

Hierarchical Sparse Dictionaries: how sparsity, topology and pooling can improve visual place recognition

Sylvain Colomer^{1,2}, Nicolas Cuperlier¹, Guillaume Bresson², Olivier Romain¹

I. EXTENDED ABSTRACT

Robotic navigation requires complex algorithms to build an accurate and robust representation of a spatial environment. These systems must therefore constantly be able to detect useful information, to encode it and to merge it with the current representation. Despite the apparition of new efficient algorithms, they still struggle to offer a solution that is efficient and light enough to be deployed on embedded systems [2]. Nevertheless, biological systems, especially mammals, can solve these problems on a daily basis with very effective methods.

Playing a key role both in the spatial cognition in mammals and in the episodic memory in humans, the hippocampal formation is a part of the brain where the interaction of several kinds of neurons results in a neural map of our spatial environment [3]. For instance, place cells [3] were discovered with the particularity to encode precise positions in an environment. Despite decades of research on this subject, a complete view of the information processing operated by these neurons is still lacking. Relying on these observations, a few neurorobotics navigation models have been proposed to better understand the neural mechanisms involved in animals navigation by mimicking the functioning of the hippocampal formation [6].

Among the different models proposed, the neural model LPMP (Log-Polar Max-Pi) allows to build place cells in an unsupervised way by merging current visual information and the absolute orientation of the robot [1]. It leans on several important principles such as one-shot learning or Hebb principle to build a neural representation of an environment. Despite its simplicity, the model has been shown to be competitive with the state of the art, and has given encouraging first results on an autonomous vehicle [1]. The tests carried out, however, also revealed computational cost limitations which prevent it from being deployed over long distances.

To solve this problem, we propose the model HSD (Hierarchical Sparse Dictionaries), a new unsupervised model of visual information encoding intended for robot navigation. Inspired from the visual processing of primate, this model proposes to cascade "Topologic" Sparse Dictionaries (TSD) and pooling layers to build a lighter representation of a place. The integration of a topology in the sparse dictionary makes possible the application of an efficient pooling and allows

us to build, like in the Hmax model [5], a more invariant code. Thus the final architecture reproduces the first layers of the visual cortex. The sparsity constraint on the first layer is used to build orientation-sensitive neurons like simple cells [4]. The pooling layer reproduces the complex cells and the topology recreates the retinotopy properties.

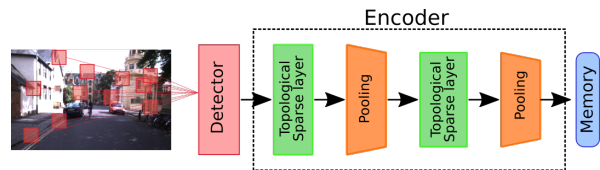


Fig. 1. **Simplified diagram of the HSD model.** The system encodes landmarks by alternating layers of topologic sparse coding and max-pooling, reducing the size of the visual feature vectors according to the number of atoms in each dictionary.

The landmark visual encoding of LPMP was replaced by HSD, leading to the new HSD+MP model for place recognition. To highlight the benefits obtained in terms of performance compared to the original model, the HSD+MP model was evaluated in terms of accuracy and computational cost on self-driving car datasets. The results showed that relying on HSD not only makes HSD-MP very efficient, with a reduction in computation time by three for equal localization performance, but also more accurate with an improvement of 10% in localization performance than LPMP while still reducing the calculation time by a factor 2. Results of these tests on large environments could also shed new lights on the activity properties of the simulated cells, whereas studies of their biological counterparts are usually conducted in small indoor environments.

REFERENCES

- [1] Y. Espada, N. Cuperlier, G. Bresson, and O. Romain, "From Neuro-robotic Localization to Autonomous Vehicles," *Unmanned Systems*, vol. 07, no. 03, pp. 183–194, July 2019.
- [2] S. Lowry, N. Sunderhauf, P. Newman, J. J. Leonard, D. Cox, P. Corke, and M. J. Milford, "Visual Place Recognition: A Survey," *IEEE Transactions on Robotics*, vol. 32, no. 1, p. 19, 2016.
- [3] M.-B. Moser, D. C. Rowland, and E. I. Moser, "Place cells, grid cells, and memory," *Cold Spring Harbor Perspectives in Biology*, vol. 7, no. 2, p. a021808, Feb 2015.
- [4] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?" *Vision Research*, vol. 37, no. 23, p. 3311–3325, Dec 1997.
- [5] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, p. 411–426, Mar 2007.
- [6] P. J. Zeno, S. Patel, and T. M. Sobh, "Review of neurobiologically based mobile robot navigation system research performed since 2000," *Journal of Robotics*, vol. 2016, p. 1–17, 2016.

¹Laboratoire ETIS UMR8051, Université Paris Seine, ENSEA, CNRS, France.

²Institut VEDECOM, 23 bis Allée des Marronniers, 78000 Versailles, France.