Improving Robot Navigation with an Explainable and Efficient Representation of the Location of Objects in a Scene

Levi Burner, Yiannis Aloimonos

University of Maryland Institute for Advanced Computer Studies

Motivation

- Robots need to navigate in all environments
- State of the art achieves autonomy in limited environments while using significant computational resources
- Biological systems navigate everywhere using low energy and relatively little computation, can we draw inspiration from them?

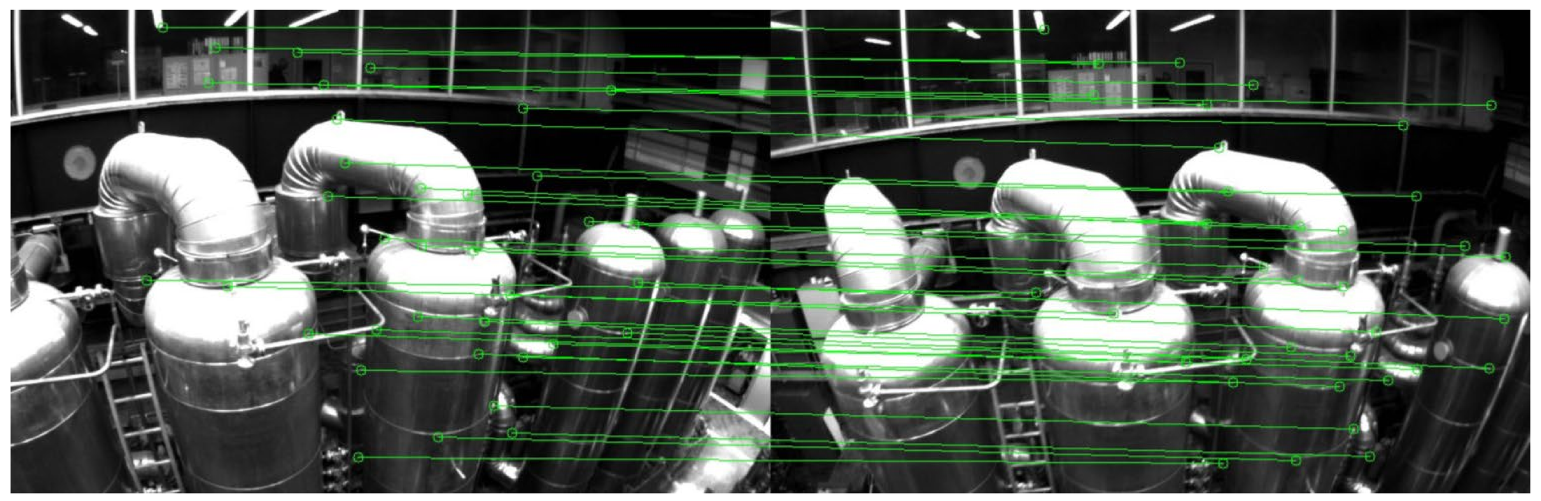


Fig 1: State of the art in robot navigation uses computationally intensive point matching between multiple views. Biologically inspired computing could help simplify the process. Figure from [1].

Background

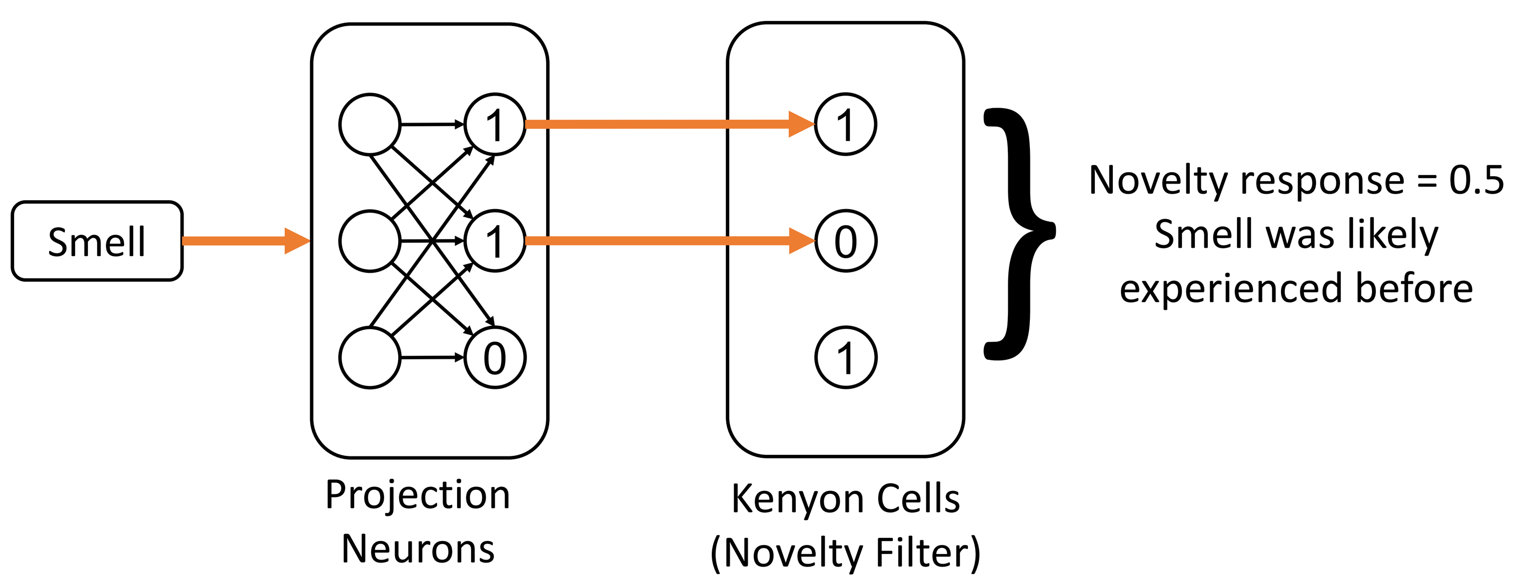


Fig 2: Fruit fly "mushroom body" generates long binary vectors which are compared with in a "novelty filter" to recall if a smell has been experienced previously [2]. Similarly, in navigation, small image patches need to be compared to memory in order to recall what other objects were close by previously.

- Generalization of such binary processing is called Hyperdimensional (HD) Computing [3]
  - Uses high dimensional vectors (>1000 dimensions)
  - Can represent sets of vectors as one vector

Fig 3: Hyperdimensional vectors can represent an entire network of objects and their locations. Instead of novelty, the HD framework can recall where the object was seen before.

Research Questions

- How many objects and their locations can be encoded in a HD vector and recalled reliably?
- What impact does the distribution of the position of objects in a scene have on this capacity?

Data and Methods

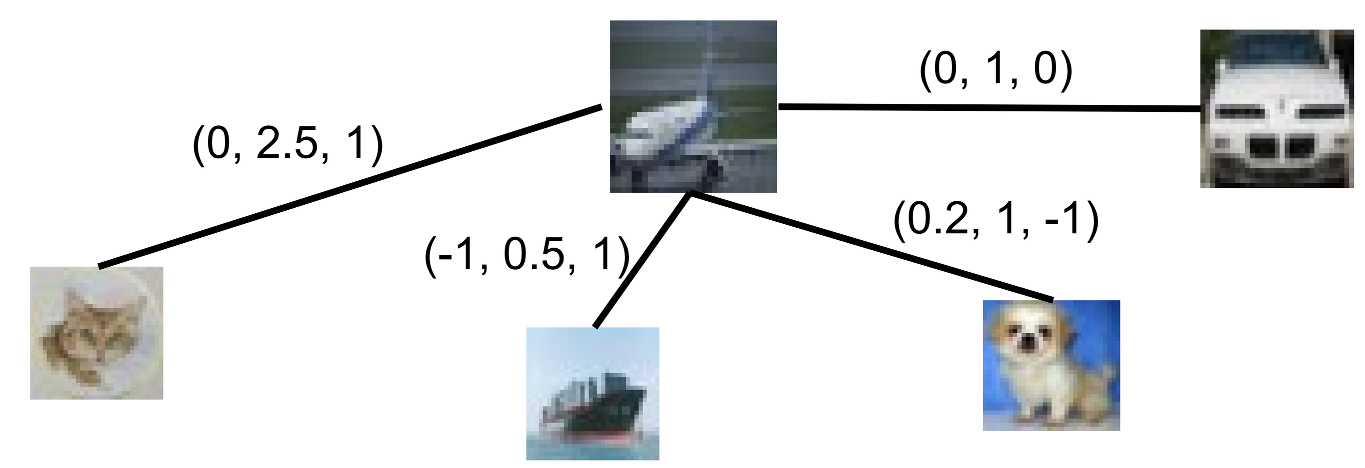


Fig 4: Generated scenes from 10,000 samples of Cifar-10 [4] with object positions drawn from uniform and gaussian distributions. Distances quantized to 100 unique values which are assigned random vectors. 1D and 2D scenes generated by fixing a dimension.

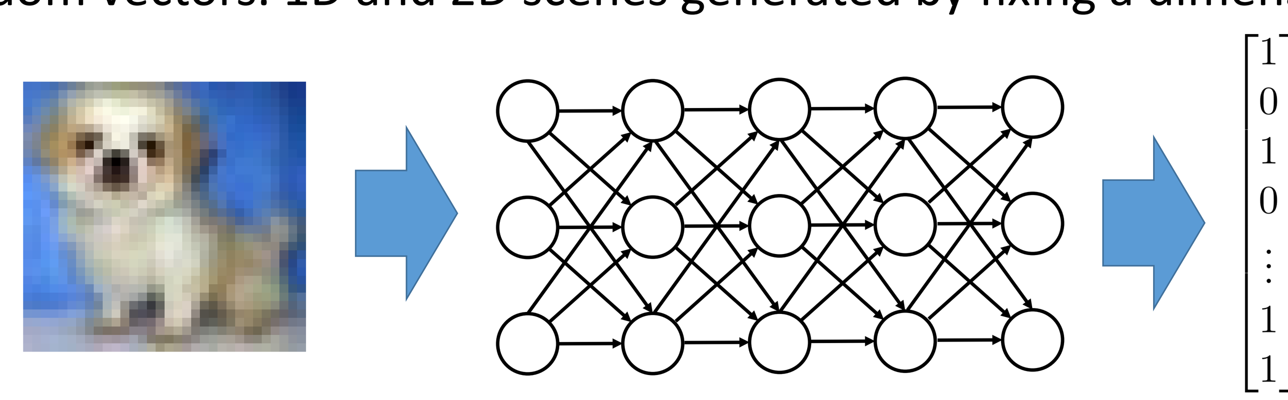


Fig 5: Deep Supervised Hashing (DSH) modified to generate 4,500 dimensional binary vectors that "summarize" an image [5]. The network is trained to assign different vectors to different classes.

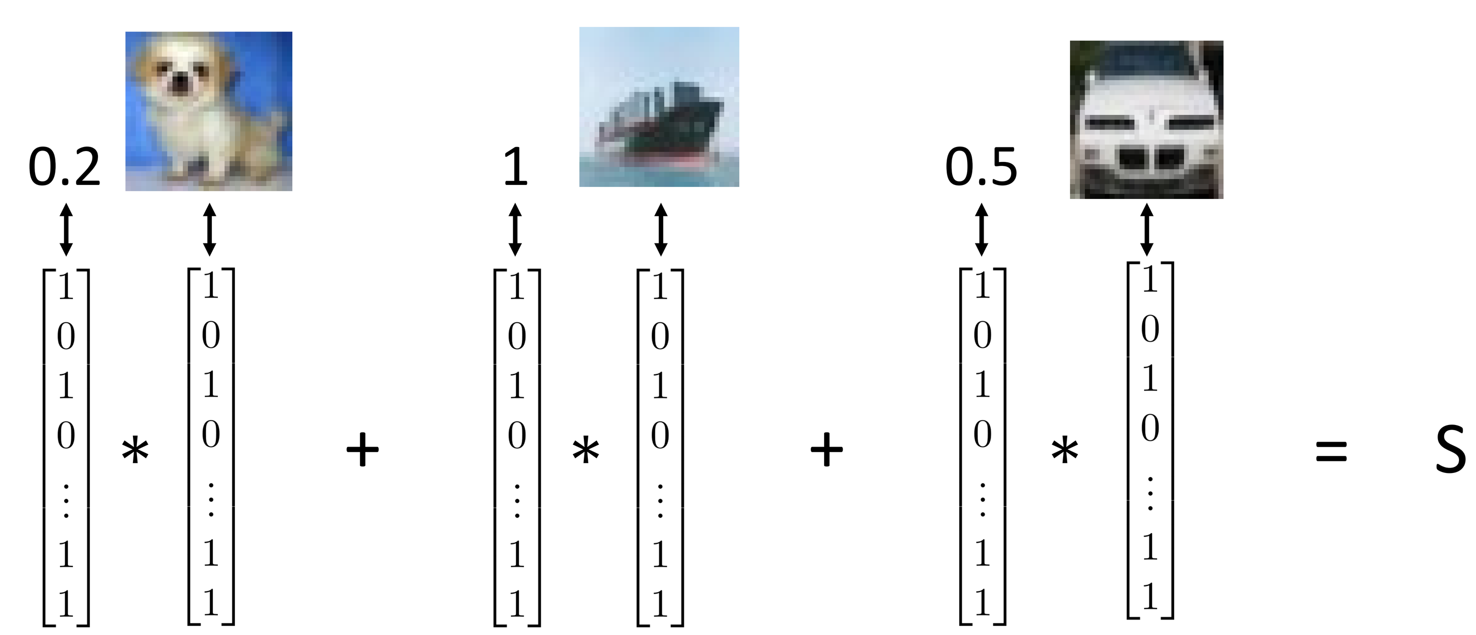


Fig 6: Image vectors and vectors representing quantized distance are "bound" (*) and "bundled" (+) to form a vector S representing all the objects and their relative distances without significant loss of information.

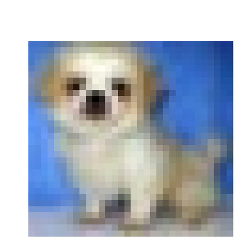
S *  ≈ 0.2

Fig 7: By properties of HD computing, scene vector "S" bound (*) again by an image vector representing "dog" results in a vector approximately equal to the one that represents the associated distance "0.2". Recall within 1 cm is considered correct.

Results

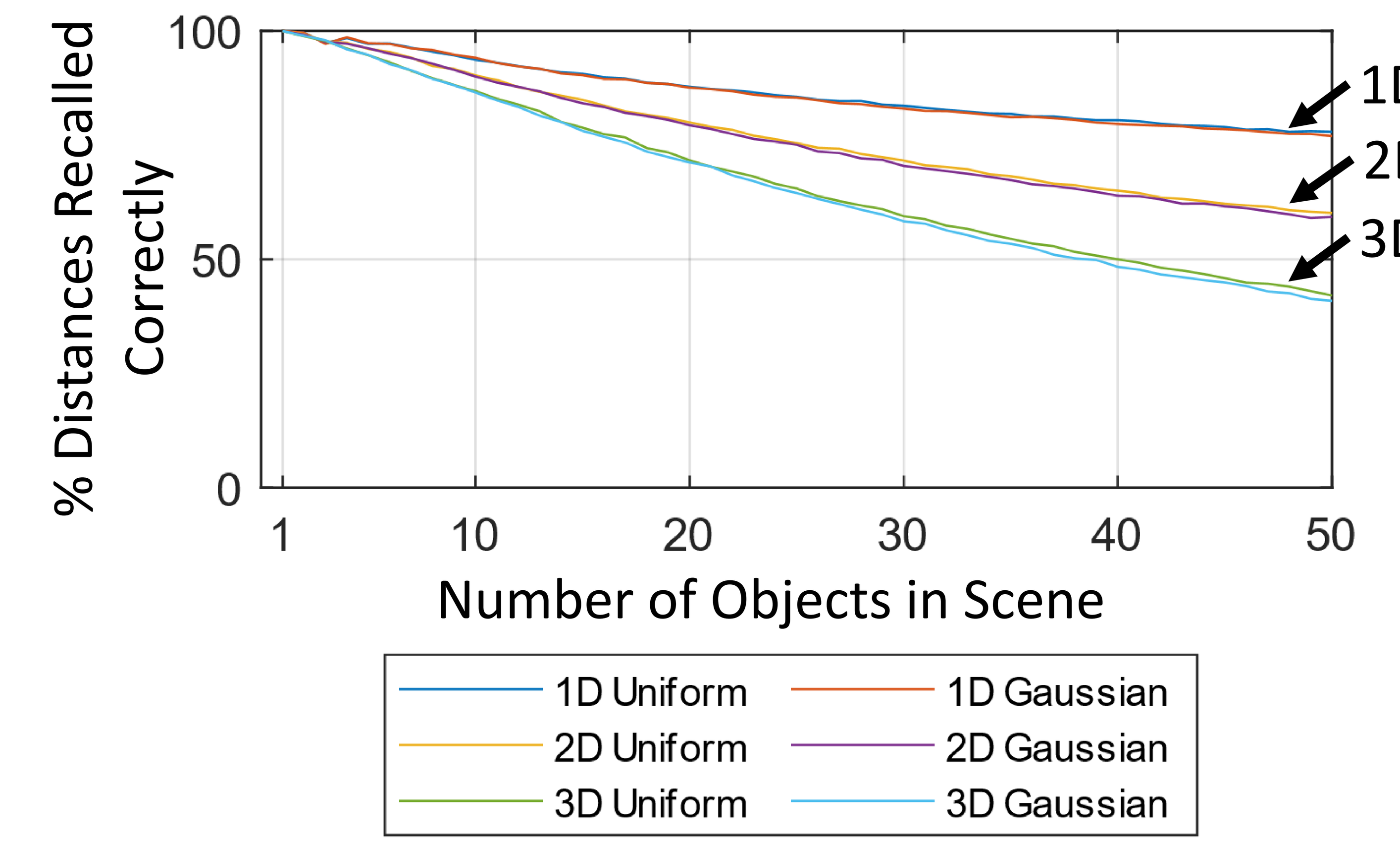


Fig 8: Number of objects in scene vector "S" versus percent of object distances correctly recalled to 1 cm tolerance. Different lines show different positional distributions used for the generated scenes. Results for scenes with N objects were calculated using 10,000 / N scenes. A reasonable recall accuracy target is 80% at which point this implementation can recall 14 objects.

Conclusions

- Encoded 14 objects and their locations with 80% recall accuracy
- 1D and 2D distributions increased recall accuracy
- Gaussian distribution resulted in marginally worse recall
- We will be pursuing HD computing for explainable and efficient robot navigation

Discussion

- Image hashing network could significantly impact accuracy
  - More experiments should be done with alternative networks
- Simulating scenes from a single viewpoint did not capture effect of differing viewpoints of the same object on the image hashing networks representation
- Because random vectors were used for to represent distance, recall accuracy can not degrade gracefully

References/Acknowledgements

[1] Tong, Q., Peiliang, L., & Shaojie, S. (2018). VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator. IEEE Transactions on Robotics, 34(4), 1–17.

[2] Dasgupta, S., Sheehan, T. C., Stevens, C. F., & Navlakha, S. (2018). A neural data structure for novelty detection. Proceedings of the National Academy of Sciences of the United States of America, 115(51), 13093–13098.

[3] Kanerva, P. (2009). Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors. Cognitive Computation, 1(2), 139–159.

[4] Krizhevsky, A. (2009). Learning Multiple Layers of Features from Tiny Images.

[5] DSH: Liu, H., Wang, R., Shan, S., & Chen, X. (2016). Deep Supervised Hashing for Fast Image Retrieval. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-December, 2064–2072.

Levi Burner's contribution to this research was supported by NSF award DGE-1632976