ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

Adaptive Velocity Tuning for Visual Motion Estimation

Volker Willert[1] and Julian Eggert[2]

1- Darmstadt University of Technology
Institute of Automatic Control, Control Theory and Robotics Lab
Landgraf-Georg-Str. 4, D-64283 Darmstadt, Germany

2- Honda Research Institute Europe GmbH
Carl-Legien-Str. 30, D-63073 Offenbach, Germany

Abstract.
In the brain, both neural processing dynamics as well as the perceptual interpretation of a stimulus can depend on sensory history. The underlying principle is a sensory adaptation to the statistics of the input collected over a certain amount of time, allowing the system to tune its detectors, e.g. by improving the sampling of the input space. Here we show how a generative formulation for the problem of visual motion estimation leads to an online adaptation of velocity tuning that is compatible with physiological sensory adaptation and observed perceptual effects.

1 Introduction

Biological sensory systems adapt to the history of the sensory input over a variety of timescales and several types of modalities. The adaptation is especially prominent after prolongued exposure to visual stimuli of a particular type, like orientation, texture, contrast or motion, and leads to a systematic bias in the perception or the estimation of the stimulus variables.

The function of this adaptation has been studied for a long time. In many cases, it leads to illusory aftereffects that come along with an increased discrimination performance, like in the cases of the orientation tilt illusion [2] or the waterfall illusion [1]. In [4], it is discussed to which extent the findings are consistent with a decoding state after the adapting sensory units that is aware / is not aware of the sensory adaptation, with the conclusion that a fixed, "unaware" decoder can account for the measured effects. However, this does not explain *how and why* the detectors adapt, but only the influence of an adaptation on a subsequent internal evaluation.

In this paper, we directly look at a functional explanation for the adaptation dynamics of the sensory units. We assume that the benefits of the adaptation would be a temporarily improved sampling of the sensory input space. In a generative setting, the target of the system is to maximize the probability of explaining the input. In the particular case of motion estimation, this optimization leads in a straightforward way to a sensory adaptation that concentrates a set of motion detectors on the relevant velocities of a scene, i.e., on the statistics of the current stimulus. This idea is visualized in Fig. 1. In consequence, this

ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

leads to attracting shifts in the speed and direction tuning of motion detectors on a short timescale, similar to those found experimentally [5], [6].

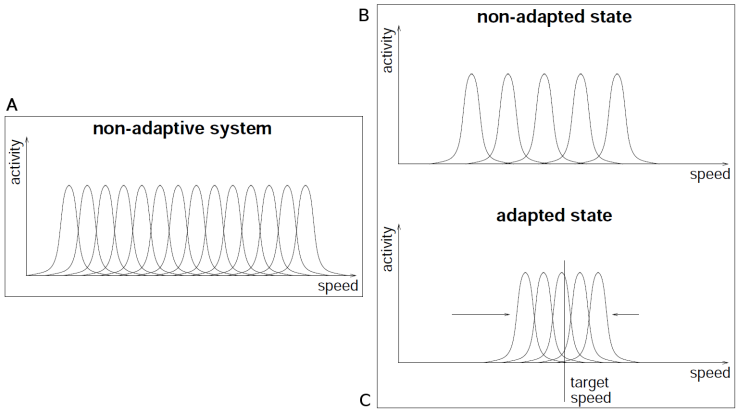

Fig. 1: (A) Set of velocity-tuned neurons with speed tuning curves densely sampling the entire velocity space. (B) Velocity preferences coarsely distributed in velocity space but (C) being smartly adaptive to be able to cluster around some relevant velocities (adapted from [3]). See text for further details.

2 The Probabilistic Filter Model for Motion Estimation

As a basis for our model, we use a modification of a probabilistic filter approach for motion estimation presented in [7]. We assume that the overall system describes a moving input by using velocity distributions $P(\mathbf{v}_{\mathbf{x}}^t)$ at retinal positions $\mathbf{x}$ for velocities $\mathbf{v}$ in a continuous velocity space. However, the velocity distributions are approximated by a limited number of discrete velocity samples at the sampling points $\mathbf{h} \in \mathbf{H}$. (In a sense, the motion field is sampled by a set of motion detectors each characterized by a fixed position $\mathbf{x}$ and a tuning velocity $\mathbf{h} \in \mathbf{H}$.)

For a generative model that takes advantage of the sensory history, we express the posterior of the entire estimated velocity field in form of a probabilistic filter model. In a Bayesian manner, we take the last velocity estimation (the previous posterior from time t) $P(\mathbf{V}^t|\mathbf{Y}^{1:t};\mathbf{H})$ to calculate the next expected velocity estimation (the predictive prior at time $t' = t + \Delta t$) $P(\mathbf{V}^{t'}|\mathbf{Y}^{1:t};\mathbf{H})$. This is combined with the likelihood that the current sensory input $\mathbf{Y}^t := \{\mathbf{I}^{t'}, \mathbf{I}^t\}$ (with the future and the current input images $\mathbf{I}^{t'}$ and $\mathbf{I}^t$) can be explained given a velocity estimation, to gain the next posterior via

$$P(\mathbf{V}^{t'}|\mathbf{Y}^{1:t'};\mathbf{H}) \propto P(\mathbf{Y}^{t'}|\mathbf{V}^{t'};\mathbf{H})\, P(\mathbf{V}^{t'}|\mathbf{Y}^{1:t};\mathbf{H}) \tag{1}$$

The likelihood can be expressed via

$$\begin{aligned} P(\mathbf{Y}^t|\mathbf{V}^t;\mathbf{H}) &= \prod_{\mathbf{x}} \ell(\mathbf{Y}^t, \mathbf{v}_{\mathbf{x}}^t;\mathbf{H}) \\ \ell(\mathbf{Y}^t, \mathbf{v}_{\mathbf{x}}^t;\mathbf{H}) &= \sum_{\mathbf{h}\in\mathbf{H}} \delta(\mathbf{v}_{\mathbf{x}}^t - \mathbf{h})\ell(\mathbf{Y}^t, \mathbf{v}_{\mathbf{x}}^t) \\ \ell(\mathbf{Y}^t, \mathbf{v}_{\mathbf{x}}^t) &= f_\ell(\mathbf{I}^{t'}_{\mathbf{x}+\mathbf{v}_{\mathbf{x}}^t}, \mathbf{I}^t_{\mathbf{x}}; \theta_\ell) \end{aligned} \tag{2}$$

with likelihood-specific fixed parameters θ_ℓ. In other words, a) the overall likelihood factorizes into position-specific likelihoods, with, b) the velocity parameterization at each position being restricted to the sampling velocities $\mathbf{h} \in \mathbf{H}$, and c) the local likelihood is gained by the patchwise comparison of the current and the future input images around positions compatible with the velocity hypothesis $\mathbf{v}_{\mathbf{x}}^t$.

The transition to get the predictive prior is done by

$$P(\mathbf{V}^{t'}|\mathbf{Y}^{1:t};\mathbf{H}) = \sum_{\mathbf{V}^t} P(\mathbf{V}^{t'}|\mathbf{V}^t;\mathbf{H})P(\mathbf{V}^t|\mathbf{Y}^{1:t};\mathbf{H}) \tag{3}$$

$$\begin{aligned} P(\mathbf{V}^{t'}|\mathbf{V}^t;\mathbf{H}) &= \prod_{\mathbf{x}} \phi(\mathbf{v}_{\mathbf{x}}^{t'}, \mathbf{V}^t;\mathbf{H}) \\ \phi(\mathbf{v}_{\mathbf{x}}^{t'}, \mathbf{V}^t;\mathbf{H}) &= \sum_{\mathbf{h}\in\mathbf{H}} \delta(\mathbf{v}_{\mathbf{x}}^{t'} - \mathbf{h})\phi(\mathbf{v}_{\mathbf{x}}^{t'}, \mathbf{V}^t) \\ \phi(\mathbf{v}_{\mathbf{x}}^{t'}, \mathbf{V}^t) &= \sum_{\mathbf{x}'} f_x(\mathbf{x}', \mathbf{x} - \mathbf{v}_{\mathbf{x}}^{t'}\Delta t; \theta_x) f_t(\mathbf{v}_{\mathbf{x}}^{t'}, \mathbf{v}_{\mathbf{x}'}^t; \theta_t) \end{aligned} \tag{4}$$

with prediction-specific parameters θ_x, θ_t. Here, we have again assumed that the predictive prior factorizes spatially and that only the sampling velocities $\mathbf{h} \in \mathbf{H}$ are allowed. The functions f_x and f_t express the "lateral" propagation of information from the last to the current timestep (see [7] for more detailed information).

Inserting 3 into 1 leads to the posterior

$$P(\mathbf{V}^{t'}|\mathbf{Y}^{1:t'};\mathbf{H}) \propto P(\mathbf{Y}^{t'}|\mathbf{V}^{t'};\mathbf{H}) \sum_{\mathbf{V}^t} P(\mathbf{V}^{t'}|\mathbf{V}^t;\mathbf{H})P(\mathbf{V}^t|\mathbf{Y}^{1:t};\mathbf{H}) \tag{5}$$

which factorizes spatially, so that using 2 and 4 we get

$$\begin{aligned} P(\mathbf{V}^{t'}|\mathbf{Y}^{1:t'};\mathbf{H}) &= \prod_{\mathbf{x}} P(\mathbf{v}_{\mathbf{x}}^{t'}|\mathbf{Y}^{1:t'};\mathbf{H}) \\ P(\mathbf{v}_{\mathbf{x}}^{t'}|\mathbf{Y}^{1:t'};\mathbf{H}) &\propto \sum_{\mathbf{h}'\in\mathbf{H}} \delta(\mathbf{v}_{\mathbf{x}}^{t'} - \mathbf{h}')\ell(\mathbf{Y}^{t'}, \mathbf{v}_{\mathbf{x}}^{t'}) \sum_{\mathbf{x}'} f_x(\mathbf{x}', \mathbf{x} - \mathbf{v}_{\mathbf{x}}^{t'}\Delta t; \theta_x) \times \\ &\quad \sum_{\mathbf{h}\in\mathbf{H}} \delta(\mathbf{v}_{\mathbf{x}'}^t - \mathbf{h}) \sum_{\mathbf{v}_{\mathbf{x}'}^t} f_t(\mathbf{v}_{\mathbf{x}}^{t'}, \mathbf{v}_{\mathbf{x}'}^t; \theta_t) P(\mathbf{v}_{\mathbf{x}'}^t|\mathbf{Y}^{1:t}) \end{aligned} \tag{6}$$

Eq. 6 is the probabilistic filter description for the transition $t \to t'$ of the velocity field distribution based on a common, discrete set of velocity samples $\mathbf{h} \in \mathbf{H}$ used at all positions $\mathbf{x}$.

ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

3 Online Adaptation of the Velocity Tuning

Now we consider the velocity sampling points $\mathbf{h}$ to be adaptive at every time step. They are considered as parameters of the velocity estimation process, that can be optimized via an approximate Expectation-Maximization (EM) mechanism.

In the EM of our system, the target is to maximize the log probability $\ln P(\mathbf{Y}^{1:t}|\mathbf{H})$ of the data $\mathbf{Y}^{1:t}$ given the parameters $\mathbf{H}$. W.l.o.g. and for derivation purposes, in the following we will assume that the same velocity sampling points $\mathbf{h}$ are used over the entire visual field. With

$$P(\mathbf{Y}^{1:t}|\mathbf{H}) = \sum_{\mathbf{V}^t} P(\mathbf{Y}^{1:t}|\mathbf{V}^t;\mathbf{H})P(\mathbf{V}^t|\mathbf{H}) \tag{7}$$

and the EM formalism that we obtain successive parameters $\mathbf{H}^t$ (i.e., the set of sampling points in velocity space that maximizes the probability of the data) by maximizing

$$Q_{\mathbf{H}^t}(\mathbf{H}) = \sum_{\mathbf{V}^t} P(\mathbf{V}^t|\mathbf{Y}^{1:t};\mathbf{H}^t)\, \ln P(\mathbf{V}^t, \mathbf{Y}^{1:t}|\mathbf{H}) \tag{8}$$

in the way

$$\mathbf{H}^{t'} = \mathrm{argmax}_{\mathbf{H}} Q_{\mathbf{H}^t}(\mathbf{H}) \qquad . \tag{9}$$

Using the probabilistic filter results 5 and 6 in 8, then calculating the derivative with respect to $\mathbf{H}$ and setting it to zero for finding the maximum according to 9, we get

$$\begin{aligned} \mathbf{h}^{t'} &\approx \begin{bmatrix} \sum_{\mathbf{x}} \alpha (I^t_{x,\mathbf{x}})^2 & \sum_{\mathbf{x}} \alpha I^t_{x,\mathbf{x}} I^t_{y,\mathbf{x}} \\ \sum_{\mathbf{x}} \alpha I^t_{x,\mathbf{x}} I^t_{y,\mathbf{x}} & \sum_{\mathbf{x}} \alpha (I^t_{y,\mathbf{x}})^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{\mathbf{x}} \alpha I^t_{t,\mathbf{x}} I^t_{x,\mathbf{x}} \\ \sum_{\mathbf{x}} \alpha I^t_{t,\mathbf{x}} I^t_{y,\mathbf{x}} \end{bmatrix} \\ \alpha &= P(\mathbf{v}^t_{\mathbf{x}} = \mathbf{h}^t | \mathbf{Y}^{1:t}) \end{aligned} \tag{10}$$

for the velocity sampling points $\mathbf{h}^{t'} \in \mathbf{H}^{t'}$.

To arrive at eqs. 10, we have linearized the input images. The $I_{x,\mathbf{x}}$, $I_{y,\mathbf{x}}$ and $I_{t,\mathbf{x}}$ denote spatial (in x-direction), spatial (in y-direction) and temporal derivatives, respectively, taken at position $\mathbf{x}$.

With 10, we have gained an explicit iterative description for the calculation of the next optimal sampling points $\mathbf{H}^{t'}$ given the current ones $\mathbf{H}^t$. It takes into account the spatial statistics ($\sum_{\mathbf{x}} ...$) of the current posterior distribution $P(\mathbf{v}^t_{\mathbf{x}} = \mathbf{h}^t|\mathbf{Y}^{1:t})$ to adapt the velocity tuning $\mathbf{h}^{t'}$ of the veclocity detection units.

What is the effect of this adaptation? Consider the probabilistic filter from 6 without velocity tuning, i.e., keeping the $\mathbf{H}$ fixed. In this case, it is sensible to choose the $\mathbf{h}$ to densely sample the entire velocity space, so that we have a set of velocity-detecting units / neurons with tuning curves which regularly and completely cover the possible range of inputs. With the adaptation according to 10, the tuning curves of the veclocity-detecting units cluster around the relevant velocities available in the input image, providing a finer sampling around these

velocities. Fig. 1 exhibits this idea. The bottom line is that the system will adjust its tuning curves as soon as relevant velocities are measured, e.g. concentrating around the velocity of a target object that appears in the visual field.

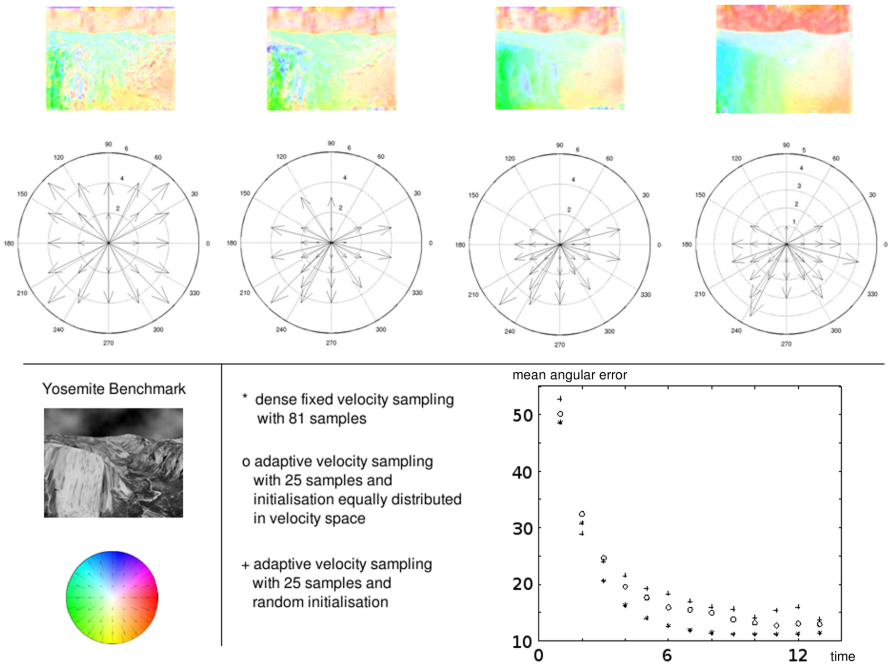

Fig. 2: Velocity estimation results of the system with adaptive velocity tuning. Within about 10 frames, the system adapts its velocity detectors to the input statistics. See text for further details.

The results of our system are shown in Fig. 2. We compared a system with a fixed array of 81 (9×9) velocity detectors at each position $\mathbf{x}$ densely covering the velocity space with an adaptive system of only 25 velocity detectors. The fixed system improves its velocity estimation by spatiotemporal filtering over consecutive video frames (here shown: the first 13 frames). In the upper row, we show the adaptive system, which exhibits very similar spatiotemporal filtering properties, shown for the Yosemite sequence in the color-coded inlets representing the motion flow. The adaptive system starts with 25 velocity sampling points equally distributed in velocity space (left circular plot). However, since in the sequence the divergent flow to the bottom dominates, within a few frames the system adapts to better cover this velocity region (right circular plot). The estimated motion flow (colored images) is quantitatively comparable to that of the 81 fixed velocity sampling points, implying a considerable reduction of resources and a concentration on the "right" velocity tunings. At the bottom right we show the mean angular error for 3 different settings: a) the fixed velocity samp-

ing case, b) the adaptive case with equally distributed initialization and c) the adaptive case with random initialization. It can be seen that, within about 10 frames, the system converges to small angular errors and that the results from the adaptive system with only 25 velocity sampling points are comparable with the fixed system with 81 sampling points.

4 Conclusions

In this work, we have presented a system that adapts the tuning of its motion detectors based on a generative framework for spatiotemporal filtering and an EM-like optimization of the velocity sampling points that maximizes the probability that the current input can be described with the estimated motion field. We have found that such a system is able to rapidly adjust (within a few frames) the tuning of its velocity detectors, allowing to achieve comparable results with much less detection units.

Evidence that this is also happening on very short (msec) [6] to short (sec) [5] timescales comes from biological findings on the tuning of velocity-detecting neurons. The neural response immediately adapts to motion stimuli which results in attracting shifts in neural speed and direction tuning, meaning that the tuning curves after the adaptation are attracted towards the true target velocities. This type of adaptation is consistent with the findings presented in this model, and would provide a functional explanation of the fast adaptation in the sense of an optimal sampling of the velocity space based on the velocity statistics of the past stimulus.

References

[1] R. Addams, An account of a peculiar optical phenomenon seen after having looked at a moving body, *London and Edinburgh Philosophical Magazine and Journal of Science*, 5:373-374, 1834

[2] J.J. Gibson and Radner, Adaptation, after-effect and contrast in the perception of tilted lines, *Journal of Experimental Psychology*, 20:453-467, 1937

[3] U. Hillenbrand, Spatiotemporal adaptation in the corticogeniculate loop. *PhD Thesis of the Technical University of Munich*, 2001

[4] P. Series, A.A. Stocker and E.P. Simoncelli, Is the Homunculus "Aware" of Sensory Adaptation? *Neural Computation*, 21:3271-3304, 2009

[5] A. Kohn and J.A. Movshon, Adaptation changes the direction tuning of macaque MT neurons, *Nature Neuroscience*, 7(7):764-772, 2004

[6] A. Schlack, B. Krekelberg and T.D. Albright, Recent History of Stimulus Speeds Affects the Speed Tuning of Neurons in Area MT, *The Journal of Neuroscience*, 27(41):11009-11018, 2007

[7] V. Willert and J. Eggert, A Stochastic Dynamical System for Optical Flow Estimation. *In proceedings of the 12th IEEE International Conference on Computer Vision, 4th Workshop on Dynamical Vision*, 711-718, 2009