Biologically Inspired Online Learning of Visual Autonomous Driving

Kristoffer Öfjäll
kristoffer.ofjall@liu.se

Michael Felsberg
michael.felsberg@liu.se

Computer Vision Laboratory
Department of Electrical Engineering
Linköping University
Linköping, Sweden

While autonomously driving systems accumulate more and more sensors as well as highly specialized visual features and engineered solutions, the human visual system provides evidence that visual input and simple low level image features are sufficient for successful driving. In this paper we propose extensions (non-linear update and coherence weighting) to one of the simplest biologically inspired learning schemes (Hebbian learning). We show that this is sufficient for online learning of visual autonomous driving, where the system learns to directly map low level image features to control signals. After the initial training period, the system seamlessly continues autonomously. This extended Hebbian algorithm, qHebb, has constant bounds on time and memory complexity for training and evaluation, independent of the number of training samples presented to the system. Further, the proposed algorithm compares favorably to state of the art engineered batch learning algorithms in a visual head pose prediction challenge, where the algorithms can be more thoroughly evaluated.

The input layer of the system consists of a single gray-scale camera and a generic, holistic representation of the whole visual field, using visual Gist [4]. No information regarding what kind of track or what kind of visual features that define the track is provided in advance. The system is supposed to learn features of the track together with correct driving behavior from the visual Gist features and the manual control signals provided during the initial training phase, see Fig. 1. Table 1 summarizes previous and present approaches to this task.

The proposed approach is based on the channel representation [2] and associative learning [3]. The channel representation of a scalar entity is a coefficient vector similar to a soft histogram (Fig. 2). A corresponding representation in biological systems is the population coding of e.g. orientation in the visual field. An associative mapping $\mathbf{y} = \mathbf{Cx}$ is learned, relating the channel representation, $\mathbf{x}$, of each image from the camera to the channel representation, $\mathbf{y}$, of the steering signal. Although the mapping is linear in the channel domain, non-linear relations can be represented between the original domains. Our two main contributions are an online learning rule for $\mathbf{C}$ with decoupled learning and forgetting rates, and a weighted variant of $\mathbf{y} = \mathbf{Cx}$, where each coefficient in $\mathbf{x}$ is weighted with the specificity with which it predicts the control signal (encoded in $\mathbf{y}$).

Method	Online Driving	Training Data Proc. Speed
CN [5]	No	Days (batch)
RFR [1]	Yes	Hours (batch)
Assoc. Hebb	No	Video rate (online)
Proposed	Yes	Video rate (online)

Table 1: Summary of approaches for visual autonomous driving, including Random Forest Regression (RFR) and Convolutional Networks (CN).

[1] Liam Ellis, Nicolas Pugeault, Kristoffer Öfjäll, Johan Hedborg, Richard Bowden, and Michael Felsberg. Autonomous navigation and sign detector learning. In *Robot Vision (WORV), 2013 IEEE Workshop on*, pages 144–151. IEEE, 2013.

[2] G. H. Granlund. An Associative Perception-Action Structure Using a Localized Space Variant Information Representation. In *Proceedings of Algebraic Frames for the Perception-Action Cycle (AFPAC)*, Germany, September 2000.

[3] Björn Johansson. *Low Level Operations and Learning in Computer Vision*. PhD thesis, Linköping University, Computer Vision, The Institute of Technology, 2004.

[4] A. Oliva and A. Torralba. Modeling the shape of the scene: a holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175, 2001.

[5] Maria Schmiterlöw. Autonomous path following using convolutional networks. Master's thesis, Linköping University, Computer Vision, The Institute of Technology, 2012.
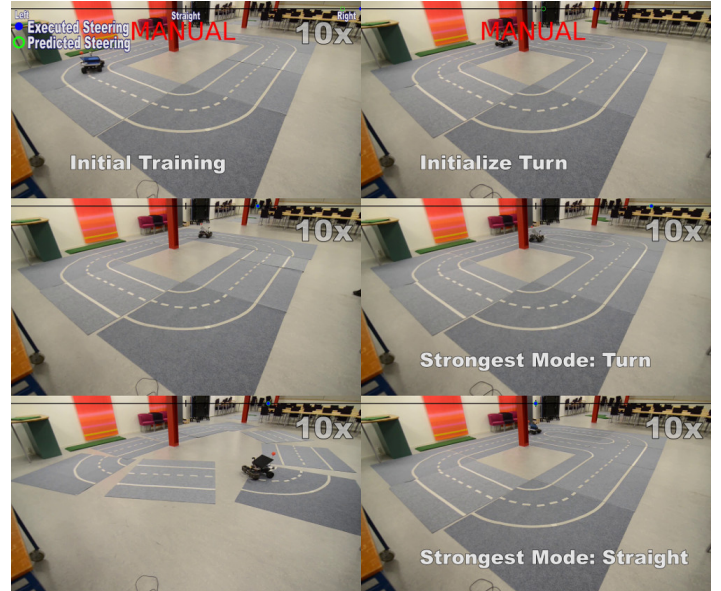
Figure 1: An autonomous vehicle learns a mapping from general low level image features to steering signal from demonstration by remotely operating the vehicle during parts of the first lap. When the remote override is released, the vehicle seamlessly continues autonomously. After the training period, the system showed robustness towards reconfiguration of the track (left column). After introducing an intersection, the vehicle can be forced to take the other way by a short application of manual control. After this new training data is acquired, both modes, *going straight* and *turning*, are present in the system. When the vehicle reaches the intersection again, it either goes straight or turns depending on which mode happen to get a stronger response (right column). However, the modes are never mixed, which would be noticed by the vehicle making half a turn at the intersection. The figure shows a selection of frames from the supplementary video available at *http://users.isy.liu.se/cvl/ofjall/bmvc2014.mp4*.
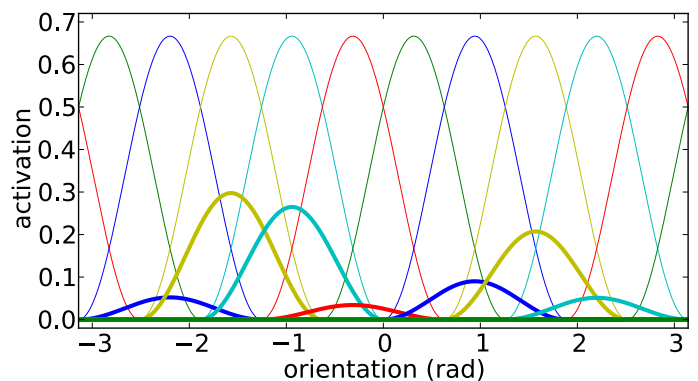

Figure 2: Illustration of a channel representation of a distribution of orientation data. The thin plots indicate the kernel functions (the channels) and the bold lines illustrate the corresponding channel coefficients as weighted kernel functions. The represented distribution has two modes.