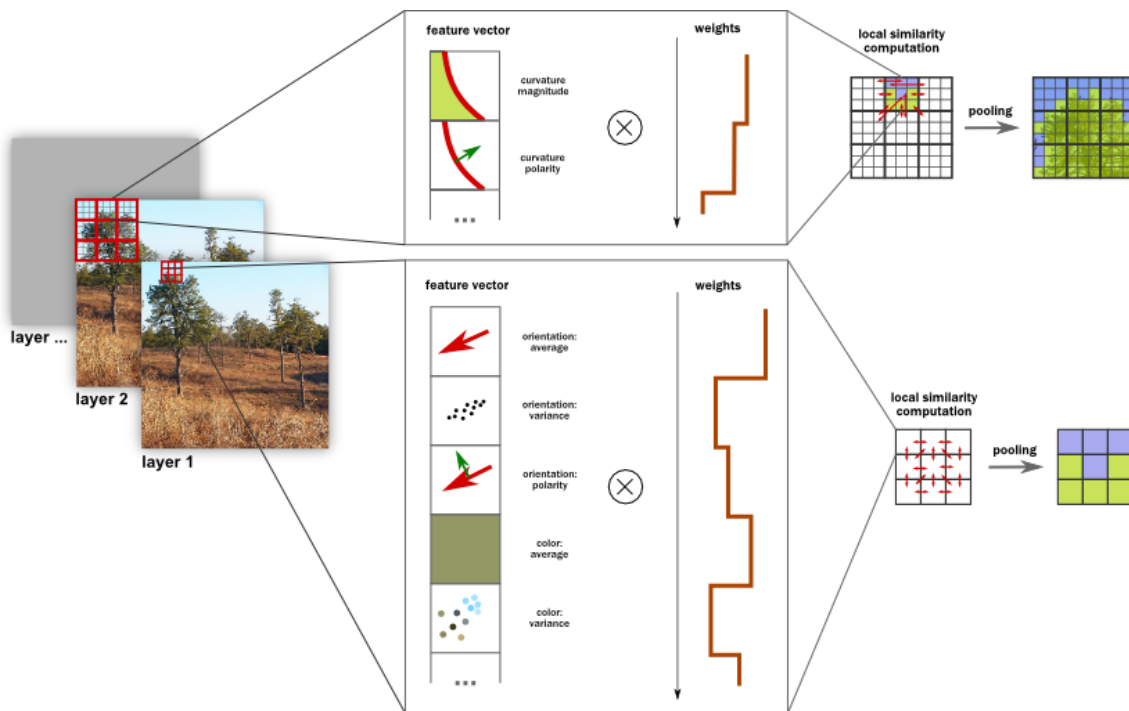


Figure. An illustration of the proposed computations. The detected features are clustered, and the new features are inferred in the next layer over larger receptive fields. (Image modified from Kubilius et al., *Front. Comput. Neurosci.*, 2014.)