Self-localization on texture statistics

Sven Eberhardt (sven2@uni-bremen.de), Christoph Zetzsche and Kerstin Schill

Abstract
The ability to localize ourselves in the outdoor world based on visual input even in absence of prior positional information is an important skill of our daily lives that comes naturally to us. However, the underlying mechanisms of this ability are poorly understood [1][2]. Here, we show how simple texture statistics can be sufficient to provide a strong prior for the self-localization tasks. We find that statistics of common outdoor features such as tree density, foliage type or road structure provide a stronger cue for self-localization than the matching and recognition of less common landmarks such as lamp posts.

Models
• Distinct biologically inspired vision models provide visual feature descriptors.
• Descriptors may either respond only to a select number of distinct features (landmark-based) or build histograms over more common features (Holistic descriptors).
• Models from different areas in human visual system modeling are tested on the localization task: Animal detection (HMax)[3], Scene recognition (Gist[4], Spatial Pyramids)[5], Image segmentation (Textons)[6]
• Output vectors tested on localization task using one-versus-all linear regression[7]

Datasets
Results validated on a wide range of datasets.
• Google StreetView data at world scale, country scale and city scale
• Virtual world screenshot data[8]
• Indoor localization[9]
• Comparison datasets: Caltech101, Serre Animal/Nonanimal, Scene15

References

Localization task
Localization differs in its invariance requirements from other tasks such as object or scene recognition tasks. Therefore, it’s not clear which feature vectors used in other areas apply to self-localization tasks.[10]

Results
Performances
• Simple texture statistics sufficient to provide a strong prior for the self-localization tasks.
• Statistics of common outdoor features: Tree density, foliage type or road structure stronger than landmarks
• Use of such common feature vectors as priors for self-localization systems
• Stable across all datasets (indoor, virtual, streetview) and scales

Location untangling
• Models that separate locations well untangle the space and cause little variation as the observer rotates.

Texton performance
• Small number of clusters (500) sufficient for accurate classification
• Texture sampling from very small image areas ($\sigma = 0.5$) outperforms larger patches