Self-localization on texture statistics

A5-[ActionSpace]

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]

Landmarks, Sparse features

Object parts (HMax)

Spatial envelope (Gist)

Texture distribution (Spatial Pyramids)

Texture histogram (Textons)

Luminance Histogram

Holistic image descriptors, Image statistics

Datasets

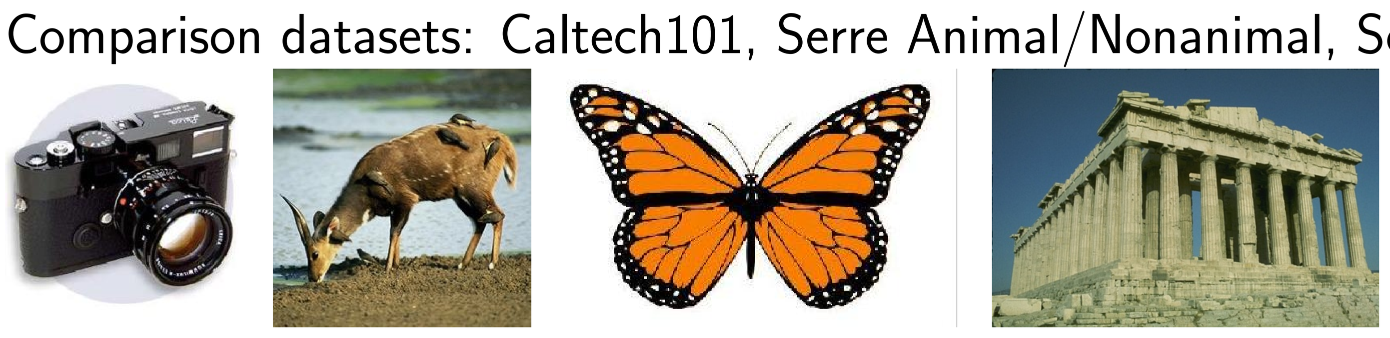
Results validated on a wide range of datasets.

- 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

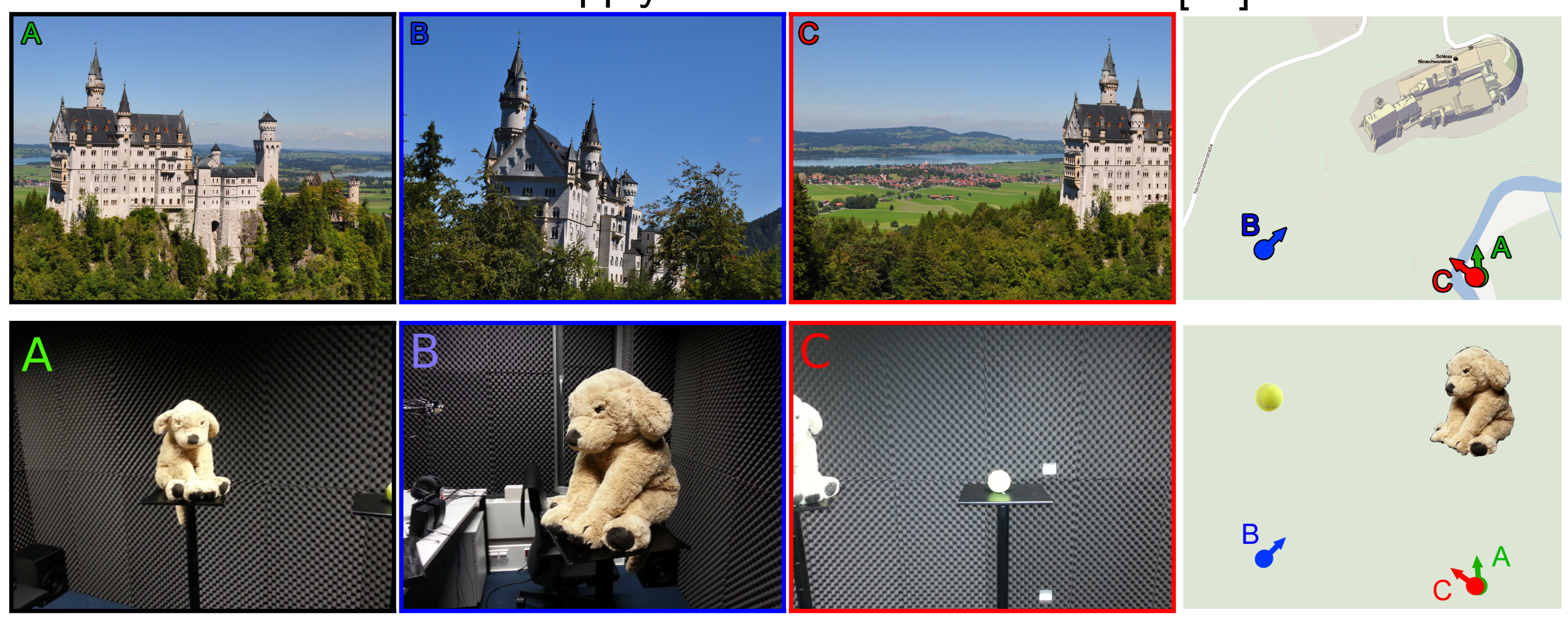


References

[1] S. Eberhardt, T. Kluth, C. Zetzsche, and K. Schill, "From pattern recognition to place identification," in Spatial cognition, international workshop on place-related knowledge acquisition research, 2012, pp. 39–44.
[2] S. Eberhardt and C. Zetzsche, "Low-level global features for vision-based localization.," in Proceedings of the KI 2013 Workshop on Visual and Spatial Cognition, 2013, pp. 5–13.
[3] T. Serre, A. Oliva, and T. Poggio, "A feedforward architecture accounts for rapid categorization.," Proceedings of the national academy of sciences, vol. 104, no. 15, pp. 6424–6429, 2007.
[4] A. Oliva and L. Ave, "Modeling the shape of the scene: A holistic representation of the spatial envelope," International journal of computer vision, vol. 42, no. 3, pp. 145–175, 2001.
[5] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in Computer vision and pattern recognition, 2006 IEEE computer society conference on, 2006, vol. 2, pp. 2169–2178.
[6]T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using threedimensional textons," International Journal of Computer Vision, 2001. [7] A. Tacchetti, P. K. Mallapragada, M. Santoro, and L. Rosasco, "GURLS: a toolbox for large scale multiclass learning.," in Big learning workshop at NIPS, 2011.
[8] Screenshots: The Elder Scrolls V: Skyrim©2011 Bethesda Softworks LLC, a ZeniMax Media company
[9] A. Pronobis and B. Caputo, "A discriminative approach to robust visual place recognition," Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on, pp. 3829-3836, 2006
[10]Photos: ©Stephen & Claire Farnsworth via flickr, license CC-BY-NC. Map: Google maps ©Google inc.

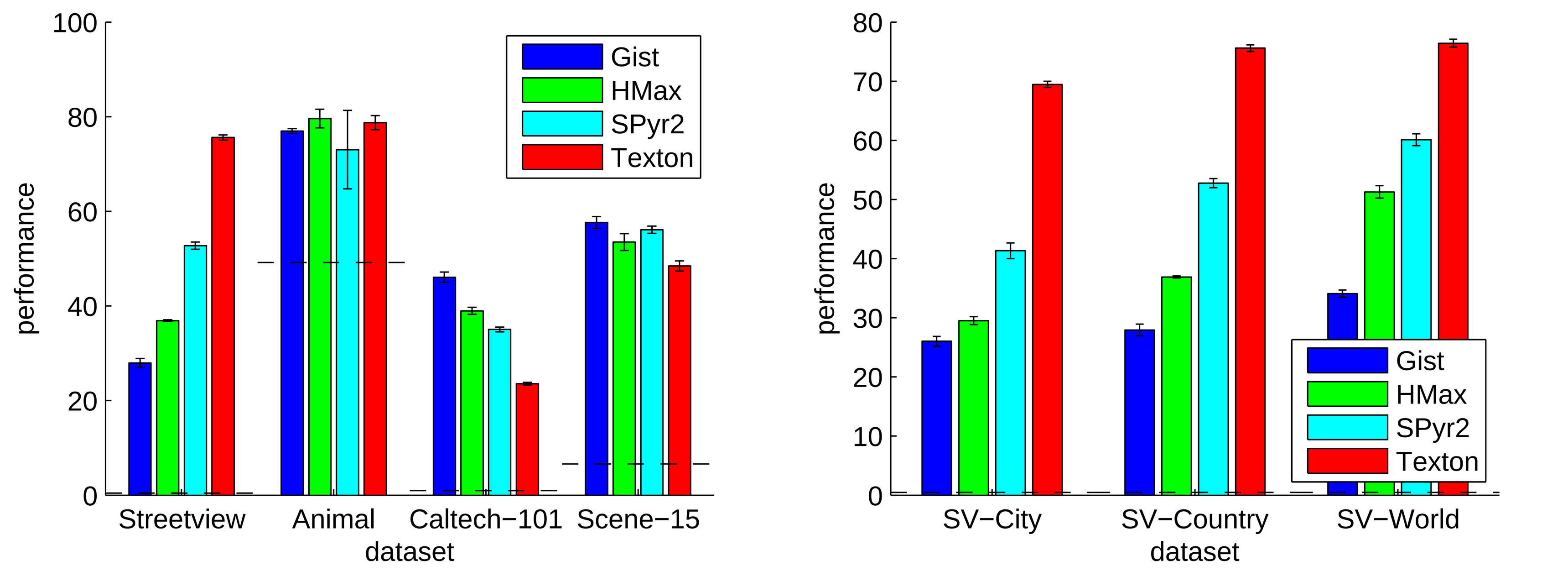
Localization task

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]



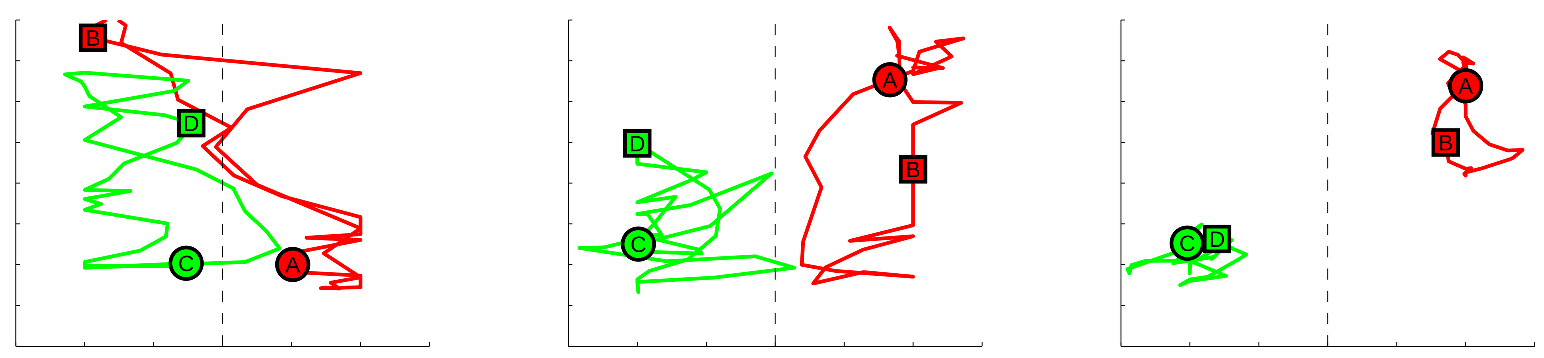
Results

Performances



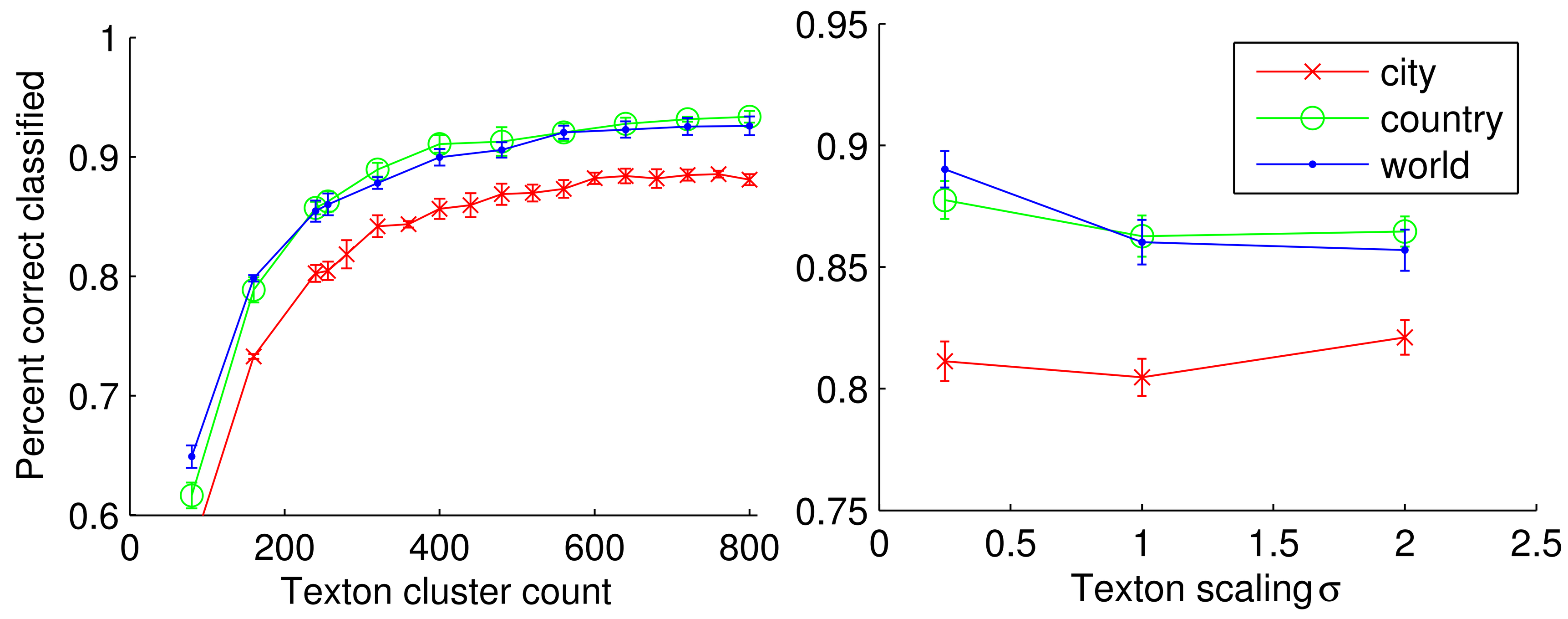
- Simple texture statistics sufficient to provide a strong prior for the self-localization tasks.
- Statistics of common outdoor features: Tree density, foliage type or road structure stronger than landmarks
- Use of such common feature vectors as priors for self-localization systems
- Stable across 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

Universität Bremen

DFG Deutsche Forschungsgemeinschaft

UNI FREIBURG