Domain-Size Pooling in Local Descriptors for Visual Correspondence

Jingming Dong, Stefano Soatto
UCLA VisionLab, {dong, soatto}@cs.ucla.edu

Abstract

We introduce a simple modification of local image descriptors, such as SIFT, that improves matching performance by 39.41% on the Oxford image matching benchmark and is implementable in a few lines of code. To put things in perspective, this is more than half of the improvement that SIFT provides over raw image intensities on the same dataset. The approach consists of pooling gradient orientations across different domain sizes, in addition to spatial locations, and yields a descriptor of the same dimension of the original, which we call DSP-SIFT.

Domain-size pooling causes DSP-SIFT to outperform a Convolutional Neural Network by 21.32%. Domain-size pooling is counter-intuitive and contrary to the practice of scale selection as taught in scale-space theory, but has solid roots in classical sampling theory.

Domain-size pooling can also be applied to the intermediate layers of many architectures such as scattering transform networks, deformable part models and convolutional neural networks.

What do we do?

Compute optimal local representation for finding correspondences

Why is it important?

Various applications: 3D reconstruction, image based localization, mapping and loop closure, objection recognition and detection.

Quantitative Evaluation

Performance with increasing image transformation

Head-to-Head (DSP-SIFT, SIFT, Convolutional Neural Network)

Performance Complexity Tradeoff

Application to Ultra Wide-baseline Matching