Towards cognitive automotive environment modelling based on vector representations

Florian Mirus, Terrence C. Stewart, Jörg Conradt
florian.mirus@bmwgroup.com, tcstewar@uwaterloo.ca, conradt@tum.de

Introduction

We outline a first step in the direction of a cognitive approach to automotive environment modelling based on Vector Symbolic Architectures (VSAs) [3]. We build a vector description of the current scene from high-level object-lists provided by individual sensor units. This rather generic representation can be applied to various different tasks like driving context classification, anomaly detection, behavior analysis and prediction with little modifications to the representation itself. Furthermore, VSAs are suitable as inputs to Spiking Neural Networks (SNNs) [2], which support efficient learning algorithms and future deployment on dedicated neuromorphic hardware. In this work, we demonstrate the general applicability of our approach on the example task of driving context classification. We aim to apply our approach to other tasks in future work.

Vector Symbolic Architectures

Vector Symbolic Architectures (VSAs) is a term coined by Ross W. Gayler [3] to cover a family of modelling approaches that represent symbols and structures by mapping them to (high-dimensional) vectors. Good representations preserve conceptual similarity by posing vectors encoding similar concepts in relative proximity within the vector space (cf. fig. 1). The core components of a VSA are a measure of similarity (here the dot product) and typically two algebraic operations, namely superposition $\oplus$ and binding $\otimes$, which create a vector similar resp. highly dissimilar to both input vectors. These algebraic operations allow for structured vector representations like

$$\mathbf{CAR} = \mathbf{TYPE}\otimes\mathbf{VEHICLE} + \mathbf{WHEELS}\otimes\mathbf{FOUR} + \mathbf{ACTUATION}\otimes\mathbf{MOTOR} + \dots$$

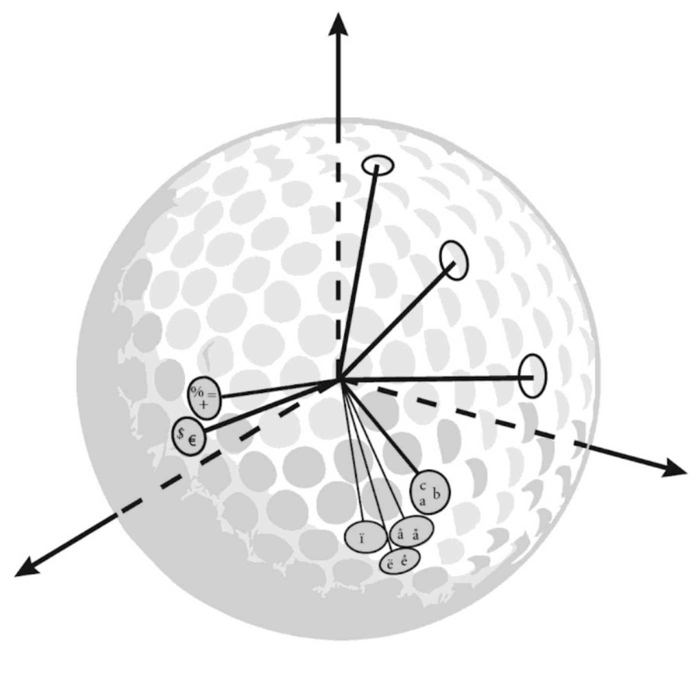

Figure 1: "Conceptual golf ball" depicting the idea of semantic vectors. Source [2].

Implementation

For highly automated vehicles, *environment perception* happens through a variety of different sensors [1] providing *preprocessed data* in the form of object-lists or raw sensory data. This observed information needs to be collected and combined into a central model of the environment, which is the basis for further reasoning and decisions. We build a vector representation from this preprocessed data through *vectorization* of sensory data and high-level object lists. In our work, elementary vectors - picked randomly from the real-valued unit sphere - take the role of atomic ingredients, i.e. all structured representations are composed from combinations of elementary vectors using the VSA's algebraic operations. The *input data* used for training and evaluating the system is real-world data gathered during test drives in the region of Munich, Germany. We manually labeled our data sets into three driving context classes: *city*, *interurban*, *highway*. We encapsulate three types of information in our vector-based scene representation: ego-vehicle dynamics , dynamic objects and traffic signs provided as preprocessed object-lists by a camera sensor system (cf. fig.2). We calculate the scene vector as

$$S = \underbrace{v\cdot\mathbf{VELOCITY} + \lambda(a_x,a_y)\otimes\mathbf{ACCELERATION} + s\cdot\mathbf{STEER}}_{\text{ego-vehicle dynamics}} + \sum_j \underbrace{(\gamma_j\cdot\mathbf{SIGN}_j)}_{\text{traffic signs}}$$
$$+\sum_i \underbrace{(\mathbf{ID}_i + \mathbf{ID}_i\otimes\mathbf{POSITION}\otimes\lambda(p_x,p_y) + \mathbf{ID}_i\otimes\mathbf{ORIENTATION}\otimes\mathbf{O}_i)}_{\text{dynamic objects}},$$

with bold items being elementary (ID) vectors, v and s denote scalar values, γ_j are decaying factors and $\lambda:\mathbb{R}^2\longrightarrow\mathbb{R}^D$ denotes an encoding function. We use *neural networks* implemented in Nengo (fig. 3) and Keras (fig. 4) for predictions based on the current scene vector S and compared them to the performance of human subjects (cf. fig. 5).

System Architecture

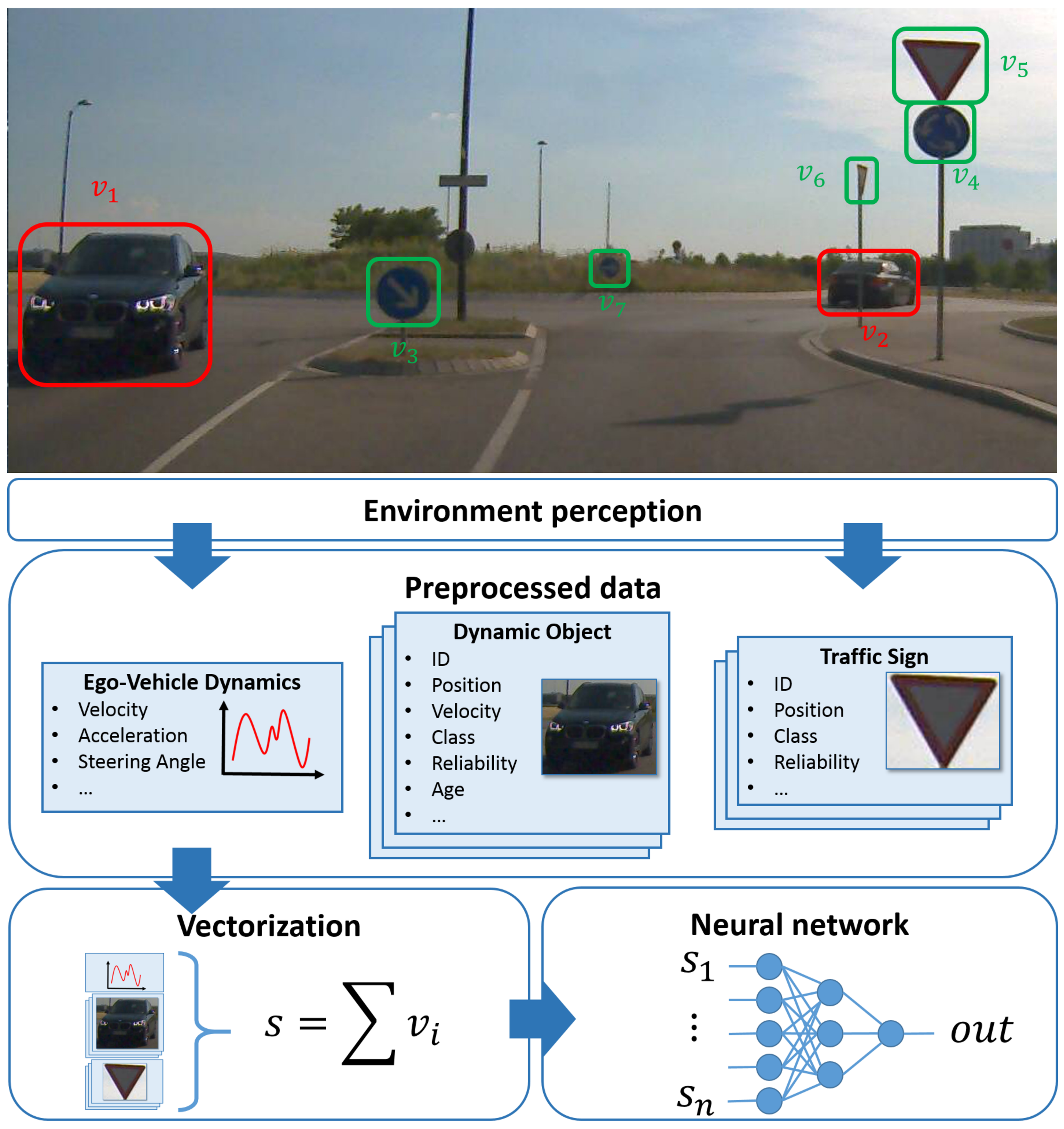

Figure 2: Schematic system overview with one example scene

Results

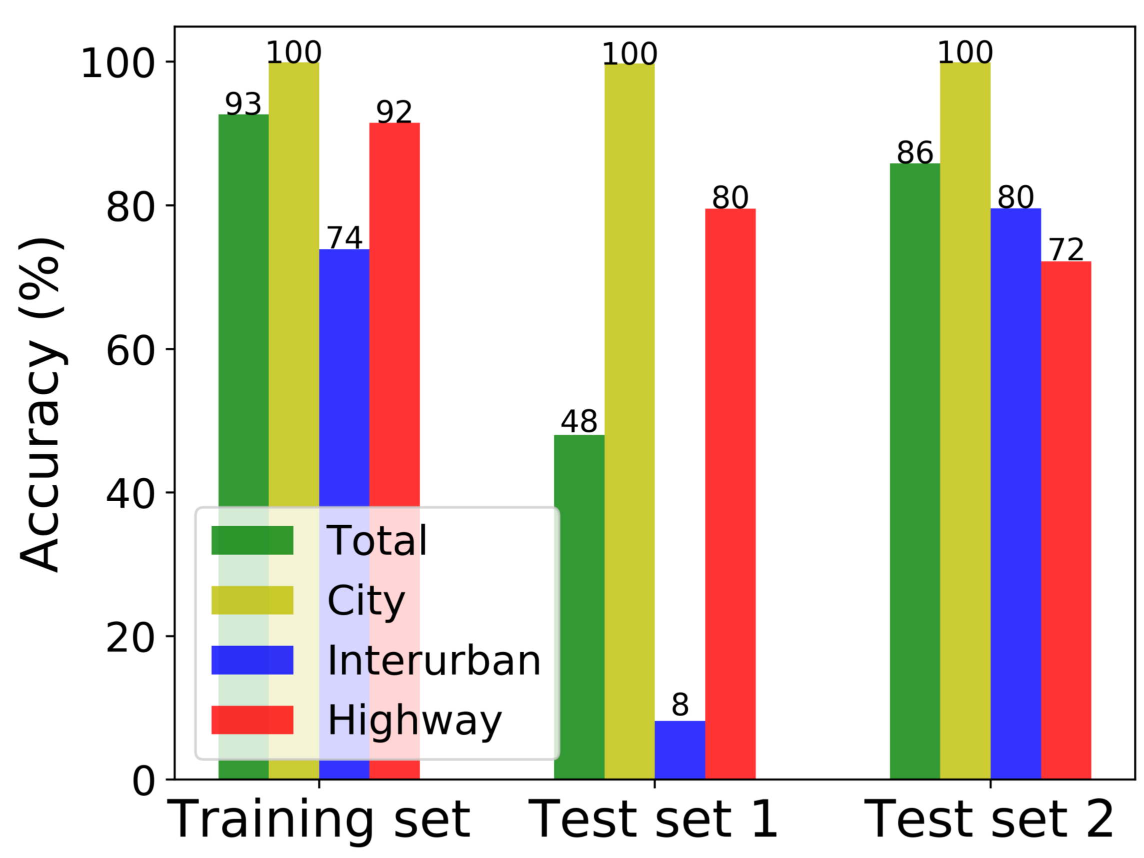

Figure 3: Nengo network

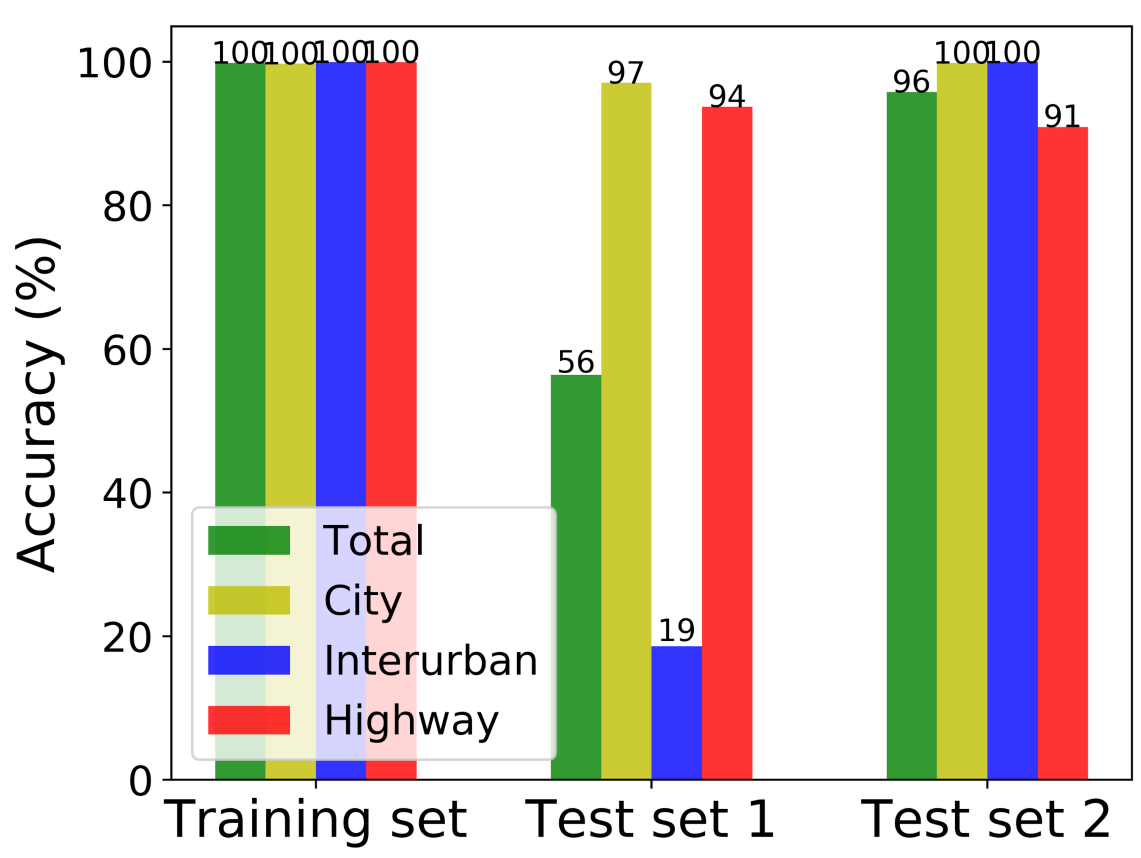

Figure 4: Keras network

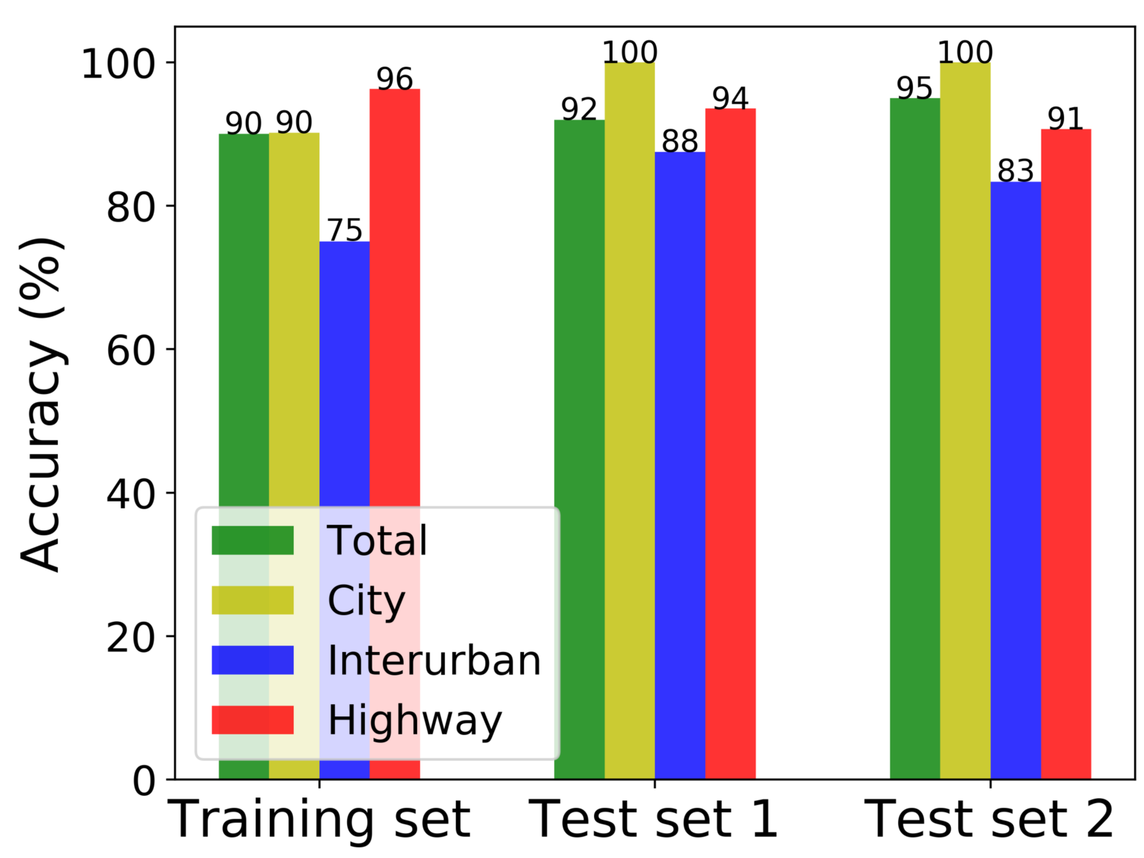

Figure 5: Human reference

References

[1] M. Aeberhard et al. "Experience, Results and Lessons Learned from Automated Driving on Germany's Highways". In: *IEEE Intelligent Transportation Systems Magazine* 7.1 (2015-Spring), pp. 42–57. ISSN: 1939-1390. DOI: 10.1109/MITS.2014.2360306.

[2] C. Eliasmith. *How to build a brain: A neural architecture for biological cognition.* New York, NY: Oxford University Press, 2013.

[3] R. Gayler. "Vector Symbolic Architectures answer Jackendoff's challenges for cognitive neuroscience". In: *ICCS/ASCS International Conference on Cognitive Science.* Ed. by P. Slezak. University of New South Wales. CogPrints, 2003, pp. 133–138.