Modeling Short-Term Adaptation Processes of Visual Motion Detectors

Volker Willert, Julian Eggert

2011

Preprint:

This is an accepted article published in Neurocomputing. The final authenticated version is available online at: https://doi.org/[DOI not available]

Modeling Short-Term Adaptation Processes of Visual Motion Detectors

Volker Willert[a], Julian Eggert[b]

[a]*Darmstadt University of Technology, Institute of Automatic Control, Control Theory and Robotics Lab, Landgraf-Georg-Str. 4, D-64283 Darmstadt, Germany*
[b]*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 some timespan, allowing the system to tune its detectors, e.g. by better sampling the input space and adjusting the response. Here, we show how a model for adaptation in visual motion processing can be set up from first principles that uses a generative formulation and casts the problem of adaptation in terms of optimal estimation over time. The model leads to an online adaptation of velocity tuning curves, inducing shifts in the velocity tuning and changes in the tuning curve widths that are compatible with observations from physiological experiments on macaque MT neurons. We also show how such an adaptation leads to a greater computational efficiency by a better sampling of the velocity space, requiring less motion detectors to achieve a desired level of estimation accuracy.

Keywords: motion estimation, velocity tuning, adaptive system, stochastic dynamical system, online optimization, maximum posterior solution

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 [1] or the waterfall illusion [2]. In [3], 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 effect that an adaptation has on a subsequent internal evaluation.

Email addresses: vwillert@rtr.tu-darmstadt.de (Volker Willert), julian.eggert@honda-ri.de (Julian Eggert)

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, which would lead to a better performance in terms of sensory estimation and/or discrimination accuracy.

Let us consider the case of basic motion estimation. For a biologically consistent setting of a model for low-level motion processing, a discrete number of motion detectors (e.g. corresponding to motion selective cells) is required that somehow covers the to-be-observed sensory range appropriately. How should they sample the velocity space? In particular, for motion we have the problem that the potential maximal motion range needs to be very large, e.g. in order to be able to capture large visual displacements originated from head or eye movement, but that the sensitivity for relative motion differences also needs to be very high to be able to detect fine motion structures within the global displacement. The immediate consequence would be a very fine sampling of velocity space with a very large number of detectors. A more economic solution would be to use a reduced set of detectors which adapt to the global displacement and use their resources to sample velocity space more finely around it to increase the discriminability of the remaining relative motion. An observation that supports this idea is that from psychophysics; it is well known that the presence of a reference frame decreases motion discrimination thresholds [4].

In physiological studies, motion adaptation is found on a timescale of tens of msec to several seconds. In [5], it is shown that the direction tuning of single motion-sensitive neurons from macaque area MT (medio temporal cortex) changes within about 40 sec. In [6], it is shown that also the speed tuning is affected by the stimulus history and adapts to it in a timeframe of about 40 msec. The two findings suggest that response characteristics of motion detectors immediately and continuously adapt to a stimulus, and that as a consequence single motion detectors (in this case single neurons) exhibit a modified motion preference, measurable by a change of the form of their motion tuning curve.

The general idea of motion adaptation is visualized in Fig. (1). In (A), we sketch a system with a dense coverage of the motion space (exemplarily, only velocity magnitude but no velocity direction is indicated in the figure) by a large number of equidistributed motion detectors. Each motion detector has a tuning curve (highlighted in red for a single detector) centered around a sampling point in velocity space. In (B), we show a similar system with an insufficient coverage of the velocity space. Such a system will be deficient in terms of motion estimation accuracy. Only by adapting its motion detectors, it will be able to improve its estimations. This is shown in (C) for the velocity sampling points. The tuning curves move closer to the object velocity and cover that part of the velocity space better. In (D), the responsiveness of the tuning curves is changed by adaptation.

In a generative setting, the target of the system is to maximise 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. In this paper, we show how the adaptation dynamics of such a system can be derived from a straightforward probabilistic formulation. We also show that the adaptation increases the motion estimation accuracy allowing the system to rely on far less motion detectors to achieve results that are comparable to a dense, but non-adaptive sampling. In consequence, the derived detector adaptation leads to attracting shifts in the speed and direction tuning of motion detectors on a short timescale, similar to those found experimentally in [5], [6].

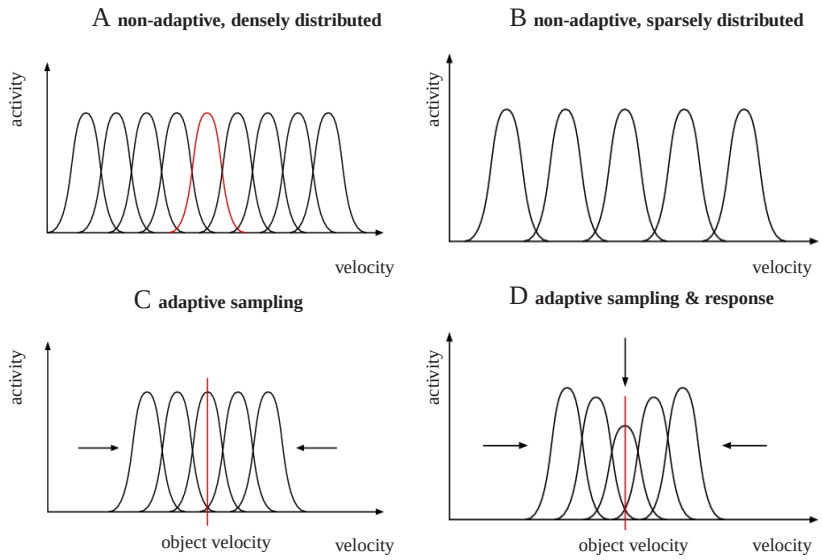

Figure 1: (A) Set of velocity-tuned neurons with velocity tuning curves densely sampling the entire velocity space. (B) Velocity preferences coarsely distributed in velocity space but (C) being *smartly adaptive to be able to cluster around some relevant velocities* [7] and (D) to be able to tune the responsiveness [6]. See text for further details.

2. The Probabilistic Filter Model for Motion Estimation

As a basis for our model, we use a modification of the probabilistic filter approach for motion estimation presented in [8]. We assume that the overall system describes a moving input by using velocity distributions $P(\mathbf{v}_\mathbf{x}^t)$ for discrete timesteps 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 $\mathbf{h} \in \mathbf{H}$.)

What we are interested in, however, is not the velocity at a single position but the *velocity field* $\mathbf{V}^t := \{\mathbf{v}_\mathbf{x}^t\}_\mathbf{x}$, which is the set of the velocities at all positions, and its (posterior) probability $P(\mathbf{V}^t|\mathbf{Y}^{1:t})$ given all visual inputs $\mathbf{Y}^{1:t} = \{\mathbf{Y}^1, \ldots, \mathbf{Y}^t\}$ until timestep t. In an incremental motion estimation process, the estimated velocity field $\mathbf{V}^t$ is influenced by the past velocity fields meaning that the old velocity estimations are used to bias the next ones, $P(\mathbf{V}^t|\mathbf{Y}^{1:t}) \rightarrow P(\mathbf{V}^{t'}|\mathbf{Y}^{1:t'}) \rightarrow P(\mathbf{V}^{t''}|\mathbf{Y}^{1:t''})$ etc. Therefore, incremental motion estimation allows to include the sensory history (e.g., the past input images), which leads to faster and more accurate estimations for continuous inputs.

The most important assumption for including the sensory history is *spatiotemporal consistency*. Intuitively, this can be understood by looking at the deterministic motion of particles at position $\mathbf{x}$ with velocities $\mathbf{v}_x^t$, which for the next timestep t' move towards position $\mathbf{x}'$. At the next timestep t' and for small time intervals of length $\Delta t = (t' - t)$, it is expected that the velocities obey $\mathbf{v}_\mathbf{x}^t \approx \mathbf{v}_{\mathbf{x}'}^{t'}$ and that the particles location follow approximately

$$\mathbf{x}' \approx \mathbf{x} + \Delta t\, \mathbf{v}_\mathbf{x}^t \approx \mathbf{x} + \Delta t\, \mathbf{v}_{\mathbf{x}'}^{t'} \,, \tag{1}$$

i.e., the particles move along with their corresponding velocities (for notational simplicity, in

the following we will omit the Δt). For a probabilistic modeling in terms of velocity transitions $\phi(\mathbf{v}_{\mathbf{x}'}^{t'}, \mathbf{v}_{\mathbf{x}}^{t})$ from current to next timestep velocities, from (1) it follows that, without incorporation of further knowledge from new measurements, there is a spatiotemporal coupling in the sense that

$$\phi(\mathbf{v}_{\mathbf{x}'}^{t'}, \mathbf{v}_{\mathbf{x}}^{t}) \approx f_x(\mathbf{x}' - \mathbf{v}_{\mathbf{x}'}^{t'}, \mathbf{x})\, f_t(\mathbf{v}_{\mathbf{x}'}^{t'}, \mathbf{v}_{\mathbf{x}}^{t})\,, \tag{2}$$

with e.g. Gaussians f_x, f_t, so that the probability for an estimated velocity $\mathbf{v}_{\mathbf{x}'}^{t'}$ increases when the locations and the velocities obey (1). Since also other locations and velocities from t can fulfill (1) approximately, we additionally have to sum over all these possibilities arriving at

$$\phi(\mathbf{v}_{\mathbf{x}'}^{t'}, \mathbf{V}^{t}) \approx \sum_{\mathbf{x}} f_x(\mathbf{x}' - \mathbf{v}_{\mathbf{x}'}^{t'}, \mathbf{x})\, f_t(\mathbf{v}_{\mathbf{x}'}^{t'}, \mathbf{v}_{\mathbf{x}}^{t})\,. \tag{3}$$

Although the velocity transitions $\phi(\mathbf{v}_{\mathbf{x}'}^{t'}, \mathbf{V}^{t})$ prefer spatiotemporal consistency, the final velocity estimate is able to capture spatiotemporal velocity changes since new measurements induce information about accelerating and decelerating stimuli.

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 [20]. In a Bayesian manner, we take the last velocity estimate (the last posterior at time t) $P(\mathbf{V}^{t}|\mathbf{Y}^{1:t};\mathbf{H})$ to calculate the next expected velocity estimate (the predictive prior at time t') $P(\mathbf{V}^{t'}|\mathbf{Y}^{1:t};\mathbf{H})$. This is combined with the likelihood that the next timestep sensory input $\mathbf{Y}^{t'} := \{\mathbf{I}^{t'}, \mathbf{I}^{t}\}$ (with the next and the last input images $\mathbf{I}^{t'}$ and $\mathbf{I}^{t}$) can be explained given a velocity estimate, to gain the next posterior via

$$P(\mathbf{V}^{t'}|\mathbf{Y}^{1:t'};\mathbf{H}) \propto \underbrace{P(\mathbf{Y}^{t'}|\mathbf{V}^{t'};\mathbf{H})}_{\text{Measurement likelihood}} \times \underbrace{P(\mathbf{V}^{t'}|\mathbf{Y}^{1:t};\mathbf{H})}_{\text{Predictive prior}}\,. \tag{4}$$

The likelihood can be expressed via

$$P(\mathbf{Y}^{t'}|\mathbf{V}^{t'};\mathbf{H}) = \prod_{\mathbf{x}'} \ell(\mathbf{Y}^{t'}, \mathbf{v}_{\mathbf{x}}^{t'};\mathbf{H})\,. \tag{5a}$$

$$\ell(\mathbf{Y}^{t'}, \mathbf{v}_{\mathbf{x}'}^{t'};\mathbf{H}) = \sum_{\mathbf{h}\in\mathbf{H}} \delta(\mathbf{v}_{\mathbf{x}'}^{t'} - \mathbf{h})\, \ell(\mathbf{Y}^{t'}, \mathbf{v}_{\mathbf{x}'}^{t'})\,. \tag{5b}$$

$$\ell(\mathbf{Y}^{t'}, \mathbf{v}_{\mathbf{x}'}^{t'}) = f_\ell(\mathbf{I}_{\mathbf{x}'}^{t'}, \mathbf{I}_{\mathbf{x}'-\mathbf{v}_{\mathbf{x}'}^{t'}}^{t};\theta_\ell)\,, \tag{5c}$$

with likelihood-specific fixed parameters θ_ℓ. That means, that the overall likelihood factorises into position-specific likelihoods (5a). Due to (5b) the velocity parameterization at each position is restricted to the sampling velocities $\mathbf{h} \in \mathbf{H}$, and with (5c) the local likelihood is gained by the patchwise comparison of the current and the future input images around positions compatible with the velocity hypothesis $\mathbf{v}_{\mathbf{x}}^{t}$, as proposed e.g. in [21]. The $\mathbf{I}_{\mathbf{x}'}^{t'}$ and $\mathbf{I}_{\mathbf{x}}^{t}$ are image patches taken from the images $\mathbf{I}^{t'}$ and $\mathbf{I}^{t}$ with a fixed window and anchored around positions $\mathbf{x}'$ and $\mathbf{x}$, respectively. In (5c) we have again made use of the spatiotemporal consistency equation (1), assuming that, for a correct motion on the image, the image patches move accordingly, such that

$$\mathbf{I}_{\mathbf{x}'}^{t'} \approx \mathbf{I}_{\mathbf{x}'-\mathbf{v}_{\mathbf{x}'}^{t'}}^{t}\,. \tag{6}$$

The transition to get the predictive prior is done by

$$
\begin{aligned}
P(\mathbf{V}^{t'}|\mathbf{Y}^{1:t};\mathbf{H}) &= \sum_{\mathbf{V}^t} P(\mathbf{V}^{t'}|\mathbf{V}^t;\mathbf{H}) P(\mathbf{V}^t|\mathbf{Y}^{1:t};\mathbf{H})\,, && (7)\\
P(\mathbf{V}^{t'}|\mathbf{V}^t;\mathbf{H}) &= \prod_{\mathbf{x}} \phi(\mathbf{v}^{t'}_{\mathbf{x}'}, \mathbf{V}^t;\mathbf{H})\,,\\
\phi(\mathbf{v}^{t'}_{\mathbf{x}'}, \mathbf{V}^t;\mathbf{H}) &= \sum_{\mathbf{h}\in\mathbf{H}} \delta(\mathbf{v}^{t'}_{\mathbf{x}'} - \mathbf{h})\, \phi(\mathbf{v}^{t'}_{\mathbf{x}'}, \mathbf{V}^t)\,,\\
\phi(\mathbf{v}^{t'}_{\mathbf{x}'}, \mathbf{V}^t) &= \sum_{\mathbf{x}} f_x(\mathbf{x}' - \mathbf{v}^{t'}_{\mathbf{x}'}, \mathbf{x};\theta_x)\; f_t(\mathbf{v}^{t'}_{\mathbf{x}'}, \mathbf{v}^t_{\mathbf{x}};\theta_t)\,, && (8)
\end{aligned}
$$

with prediction-specific parameters θ_x, θ_t. Here, we have assumed that the predictive prior factorises 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 (cf. Sec. 4 for more detailed information).

Inserting (7) into (4) leads to the posterior

$$
P(\mathbf{V}^{t'}|\mathbf{Y}^{1:t'};\mathbf{H}) \propto P(\mathbf{Y}^{t'}|\mathbf{V}^{t'};\mathbf{H}) \sum_{\mathbf{V}^t} P(\mathbf{V}^{t'}|\mathbf{V}^t;\mathbf{H}) P(\mathbf{V}^t|\mathbf{Y}^{1:t};\mathbf{H})\,, \tag{9}
$$

which factorises spatially, such that using (5) and (8) we obtain

$$
P(\mathbf{V}^{t'}|\mathbf{Y}^{1:t'};\mathbf{H}) = \prod_{\mathbf{x}'} P(\mathbf{v}^{t'}_{\mathbf{x}'}|\mathbf{Y}^{1:t'};\mathbf{H})\,, \tag{10}
$$

with the single velocity posterior

$$
\begin{aligned}
P(\mathbf{v}^{t'}_{\mathbf{x}'}|\mathbf{Y}^{1:t'};\mathbf{H}) &\propto \ell(\mathbf{Y}^{t'}, \mathbf{v}^{t'}_{\mathbf{x}'};\mathbf{H}) \sum_{\mathbf{V}^t} \sum_{\mathbf{x}} \phi(\mathbf{v}^{t'}_{\mathbf{x}'}, \mathbf{V}^t;\mathbf{H})\, P(\mathbf{V}^t|\mathbf{Y}^{1:t};\mathbf{H})\\
&= \sum_{\mathbf{h}'\in\mathbf{H}} \delta(\mathbf{v}^{t'}_{\mathbf{x}'} - \mathbf{h}')\; f_\ell(\mathbf{I}^{t'}_{\mathbf{x}'}, \mathbf{I}^t_{\mathbf{x}'-\mathbf{v}^{t'}_{\mathbf{x}'}};\theta_\ell) \sum_{\mathbf{x}} \sum_{\mathbf{v}^t_{\mathbf{x}}} \sum_{\mathbf{h}\in\mathbf{H}} \delta(\mathbf{v}^t_{\mathbf{x}} - \mathbf{h}) \times\\
&\quad \underbrace{f_x(\mathbf{x}' - \mathbf{v}^{t'}_{\mathbf{x}'}, \mathbf{x};\theta_x)\, f_t(\mathbf{v}^{t'}_{\mathbf{x}'}, \mathbf{v}^t_{\mathbf{x}};\theta_t)}_{\phi(\mathbf{v}^{t'}_{\mathbf{x}'}, \mathbf{v}^t_{\mathbf{x}}) \text{ defined in (2)}}\; P(\mathbf{v}^t_{\mathbf{x}}|\mathbf{Y}^{1:t};\mathbf{H})\,. && (11)
\end{aligned}
$$

Eq. (11) 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 $\mathbf{x}$ as well as $\mathbf{x}'$. We further define

$$
\phi(\mathbf{v}^{t'}_{\mathbf{x}'}, \mathbf{v}^t_{\mathbf{x}};\mathbf{H}) := \sum_{\mathbf{h}\in\mathbf{H}} \delta(\mathbf{v}^{t'}_{\mathbf{x}} - \mathbf{h})\, \phi(\mathbf{v}^{t'}_{\mathbf{x}'}, \mathbf{v}^t_{\mathbf{x}})\,. \tag{12}
$$

In summary, using definition (12), we now can introduce a compact notation

$$
P(\mathbf{v}^{t'}_{\mathbf{x}'}|\mathbf{Y}^{1:t};\mathbf{H}) := \sum_{\mathbf{x}} \sum_{\mathbf{v}^t_{\mathbf{x}}} \phi(\mathbf{v}^{t'}_{\mathbf{x}'}, \mathbf{v}^t_{\mathbf{x}};\mathbf{H})\, P(\mathbf{v}^t_{\mathbf{x}}|\mathbf{Y}^{1:t};\mathbf{H})\,, \tag{13}
$$

as the single velocity predictive prior, i.e., the velocity prediction for the next timestep t' gained by propagation of the last posterior and using only the measurements $\mathbf{Y}^{1:t}$ until t, which will be used in the derivations of the next section.

3. Online Adaptation of the Velocity Tuning

The incremental filtering uses the sensory history to adapt the motion estimations. In addition, we can consider different parameter sets of the velocity estimation process to be adaptive at every time step t. Following the argumentation from the introduction and motivated by the biological findings, we concentrate on the tuning curve positions and tuning curve widths of the motion detectors.

In the system introduced in Sec. 2, the velocity detecting units measure the current velocity estimate directly from the input images as expressed by $\ell(\mathbf{Y}^{t'}, \mathbf{v}^{t'}_{\mathbf{x}'}; \theta_\ell)$ in (5). The function $\delta(\mathbf{v}^{t'}_{\mathbf{x}'} - \mathbf{h})$ from (5) restricts the motion detectors to velocities $\mathbf{v}^{t'}_{\mathbf{x}'} = \mathbf{h}$, so that we assume each motion detecting unit to be realised as follows

$$\ell(\mathbf{Y}^{t'}, \mathbf{v}^{t'}_{\mathbf{x}'} = \mathbf{h}; \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(-\frac{1}{2\sigma^2} \sum_{\mathbf{y}'} w(\mathbf{y}' - \mathbf{x}')(I^{t'}_{\mathbf{y}'} - I^{t}_{\mathbf{y}'-h})^2\Big)\,, \tag{14}$$

i.e., their velocity tuning curve is Gaussian-shaped with a center at $\mathbf{h}$ and standard deviation $\theta_\ell = \sigma$. Here, $w(\mathbf{y}' - \mathbf{x}')$ is a Gaussian weighting function of the spatial neighborhood around $\mathbf{x}'$ and $I^t_{\mathbf{x}}$ is the intensity of the input $\mathbf{I}^t$ at position $\mathbf{x}$. The image consists of X discrete 2D pixel positions $\mathbf{x}$. We now assume the motion estimation to be based on a set of N motion detecting units with corresponding velocity sampling points (resp. tuning curve centers) $\mathbf{H} = \{\mathbf{h}_n\}_n$ and standard deviations $\Sigma = \{\sigma_n\}_n$ that parameterise the tuning curve widths. Both, the tuning curve centers as well as the tuning curve widths are allowed to be adaptive and can adjust to the current and past motion estimation statistics of the input.

The temporal adaptation of the parameter sets $\{\mathbf{H}, \Sigma\}$ is realised via an extended online variant of the Expectation-Maximisation algorithm described in the next section. For this purpose, we introduce additional priors for the parameters, $P(\mathbf{H})$ and $P(\Sigma)$, respectively, to keep the parameters in suitable adaptation ranges.

3.1. Maximum posterior solution

We consider the velocity tuning curve centers h_n and the tuning curve widths σ_n to be adaptive at every time step. They are parameters of the velocity estimation process, which can be optimised via an approximate Expectation-Maximisation (EM) mechanism.

The target is to maximise the log probability function $\ln P(\mathbf{Y}^{1:t}|\mathbf{H}, \Sigma)$ of the data $\mathbf{Y}^{1:t}$ up to time t given the parameters $\mathbf{H}, \Sigma$ with respect to the optimal parameters $\mathbf{H}^{t'}, \Sigma^{t'}$ for the next time step t'. The likelihood function is given by

$$P(\mathbf{Y}^{1:t}|\mathbf{H}^t, \Sigma^t) = \sum_{\mathbf{V}^t} P(\mathbf{Y}^{1:t}, \mathbf{V}^t|\mathbf{H}^t, \Sigma^t)\,. \tag{15}$$

W.l.o.g. and for derivation purposes, in the following we will assume that the same velocity sampling points $\mathbf{H}$ and standard deviations Σ are used over the entire visual field. That is, the sampling points and standard deviations are independent on the positions along the visual field for which the velocity detection units are responsible. Since the probabilistic motion model given in (11) provides the posterior distribution $P(\mathbf{V}^t|\mathbf{Y}^{1:t}; \mathbf{H}^t, \Sigma^t)$ only up to time t, an online EM algorithm is applied that only considers the expected value under the posterior given only past data for parameter optimisation (instead of the complete-data log likelihood). The expectation for the log likelihood evaluated for the parameter sets $\mathbf{H}, \Sigma$, denoted $Q_{\mathbf{H}^t, \Sigma^t}(\mathbf{H}, \Sigma)$, is given by

$$Q_{\mathbf{H}^t, \Sigma^t}(\mathbf{H}, \Sigma) = \sum_{\mathbf{V}^t} P(\mathbf{V}^t|\mathbf{Y}^{1:t}; \mathbf{H}^t, \Sigma^t)\, \ln P(\mathbf{Y}^{1:t}, \mathbf{V}^t|\mathbf{H}, \Sigma)\,. \tag{16}$$

Introducing suitable priors for the parameters $P(\mathbf{H})$ and $P(\Sigma)$ we can calculate successively these parameters by maximising the expectation while considering the priors (which is called maximum posterior solution). This yields

$$\{\mathbf{H}^{t'}, \Sigma^{t'}\} = \mathrm{argmax}_{\mathbf{H},\Sigma}\Big(Q_{\mathbf{H}^t,\Sigma^t}(\mathbf{H},\Sigma) + \lambda_{\mathbf{h}} \ln P(\mathbf{H}) + \lambda_\sigma \ln P(\Sigma)\Big) \quad , \tag{17}$$

with $\lambda_{\mathbf{h}}$ and λ_σ being weighting factors to adjust the influence of the parameter priors.

3.2. Priors for the tuning parameters

The prior for the set of velocity sampling points is chosen as a product of Gaussian-like exponentials

$$P(\mathbf{H}) \propto \prod_n \exp\Big(-\frac{1}{2}||\mathbf{h}_n - \mathbf{h}_{n,0}||^2\Big) . \tag{18}$$

This prior becomes maximal if $\mathbf{H} = \mathbf{H}_0$, with the N reference sampling points $\mathbf{h}_{n,0} \in \mathbf{H}_0$ sitting on a squared 2D-grid distributed equidistantly in velocity space around zero. This prior prevents the velocity samples to get too close together and to collapse into one single velocity sample point.

The prior for the set of standard deviations is chosen as a product of Gaussians such that the sensitivities of the velocity detection units do not become too diverse during adaptation and keep being around some reference value σ_0

$$P(\Sigma) \propto \prod_n \frac{1}{\sqrt{2\pi}\sigma_n} \exp\Big(-\frac{1}{2\sigma_n^2}||\sigma_n - \sigma_0||^2\Big) . \tag{19}$$

3.3. Derivation of the adaptation rules

Next, we derive the adaptation rules by an approximate maximum posterior solution. To this end, the following partial derivatives with respect to the parameter sets $\mathbf{H}$ and Σ have to be set to the zero vector

$$\frac{\partial Q_{\mathbf{H}^t,\Sigma^t}(\mathbf{H},\Sigma)}{\partial \mathbf{H}} + \lambda_{\mathbf{h}} \frac{\partial \ln P(\mathbf{H})}{\partial \mathbf{H}} \stackrel{!}{=} \mathbf{0} , \tag{20}$$

$$\frac{\partial Q_{\mathbf{H}^t,\Sigma^t}(\mathbf{H},\Sigma)}{\partial \Sigma} + \lambda_\sigma \frac{\partial \ln P(\Sigma)}{\partial \Sigma} \stackrel{!}{=} \mathbf{0} . \tag{21}$$

Starting with the partial derivative $\partial Q/\partial \mathbf{H}$ and assuming $P(\mathbf{Y}^{1:t}|\mathbf{H})$ to be independent of the parameters $\mathbf{H}$, we obtain

$$\frac{\partial Q}{\partial \mathbf{H}} = \frac{\partial}{\partial \mathbf{H}} \sum_{\mathbf{V}^t} P(\mathbf{V}^t|\mathbf{Y}^{1:t};\mathbf{H}^t) \ln P(\mathbf{V}^t|\mathbf{Y}^{1:t};\mathbf{H}) + \underbrace{\frac{\partial}{\partial \mathbf{H}} \sum_{\mathbf{V}^t} P(\mathbf{V}^t|\mathbf{Y}^{1:t};\mathbf{H}^t) \ln P(\mathbf{Y}^{1:t}|\mathbf{H})}_{=\mathbf{0}} . \tag{22}$$

Using the factorisation assumption (10) and replacing the posterior with fixed parameters with the abbreviation $\alpha(\mathbf{v}_{\mathbf{x}}^t;\mathbf{H}^t) = P(\mathbf{v}_{\mathbf{x}}^t|\mathbf{Y}^{1:t};\mathbf{H}^t,\Sigma^t)$ simplifies the derivative as follows

$$\frac{\partial Q}{\partial \mathbf{H}} = \frac{\partial}{\partial \mathbf{H}} \sum_{\mathbf{V}^t} \prod_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^t;\mathbf{H}^t) \ln \prod_{\mathbf{x}} P(\mathbf{v}_{\mathbf{x}}^t|\mathbf{Y}^{1:t};\mathbf{H}) \tag{23}$$

$$\approx \frac{\partial}{\partial \mathbf{H}} \sum_{\mathbf{x}} \sum_{\mathbf{v}_{\mathbf{x}}^t} \alpha(\mathbf{v}_{\mathbf{x}}^t;\mathbf{H}^t) \ln P(\mathbf{v}_{\mathbf{x}}^t|\mathbf{Y}^{1:t};\mathbf{H}) \tag{24}$$

$$= \sum_{\mathbf{x}} \sum_{\mathbf{v}_{\mathbf{x}}^t} \alpha(\mathbf{v}_{\mathbf{x}}^t;\mathbf{H}^t) \frac{\partial}{\partial \mathbf{H}} \ln \frac{1}{Z} \ell(\mathbf{Y}^t,\mathbf{v}_{\mathbf{x}}^t;\mathbf{H}) P(\mathbf{v}_{\mathbf{x}}^t|\mathbf{Y}^{1:t-1};\mathbf{H}) . \tag{25}$$

Here, $P(\mathbf{v}_\mathbf{x}^t|\mathbf{Y}^{1:t};\mathbf{H})$ is the single velocity predictive prior from (13) and the quantity Z is a normalisation constant in order to fulfil $\sum_{\mathbf{v}_\mathbf{x}^t} P(\mathbf{v}_\mathbf{x}^t|\mathbf{Y}^{1:t};\mathbf{H}) = 1$. Next, we assume the adaptation of the velocity sampling to have negligible influence on the predictive prior, such that it is calculated with the last velocity sampling $\mathbf{H}^{t-1}$, $P(\mathbf{v}_\mathbf{x}^t|\mathbf{Y}^{1:t-1};\mathbf{H}) \approx P(\mathbf{v}_\mathbf{x}^t|\mathbf{Y}^{1:t-1};\mathbf{H}^{t-1})$. This approximation simplifies the derivative and we arrive at

$$\frac{\partial Q}{\partial \mathbf{H}} = \sum_\mathbf{x} \sum_{\mathbf{v}_\mathbf{x}^t} \alpha(\mathbf{v}_\mathbf{x}^t;\mathbf{H}^t)\Big(\underbrace{\frac{\partial}{\partial \mathbf{H}} \ln \frac{1}{Z}}_{=\mathbf{0}} + \frac{\partial}{\partial \mathbf{H}} \ln \ell(\mathbf{Y}^t, \mathbf{v}_\mathbf{x}^t;\mathbf{H}) + \underbrace{\frac{\partial}{\partial \mathbf{H}} \ln P(\mathbf{v}_\mathbf{x}^t|\mathbf{Y}^{1:t-1};\mathbf{H}^{t-1})}_{=\mathbf{0}} \Big). \tag{26}$$

Now, the adaptation tries to tune the parameters of the velocity detection units such that the expectation of the new observation likelihoods under the past posterior is maximised only but not the expectation of the transition probability. Hence, the adaptation is more sensitive to changes in the velocity statistics of the visual scene.

In the following, we take the remaining term from (26) and insert the observation likelihood (5). Evaluating the δ-functions that constrain the velocities $\mathbf{v}_\mathbf{x}^t$ to the discrete sampling points $\mathbf{h}_n$ leads to a simplified partial derivative

$$\begin{aligned} \frac{\partial Q}{\partial \mathbf{H}} &= \frac{\partial}{\partial \mathbf{H}} \sum_\mathbf{x} \sum_{\mathbf{v}_\mathbf{x}^t} \alpha(\mathbf{v}_\mathbf{x}^t;\mathbf{H}^t) \ln \sum_n \delta(\mathbf{v}_\mathbf{x}^t - \mathbf{h}_n) \ell(\mathbf{Y}^t, \mathbf{v}_\mathbf{x}^t) & (27) \\ &\approx \frac{\partial}{\partial \mathbf{H}} \sum_\mathbf{x} \sum_n \alpha(\mathbf{v}_\mathbf{x}^t = \mathbf{h}_n^t;\mathbf{H}^t) \ln \ell(\mathbf{Y}^t, \mathbf{v}_\mathbf{x}^t = \mathbf{h}_n)\,. & (28) \end{aligned}$$

If we specify the observation likelihood (5) with $\theta_\ell = \sigma_n$ as given in (14) and linearise it with the first order Taylor expansion, similar to the likelihood given in [18] using the Lucas-Kanade approach [19],

$$\begin{aligned} \ell(\mathbf{Y}^t, \mathbf{v}_\mathbf{x}^t = \mathbf{h}_n) &= \frac{1}{\sqrt{2\pi}\sigma_n} \exp\Big(-\frac{1}{2\sigma_n^2} \sum_\mathbf{y} w(\mathbf{y}-\mathbf{x})(I_\mathbf{y}^t - I_{\mathbf{y}-\mathbf{h}_n}^{t-1})^2\Big) & (29) \\ &\approx \frac{1}{\sqrt{2\pi}\sigma_n} \exp\Big(-\frac{1}{2\sigma_n^2} (\nabla^\top I_\mathbf{x}^t \mathbf{h}_n + I_{t,\mathbf{x}}^t)^2\Big), & (30) \end{aligned}$$

we arrive at the final expression for the derivatives of Q

$$\begin{aligned} \frac{\partial Q}{\partial \mathbf{H}} &= \Big\{\frac{\partial Q}{\partial \mathbf{h}_n}\Big\}_n, & (31) \\ \frac{\partial Q}{\partial \mathbf{h}_n} &= \sum_\mathbf{x} \alpha(\mathbf{v}_\mathbf{x}^t = \mathbf{h}_n^t;\mathbf{H}^t) \frac{\partial}{\partial \mathbf{h}_n} \ln \ell(\mathbf{Y}^t, \mathbf{v}_\mathbf{x}^t = \mathbf{h}_n) & (32) \\ &= \sum_\mathbf{x} \alpha(\mathbf{v}_\mathbf{x}^t = \mathbf{h}_n^t;\mathbf{H}^t) \frac{\partial}{\partial \mathbf{h}_n} \ln \frac{1}{\sqrt{2\pi}\sigma_n} \exp\Big(-\frac{1}{2\sigma_n^2} (\nabla^\top I_\mathbf{x}^t \mathbf{h}_n + I_{t,\mathbf{x}}^t)^2\Big) & (33) \\ &= -\frac{1}{\sigma_n^2} \sum_\mathbf{x} \alpha(\mathbf{v}_\mathbf{x}^t = \mathbf{h}_n^t;\mathbf{H}^t)(\nabla^\top I_\mathbf{x}^t \mathbf{h}_n + I_{t,\mathbf{x}}^t) \nabla I_\mathbf{x}^t \quad . & (34) \end{aligned}$$

The intensity gradient $\nabla I_\mathbf{x}^t = \Big[I_{x,\mathbf{x}}, I_{y,\mathbf{x}}\Big]^\top$ denotes spatial derivatives of the image intensities in x- and y-direction and $I_{t,\mathbf{x}}$ denotes temporal derivatives, respectively, taken at position $\mathbf{x}$. Since for the computation of the spatial derivatives in (30) the images are pre-smoothed with the same Gaussian weighting $w(\mathbf{y}-\mathbf{x})$ as used in (29), we neglect the weighting in the linearisation (30).

Next, we need the derivatives of the prior

$$\frac{\partial \ln P(\mathbf{H})}{\partial \mathbf{H}} = \Big\{\frac{\partial \ln P(\mathbf{H})}{\partial \mathbf{h}_n}\Big\}_n , \tag{35}$$

$$\frac{\partial \ln P(\mathbf{H})}{\partial \mathbf{h}_n} = \frac{\partial}{\partial \mathbf{h}_n} \ln \prod_n \exp\Big(-\frac{1}{2}\|\mathbf{h}_n - \mathbf{h}_{n,0}\|^2\Big) = (\mathbf{h}_{n,0} - \mathbf{h}_n) . \tag{36}$$

Finally, in order to maximise the objective function defined in (17)

$$\{\mathbf{H}^{t'}\} = \text{argmax}_{\mathbf{H}}\Big(Q_{\mathbf{H}^t,\Sigma^t}(\mathbf{H},\Sigma) + \lambda_{\mathbf{h}} \ln P(\mathbf{H})\Big), \tag{37}$$

we insert the derivatives (34) and (36) into the constraint (20)

$$\frac{\partial Q_{\mathbf{H}^t,\Sigma^t}(\mathbf{H},\Sigma)}{\partial \mathbf{H}} + \lambda_{\mathbf{h}} \frac{\partial \ln P(\mathbf{H})}{\partial \mathbf{H}} \stackrel{!}{=} \mathbf{0} , \tag{38}$$

$$\Big\{\frac{\partial Q_{\mathbf{H}^t,\Sigma^t}(\mathbf{H},\Sigma)}{\partial \mathbf{h}_n} + \lambda_{\mathbf{h}} \frac{\partial \ln P(\mathbf{H})}{\partial \mathbf{h}_n}\Big\}_n \stackrel{!}{=} \mathbf{0} , \tag{39}$$

$$\Big\{ -\frac{1}{\sigma_n^2} \sum_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^t = \mathbf{h}_n^t; \mathbf{H}^t)(\nabla^\top I_{\mathbf{x}}^t \mathbf{h}_n + I_{t,\mathbf{x}}^t)\Big[\nabla I_{\mathbf{x}}^t\Big] + \lambda_{\mathbf{h}}\left[\mathbf{h}_{n,0} - \mathbf{h}_n\right]\Big\}_n \stackrel{!}{=} 0 , \tag{40}$$

and solve for $\mathbf{h}_n = \mathbf{h}_n^{t'}$ to get the adaptation rule for the set of new velocity sampling points $\mathbf{H}^{t'}$ as follows

$$\mathbf{h}_n^{t'} = 2^k \begin{bmatrix} \sum_{\mathbf{x}} \alpha_n (I_{x,\mathbf{x}}^{k,t})^2 + \lambda_{\mathbf{h}}\sigma_n^2 & \sum_{\mathbf{x}} \alpha_n I_{x,\mathbf{x}}^{k,t} I_{y,\mathbf{x}}^{k,t} \\ \sum_{\mathbf{x}} \alpha_n I_{x,\mathbf{x}}^{k,t} I_{y,\mathbf{x}}^{k,t} & \sum_{\mathbf{x}} \alpha_n (I_{y,\mathbf{x}}^{k,t})^2 + \lambda_{\mathbf{h}}\sigma_n^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_{\mathbf{x}} \alpha_n I_{t,\mathbf{x}}^{k,t} I_{x,\mathbf{x}}^{k,t} + \lambda_{\mathbf{h}}\sigma_n^2 \frac{h_{x0}}{2^k} \\ -\sum_{\mathbf{x}} \alpha_n I_{t,\mathbf{x}}^{k,t} I_{y,\mathbf{x}}^{k,t} + \lambda_{\mathbf{h}}\sigma_n^2 \frac{h_{y0}}{2^k} \end{bmatrix} , \tag{41}$$

using the abbreviation $\alpha_n = \alpha(\mathbf{v}_{\mathbf{x}}^t = \mathbf{h}_n^t; \mathbf{H}_n^t)$. Since we have linearised the exponent of the observation likelihood in (29) the adaptation rule only holds for small velocities. To circumvent this problem, for each $\mathbf{h}_n^t$ we downsample the derivatives of the image intensities with a Gaussian pyramid [9] to a proper scale k. The scale depends on the amplitude of the velocity sampling point $\|\mathbf{h}_n^t\|$ such that the linearisation assumption of small movements holds, like $round(\|\mathbf{h}_n^t\|/2^k) = 1$ with $k = 0, ..., K$.

The derivation of the update rules to achieve an adapted set of standard deviations $\Sigma^{t'}$ works in an analogous way. The standard deviations enter the formulations in the likelihood $\ell(\mathbf{Y}^t, \mathbf{v}_{\mathbf{x}}^t; \mathbf{H}, \Sigma)$. In order to maximise the objective function defined in (17) with respect to Σ,

$$\Sigma^{t'} = \text{argmax}_{\Sigma}\Big(Q_{\mathbf{H}^t,\Sigma^t}(\mathbf{H},\Sigma) + \lambda_{\sigma} \ln P(\Sigma)\Big), \tag{42}$$

we then insert the derivatives

$$\frac{\partial \ln P(\Sigma)}{\partial \Sigma} = \Big\{\frac{\partial \ln P(\Sigma)}{\partial \sigma_n}\Big\}_n , \tag{43}$$

$$\frac{\partial \ln P(\Sigma)}{\partial \sigma_n} = \frac{\partial}{\partial \sigma_n} \ln \prod_m \frac{1}{\sqrt{2\pi}\sigma_m} \exp\Big(-\frac{1}{2\sigma_m^2}\|\sigma_m - \sigma_0\|^2\Big) = -\frac{\sigma_0^2}{\sigma_n^3} , \tag{44}$$

and, analogous to eq. (31),

$$\frac{\partial Q}{\partial \Sigma} = \Big\{\frac{\partial Q}{\partial \sigma_n}\Big\}_n , \tag{45}$$

$$\frac{\partial Q}{\partial \sigma_n} = \sum_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^t; \mathbf{H}^t) \frac{\partial}{\partial \sigma_n} \ln \ell(\mathbf{Y}^t, \mathbf{v}_{\mathbf{x}}^t; \mathbf{H}, \Sigma) \tag{46}$$

$$= \sum_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^t; \mathbf{H}^t) \frac{\partial}{\partial \sigma_n} \ln \frac{1}{\sqrt{2\pi}\sigma_n} \exp\Big(-\frac{1}{2\sigma_n^2}(\nabla^\top I_{\mathbf{x}}^t \mathbf{h}_n + I_{t,\mathbf{x}}^t)^2\Big) \tag{47}$$

$$= \frac{1}{\sigma_n^3}\Big(\sum_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^t; \mathbf{H}^t)(\nabla^\top I_{\mathbf{x}}^t \mathbf{h}_n + I_{t,\mathbf{x}}^t)^2 - \sum_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^t; \mathbf{H}^t)\sigma_n^2\Big), \tag{48}$$

in the constraint (21)

$$\frac{\partial Q_{\mathbf{H}^t,\Sigma^t}(\mathbf{H}, \Sigma)}{\partial \Sigma} + \lambda_\sigma \frac{\partial \ln P(\Sigma)}{\partial \Sigma} \stackrel{!}{=} \mathbf{0}, \tag{49}$$

$$\Big\{\frac{\partial Q_{\mathbf{H}^t,\Sigma^t}(\mathbf{H}, \Sigma)}{\partial \sigma_n} + \lambda_\sigma \frac{\partial \ln P(\Sigma)}{\partial \sigma_n}\Big\}_n \stackrel{!}{=} \mathbf{0}, \tag{50}$$

$$\Big\{\frac{1}{\sigma_n^3}\Big(\sum_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^t; \mathbf{H}^t)(\nabla^\top I_{\mathbf{x}}^t \mathbf{h} + I_{t,\mathbf{x}}^t)^2 - \sum_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}; \mathbf{H}^t)\sigma_n^2\Big) - \lambda_\sigma \frac{\sigma_0^2}{\sigma_n^3}\Big\}_n \stackrel{!}{=} 0, \tag{51}$$

and obtain the adaptation rule for the set of new standard deviations $\Sigma^{t'}$ as follows

$$(\sigma_n^2)^{t'} = \Big(\sum_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^t; \mathbf{H}^t)(\nabla^\top I_{\mathbf{x}}^{k,t} \mathbf{h}_n^t / 2^k + I_{t,\mathbf{x}}^{k,t})^2 - \lambda_\sigma \sigma_0^2\Big) / \Big(\sum_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^t; \mathbf{H}^t)\Big). \tag{52}$$

3.4. Interpretation of the adaptation rules

With (41) and (52), we have gained an explicit iterative description for the calculation of the next optimal sampling points $\mathbf{H}^{t'}$ and standard deviations $\Sigma^{t'}$ given the current set of adaptive $(\mathbf{H}^t, \Sigma^t)$ and fixed $(\lambda_{\mathbf{h}}, \lambda_\sigma)$ parameters. It takes into account the spatial statistics $(\sum_{\mathbf{x}} ...)$ of the current posterior distribution $\alpha(\mathbf{v}_{\mathbf{x}}^t; \mathbf{H}^t) = P(\mathbf{v}_{\mathbf{x}}^t | \mathbf{Y}^{1:t}; \mathbf{H}^t, \Sigma^t)$ to adapt the velocity tuning $\mathbf{h}_n^{t'} \in \mathbf{H}^{t'}$ and the standard deviation $\sigma_n^{t'} \in \Sigma^{t'}$ of each $n = 1 \ldots N$ of the velocity detection units.

The posterior probabilities $\alpha(\mathbf{v}_{\mathbf{x}}^t; \mathbf{H}^t)$ can be interpreted as a model for the neural activity of velocity sensitive MT neurons. Their responses are driven by the responses of the motion detection units $\ell(\mathbf{Y}^{t'}, \mathbf{v}_{\mathbf{x}'}^{t'} = \mathbf{h}; \sigma)$ which can be interpreted as a model for velocity tuned V1 cells. The proposed adaptive Bayesian filter (11) realises a spatiotemporal dynamics that can be interpreted as a model for the dynamics of the neural MT activity via a strong closed-loop feedback interaction.

The adaptation of the velocity detection units (V1 cells), namely adaptation of preferred velocities $\mathbf{H}^{t'}$ and sensitivity range $\Sigma^{t'}$, is driven by the current probability of the posterior $\alpha(\mathbf{v}_{\mathbf{x}}^t; \mathbf{H}^t)$ (activity of MT cells), the current spatiotemporal luminance changes in the visual input $(\nabla^\top I_{\mathbf{x}}^{k,t}, I_{t,\mathbf{x}}^t)$ and weak ($\lambda_{\mathbf{h}}$ and λ_σ are small) time-independent priors $(\mathbf{H}_0, \Sigma_0)$ on the tuning. This leads to an adaptation of the preferred velocities (41) and changes the sensitivity (52) of the MT cells such that they best explain the luminance changes in the visual input and simultaneously maximise their activity. The priors lead to a trade-off between a neural population of MT cells with their preferred velocities being equally distributed in velocity space versus MT cells with their preferred velocities being distributed around the statistically relevant velocities. The

clustering around some certain velocities influences the adaptation of the sensitivity range of the corresponding neurons and vice versa. The dependencies between sampling and sensitivity are shown in the next section 4.

Since we assume the posterior to factorise locally ($\prod_{\mathbf{x}} ...$) the updates are dependent on local information only. For this reason, the update can be implemented efficiently in a purely parallel way. The following pseudo-code shows the compact form of the derived adaptive spatiotemporal filter suitable for an algorithmic implementation. All experiments described in the next section have been realised in Matlab.

Algorithm 1 Pseudo-code for the adaptive spatiotemporal filter

Initialise the priors $\alpha(\mathbf{v}_{\mathbf{x}}^0)$ and the parameter sets $\mathbf{H}_0$, Σ_0, $\lambda_{\mathbf{h}}$, λ_σ.
for $t = 1$ to T **do**
  for $\mathbf{x} = 1$ to X **do**
    Compute the posterior $\alpha(\mathbf{v}_{\mathbf{x}}^t; \mathbf{H}^t) = P(\mathbf{v}_{\mathbf{x}}^t | \mathbf{Y}^{1:t}; \mathbf{H}^t, \Sigma^t)$ according to equation (11)
  end for
  Update the parameters $\mathbf{h}_n^{t'}$ and $(\sigma_n^2)^{t'}$ of the sets $\mathbf{H}$ and Σ for the next time step t' according to equations (41) and (52).
end for

4. Experiments

What is the effect of the adaptation rules derived in the last section? Consider the probabilistic filter from (11) without adaptation, 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 completely covering the possible range of inputs, e.g. at regular intervals in velocity space. With the adaptation according to (41) and (52), the tuning curves of the velocity-detecting units cluster around the relevant velocities available in the input image, providing a finer sampling around these velocities and different maximum values of response. Fig. (1) (C) and (D) 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. To verify this hypothesis, we provide several experiments.

4.1. Temporal effects of the adaptation process

First, we discuss the difference between the effects of the spatiotemporal integration of velocity information realized by the recurrent filter *without* (see Fig. 2 (C)) and *with* (see Fig. 2 (D),(E),(F)) additional adaptation of the parameters of the velocity detecting units. For this purpose, we start with a moving bar as visual stimulus with a motion direction of 45° in relation to the bar orientation and a constant speed of one pixel per frame as shown in Fig. 2 (A), similar to the experiments done by Pack and Born [14]. With such a stimulus, the velocity detecting units (neurons) that are responsible for the motion estimations along the edge of the bar are heavily faced with the aperture problem [15]. To recover the motion from the population of neurons, the population vector method is used [16]. We define the velocity estimation result of all detection units with an identical receptive field at position $\mathbf{x}$ as the mean $\boldsymbol{\mu}_{\mathbf{x}}^t$ of velocity samples $\mathbf{h}$ weighted by the strength of the response (the posterior probability $\alpha(\mathbf{v}_{\mathbf{x}}^t = \mathbf{h}_n^t; \mathbf{H}^t)$):

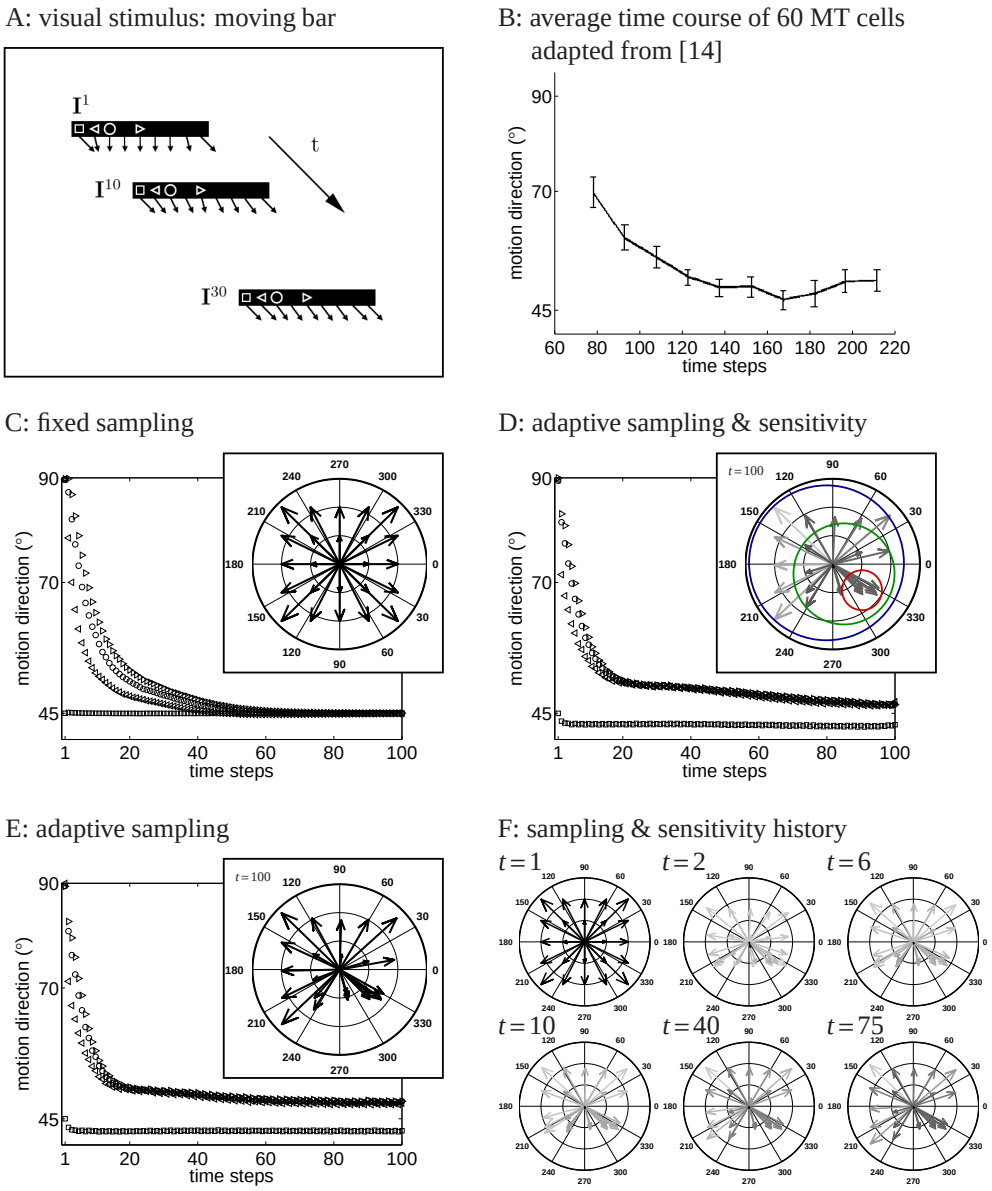

Figure 2: (A) Snapshots $\mathbf{I}^t$ at different points in time t of a moving bar. Results for four different positions along the bar ($\square, \triangleleft, \bigcirc, \triangleright$) are given. The black arrows show the velocity estimations correspondent to simulation (D). (B) Average time course of the activity of 60 MT cells of macaque brain for the stimulus shown in (A), adapted from [14]. (C) Simulated time course of the velocity detection units responsible for marked positions ($\square, \triangleleft, \bigcirc, \triangleright$). The polar plot shows the fixed sampling of velocity space by 25 units with constant and equal sensitivity. (D) Simulations for the same units as in C but now they are able to adapt their preferred velocity and sensitivity. The polar plot shows the adaptation result after 100 time steps. Each *arrow* indicates the velocity a unit is most responsive for and the *gray value* visualises the sensitivity (the brighter, the broader the tuning). (F) shows the history of adaptation at times $t = 1, 2, 6, 10, 40, 75$ correspondent to (D). (E) Same simulation as in (D) but *without* sensitivity adaptation.

$\boldsymbol{\mu}_{\mathbf{x}}^t = \sum_n \alpha(\mathbf{v}_{\mathbf{x}}^t = \mathbf{h}_n^t; \mathbf{H}^t)\mathbf{h}_n$ [1]. If we assume a fixed sampling of the velocity space with 25 velocity hypotheses the recurrent filter is able to completely disambiguate the motion ambiguities caused by the aperture problem by integrating motion information from different points in space and time using spatiotemporal transition probabilities as defined in (11). This is shown in Fig. 2 (C) by the time course of $\mu_{\mathbf{x}}^t$ for four different positions along the bar $\mathbf{x} = \{\square, \triangleleft, \bigcirc, \triangleright\}$. After ≈ 70 time steps the velocity estimates of all positions along the bar converge to the true motion. Similar results without incorporating adaptation have also been shown by other models, e.g. [10], using a non-probabilistic recurrent approach. However, the presented probabilistic model enables to include adaptation in a formalised way. If we now allow the velocity detection units to adapt their parameters according to (41) and (52), the sampling of the velocity space, the sensitivity of the detection units and the time courses of the velocity estimates $\boldsymbol{\mu}_{\mathbf{x}}^t$ change as shown in Fig. 2 (D). One can see, that the detection units tuned to velocities surrounding the movement of the stimulus adapt their velocity tuning to cluster around the true velocity of the moving bar (red circle in Fig. 2 (D)). Hence, the velocity resolution around this cluster center increases. Units that are tuned to other velocities that are not able to explain the stimulus movement only slightly adapt their velocity tuning. In the direct surrounding of the stimulus movement (red circle in Fig. 2 (D)) the tuning widths become more narrow, in the near surrounding (green circle in Fig. 2 (D)) some tuning widths become even more narrow (some less narrow) compared to the direct surrounding, and in the far sourrounding (blue circle in Fig. 2 (D)) the tuning widths broaden. Looking at the temporal history of the adaptation process in Fig. 2 (F) it can be seen, that within the first $t = 5, \ldots, 10$ time steps the *velocity tuning* $\mathbf{h}_n^t$ adapts and afterwards ($t = 5, \ldots, 75$) the *sensitivity range* σ_n^t adapts. After $t \approx 80$ no adaptation happens any more as long as the stimulus does not change its movement. This adaptation process also influences the temporal history of the velocity estimations as can be seen in Fig. 2 (D). The estimations converge much faster (after $t \approx 20$ compared to $t \approx 70$ for a fixed sampling) to one common estimate for three different bar positions $\mathbf{x} = \{\triangleleft, \bigcirc, \triangleright\}$ but under-estimate the motion direction about 4°. Between $t = 20, \ldots, 80$ the estimates keep being together and decrease the under-estimate of the motion direction because the differences in the tuning widths of the detection units become more and more pronounced. After convergence a slight under-estimate of about 2° remains. At the beginning the tuning widths are equal $\sigma_{1:N}^1 = \sqrt{0.02}$ (which is also chosen as the prior $\sigma_{1:N}^0 = \sqrt{0.02}$) and after 100 time steps they differ between $\sigma_{min}^{100} = \sqrt{0.005}$ and $\sigma_{max}^{100} = \sqrt{0.04}$. For position $\mathbf{x} = \square$ the population coding now slightly over-estimates the true velocity about $-2°$ because the samples are no longer symmetrically distributed around the true velocity. In (E) the same simulation as in (D) has been executed but *without* tuning width adaptation. The sampling converges to the same distribution in velocity space as in (D) but the velocity estimates converge to slightly different fixed points with increased under-/over-estimates of $3.5°/2.3°$. Such estimation offsets with a mean difference of $5.3° \pm 10.5°$ has also been observed in the experiments by Pack and Born [14] measuring the time course of the response of velocity sensitive MT neurons of macaque brain. A copy of the result of their experiments is shown in Fig. 2 (B). As shown in Fig. 2 (D), our spatiotemporal integration framework in conjunction with the proposed adaptive velocity tuning method is able to simulate such temporal behavior in a qualitative way. It is important to mention, that the choice of the parameters $\lambda_{\mathbf{h}}$, λ_σ, $\mathbf{H}_0$ and Σ_0 influences the temporal

[1] In the terminology of the introduction, such a population-based estimation constitutes an *aware* decoder, since it relies on the adapted tuning curve centers $\mathbf{h}_n$. An *unaware* decoder would then correspondingly calculate the estimated velocity using $\boldsymbol{\mu}_{\mathbf{x}}^t = \sum_n \alpha(\mathbf{v}_{\mathbf{x}}^t = \mathbf{h}_n^t; \mathbf{H}^t)\mathbf{h}_{0,n}$.

history of both the adaptation and the estimations.[2]

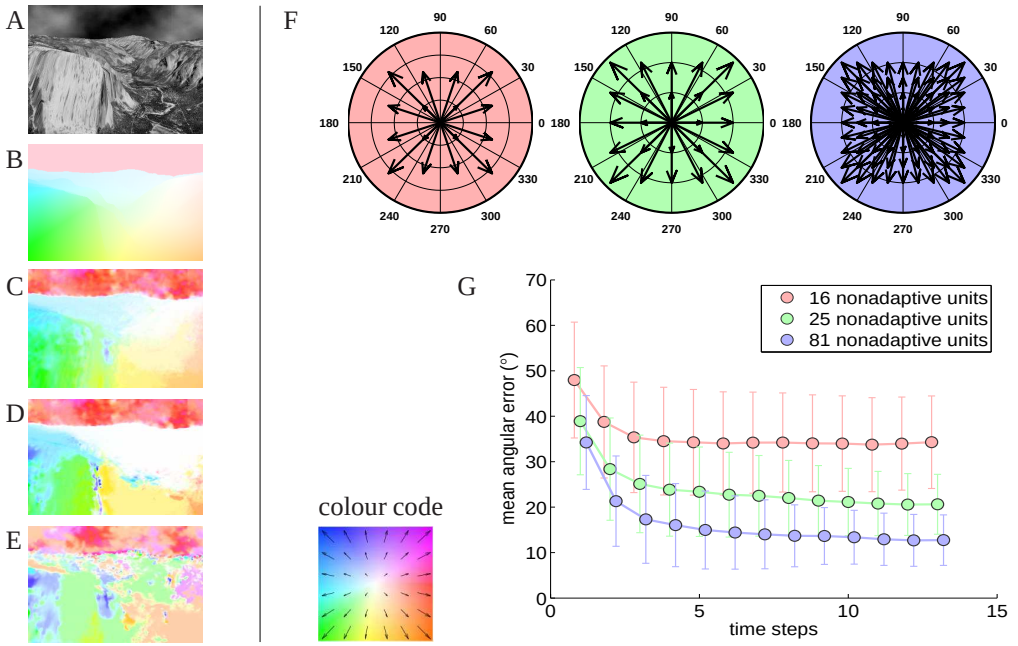

Figure 3: (A) The first frame of the Yosemite sequence. (B) The grund truth optical flow for frame 13. The colors visualise different movement directions and the saturation visualises the speed amplitude (the more saturation, the larger the amplitude). (C-E) Simulation results of the optical flow for 81, 25 and 16 velocity detection units per image location. (F) Polar plots of the fixed sampling distribution in velocity space for three different numbers of detection units 16 (red), 25 (green) and 81 (blue). Each arrow shows a different sampling location. The absolute value of an arrow corresponds to the speed and the direction of an arrow equals the motion direction. (G) Time course of the mean angular error and the corresponding standard deviations shown by the error bars for three different numbers of detection units 16 (red), 25 (green) and 81 (blue). See text for further details.

4.2. Sampling number versus accuracy

Next, we analyse the accuracy of the estimation for fixed and adaptive velocity samples and responsiveness. For fixed $\mathbf{H}$ and Σ but different numbers of detection units the results are shown in Fig. 3. We compare the accuracy of the motion estimation for the Yosemite sequence as shown in Fig. 3 (A) using fixed arrays of 81 (9x9) (blue), 25 (5x5) (green) and 16 (4x4) (red) velocity detectors at each position $\mathbf{x}$ covering the velocity space as shown in Fig. 3 (F). As a measure of accuracy we use the mean angular error between ground truth optical flow (see Fig. 3 (B)) and estimated flow as defined in [13]. Along 13 frames of the Yosemite sequence (see Fig. 3 (G)), the performance increases along time because of the spatiotemporal couplings of the recurrent filter, but drops for low numbers of velocity samples because of a too coarse coverage of velocity space. The corresponding flow results after 13 time steps for 81, 25 and 16 samples is shown in Fig. 3 (C), (D) and (E), respectively.

[2]Here, we have chosen: $\lambda_{\mathbf{h}} = 5$, $\lambda_\sigma = 400$, $\Sigma_0 = \sqrt{0.02}\mathbf{I}$ and $\mathbf{H}_0$ equally distributed on a square grid between $\mathbf{h}_1 = (-2,-2)^\top$ and $\mathbf{h}_{25} = (2,2)^\top$.

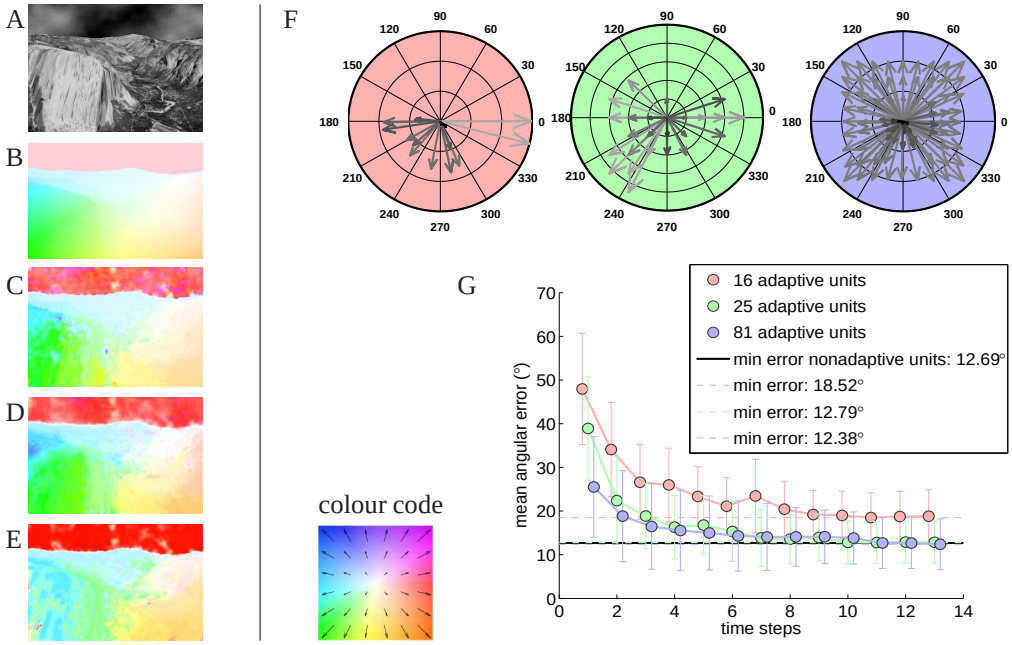

Figure 4: (A) The first frame of the Yosemite sequence. (B) The grund truth optical flow for frame 13. The colors visualise different movement directions and the saturation visualises the speed amplitude (the more saturation, the larger the amplitude). (C-E) Simulation results of the optical flow for 81, 25 and 16 adaptable velocity detection units per image location. (F) Polar plots of the adaptable sampling distribution in velocity space for three different numbers of detection units 16 (red), 25 (green) and 81 (blue) at time step 13. Each arrow shows a different sampling location. The absolute value of an arrow corresponds to the speed and the direction of an arrow equals the motion direction. The gray values visualise the sensitivity (the brighter, the broader the tuning). (G) Time course of the mean angular error and the corresponding standard deviations shown by the error bars for three different numbers of detection units 16 (red), 25 (green) and 81 (blue). The dashed lines show the mean angular errors for different sampling numbers after convergence. See text for further details.

Now, we allow the parameters $\mathbf{H}$ and Σ to be adaptive and again have a look at the time course of the mean angular error for different numbers of *adaptable* detection units. As can be seen in Fig. 4 (G) the adaptation mechanism leads to velocity samplings (see Fig. 4 (F)) that smartly cluster around the velocities that best fit to the flow statistics of the scene. Within about 10 frames, the system converges to small angular errors. For 81 adaptable detection units the mean angular error 12.38° slightly decreases, compared to the error 12.69° for 81 non-adaptable detection units (the baseline shown as a black line in Fig. 4 (G)). For 25 adaptable detection units the mean angular error 12.79° only slightly increases compared to 12.69° (see Fig. 4 (G)), although the number of samples has been reduced dramatically by the factor of $81/25 = 3.24$ (see Fig. 4 (F)). Not before the sampling number is reduced even more, like 16 samples, the accuracy breaks down to a mean angular error of 18.52° but is still much better than the results for 25 fixed samplings (compare to Fig. 3 (G)). This is also reflected in the results of the flows for differently sampled motion detectors shown in Fig. 4 (C), (D) and (E) for the sample numbers 81, 25 and 16, respectively. What can also be observed is, that the adaptation leads to a spread of the sensitivity (i.e., the σ_n's) of the detection units. Sometimes, this happens in a way that the more samples cluster around some region the more the sensitivity decreases as shown in Fig. 4 (F). To

conclude, the reduction of sampling numbers by a factor of $f \approx 3$ does not affect the accuracy significantly, if the velocity detection units are allowed to adapt. Nevertheless, the computational costs are reduced by the factor f^2 since the number of state variables $\mathbf{v_x}$ to be predicted to the next time step influences the computational complexity quadratically $X \times N^2$.

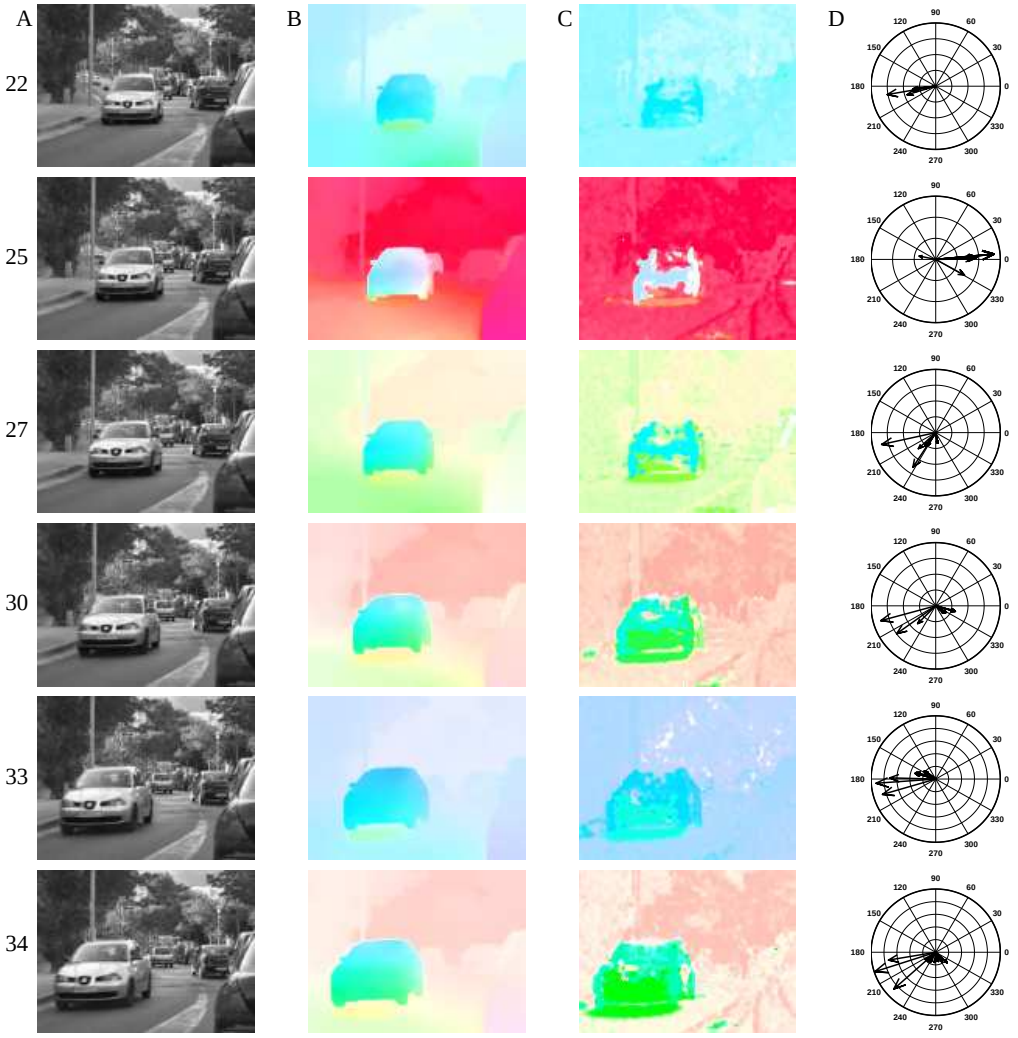

Figure 5: Optical flow results for a real world sequence at different timesteps (column A) with ground truth optical flow (column B), estimated flow using the proposed adaptive filter (column C) and adaptation history of the velocity tuning along time (column D). The different colors in (B) and (C) visualise different movement directions and the saturation visualises the speed amplitude (the more saturation, the larger the amplitude). The different arrows in (D) visualise the currently optimal velocity samples reflecting the motion statistics given by the input. See text for further details.

4.3. Multiple motions and movement changes

Finally, we tested the adaptive filter for optical flow estimation with a real world sequence taken from a moving car [17] shown in Fig. (5) (A) (selected frames from the interval {22, ..., 34}

are shown). The ground truth optical flow shown in Fig. (5) (B) consists of the movement of the static scene induced by the camera motion and the motion of the approaching car. Due to the rapid changes of the camera motion the optical flow that corresponds to the background changes quite heavily from frame to frame. The motion amplitude of the flow of the approaching car is quite large compared to the amplitude of the background motion. Additionally, the amplitude of the car flow increases quadratically because the shorter the distance of an object is, the larger is the amplitude of the projected movement[3] and the larger the image region is that is covered by the object flow. Despite of this highly dynamic image motion, Fig. (5) (C) shows that the proposed adaptive filter is capable of estimating the optical flow using only 25 adaptable velocity-detection units. This is also reflected in the time history of the sampling of the velocity space seen in Fig. (5) (D). The adaptation keeps up with the changes of the motion statistics and gathers around the preferred motions of the car and the background. Since the movement of the background is small compared to the car motion the samplings responsible for the background gather around zero velocity. Hence some are not visible in the plots of Fig. (5) (D) because the amplitude is to small.

5. Discussion

5.1. Modeling level

We present a functional model of tuning curve adaptation for motion estimation units. No attempt was made to model the single units with neurobiologically realistic units, as could e.g. be done using models of spiking neurons. The reason is that we introduced the adaptation dynamics as a consequence of an online optimization on the parameters describing the neuronal tuning curves of each unit, in order to directly show the implications of the optimization process on these parameters (the motion tuning curve center and width) during the adaptation. Nevertheless, the considerations remain valid irrespective of a detailed neuronal model, if the single motion detectors exhibit a smoothly varying response that is dominated by their underlying motion measurement and the adaptation drive. This is generally the case if the processing units are not single neurons, but rather neuronal populations. Other authors (see e.g. [22]) have proposed more detailed functional models for the motion estimation cascade through V1 and MT, describing a motion measurement process including contrast gain control, but their models do only address contrast gain adaptation.

5.2. Physiological plausibility

Our model can be interpreted in terms of a connectionist neural network composed of artificial "neurons" for each motion estimation unit, i.e., for each unit parameterized by a retinal position $\mathbf{x}$ of its receptive field, a tuning motion vector $\mathbf{h}_n$ and a tuning curve width σ_n (following the notation from Eq. (14)). Each motion estimation unit would be composed of a subpopulation of several biological neurons with similar tuning properties, and the unit response would be the averaged subpopulation response. In fact, the model has been implemented as a network with as many neuronal units as image positions times the number of velocity sampling points, which amounts e.g. to $316 * 252 * 81 = 6450192$ units for the Yosemite sequence when using 81 different motion sampling points at each position. Each unit receives mainly two types of inputs: A

[3] This is similar to a nonlinear accelerating object that moves front-to-parallel to a static camera.

driving input from sensory motion measurement that provides the amount of evidence from its receptive field for the motion that the unit is tuned to, and lateral inputs that convey information about the estimation confidence of the other units. The final motion estimation is contained in the population of all different motion estimating units at each position, and extracted using a population code.

5.3. Experimental evidence

The clearest benefit of adaptation is increased sensitivity. For input contrast adaptation, this has been discussed in extent in numerous publications (see, e.g., [28] for a review). As an example, light adaptation in the retina [23] enables to cover light intensity ranges that vary over several orders of magnitude with detectors that are severely limited in their dynamic range. A similar effect can be observed for adaptation-induced changes in contrast gain in motion area MT, see e.g. [24] and [30], [31]. The findings support the notion of an adaptation mechanism that occurs in MT and which emerges from intracortical interaction within MT and that depends specifically from the tuning of the neurons that provide the inputs for the adaptation [30], in analogy to the lateral propagation of information in our model. Physiological evidence of adaptation of motion tuning curve parameters (motion direction, speed and tuning curve width) have only recently been reported by a selected number of publications, see [5, 32, 6]. In [32], it was found that MT neurons exhibit speed tuning adaptation, and that this adaptation enhances speed discrimination around the speeds that currently prevail in the environment. To the contrary, the psychophysical evidence for improved perceptual performance as a consequence of response adaptation is less clear. Adaptation, however, has been hypothesized to be the source for a number of well measurable perception aftereffects [1, 2].

5.4. Adaptation and coding efficiency

Grounded on our approach, we propose to consider adaptation as an effect of the tendency of a neural system to optimally cover a large sensory range with the available limited resources. In the case of motion estimation, the need to cover a large range of motion directions and velocities at all retinal positions rapidly leads to a huge number of necessary motion detection units. We have shown that with adaptation, this number can be considerably reduced (e.g. by a factor of 3 without losing motion estimation performance, see Fig. 4). In this sense, the adaptation takes advantage of large-scale statistical properties of the simuli (i.e., properties that cannot be measured by single units) to enable the system to improve coding efficiency, a principle that has been orignally postulated by [26]. However, in our model an improved coding efficiency means that the estimation of the *stimulus properties* is improved. In future, it might be interesting to also include further criteria like metabolic efficiency (see e.g. [25] for a review on this topic and [28] for the relation to visual adaptation) into the optimization process. In our model, the consequences of the adaptation are that the activities in the population code get more pronounced, i.e., units with motion tuning matching to the input increase their activity whereas units with non-matching tuning decrease it, maintaining on average a constant activity level over the population at each retinal position, so that metabolic efficiency also would remain constant.

6. Summary

In this work, we have presented a theoretical derivation and a system implementation of a model for motion estimation that automatically adapts the tuning of its motion detectors on a

short timescale. It is based on the idea of incremental motion estimation, making use of a probabilistic interpretation of neuronal dynamics. The model consists of "populations" of motion detectors for each retinal position that encode motion estimation probabilities for a discretely sampled set of velocities. The motion detectors are recurrently connected and propagate their local estimations according to a spatiotemporal consistency constraint. For adaptation, the sampling locations in velocity space as well as the tuning curve widths are tuned systematically to better cover the motion measurement.

The formal basis of the presented work is 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 show that such a system is able to rapidly adjust (within a few frames) the tuning of its velocity detectors, allowing to achieve comparable motion estimation 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.

In our work, for simplicity we have assumed that all units from the motion detecting population make use of *the same* set of discrete, but adaptable motion tuning parameters $\mathbf{H}, \Sigma$, even if they are located at different retinal positions. This means, that all neurons feel the entire motion statistics and adjust accordingly. In a more plausible setting, this assumption would have to be relaxed, allowing e.g. units at very distant locations to tune independently.

Acknowledgements We would like to thank Valentina Ansel for preparing the figures shown in the introduction and Thomas Gußner and Jörg Deigmöller for proof-reading.

References

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

[2] 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.

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

[4] H. Takemura and I. Murakami, Motion detection sensitivity enhanced by induced motion, *Journal of Vision*, 8(6), 2008.

[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] U. Hillenbrand, Spatiotemporal adaptation in the corticogeniculate loop. *PhD Thesis TUM*, 2001.

[8] 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.

[9] B. Jaehne, Digitale Bildverarbeitung. *6. Edition, Springer-Verlag*, 2005.

[10] P. Bayerl and H. Neumann, Disambiguating Visual Motion through Contextual Feedback Modulation. *Neural Computation*, 16(10):2041-2066, 2004.

[11] P. Burgi, A. Yuille and N. Grzywacz, Probabilistic Motion Estimation Based on Temporal Coherence. *Neural Computation*, 12:1839-1867, 2000.

[12] P. Bayerl and H. Neumann, A Fast Biologically Inspired Algorithm for Recurrent Motion Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2):246-260, 2007.
[13] J. Barron, D. Fleet and S. Beauchemin, Performance of Optical Flow Techniques. *International Journal of Computer Vision*, 12:43-77, 1994.
[14] C. Pack and R. Born, Temporal Dynamics of a Neural Solution to the Aperture Problem in Visual Area MT of Macaque Brain. *Nature*, 409:1040-1042, 2001.
[15] C. Pack, M. Livingstone, K. Duffy and R. Born, End-stopping and the Aperture Problem: Two-dimensional Motion Signals in Macaque V1. *Neuron*, 39:671-680, 2003.
[16] A. Georgopoulos, A. Schwartz and R. Kettner, Neural Population Coding of Movement Direction. *Science*, 233(4771):1416-1419, 1986.
[17] C. Liu, W. T. Freeman, E. H. Adelson and Y. Weiss, Human-Assisted Motion Annotation, *IEEE Conference on Computer Vision and Pattern Recognition*, http://people.csail.mit.edu/celiu/motionAnnotation/index.html, 2008.
[18] V. Willert, M. Toussaint, J. Eggert and E. Körner, Probabilistic Exploitation of the Lucas and Kanade Smoothness Constraint, *International Conference on Machine Learning and Applications*, 259-266, 2008.
[19] B. D. Lucas and T. Kanade, An Iterative Image-Registration Technique with an Application to Stereo Vision, *International Joint Conference on Artificial Intelligence*, 674-679, 1981.
[20] C. M. Bishop, Pattern Recognition and Machine Learning, *Springer Science+Business Media*, 2006.
[21] Y. Weiss and D.J. Fleet, Velocity likelihoods in biological and machine vision, *Probabilistic Models of the Brain: Perception and Neural Function, MIT Press*, 77-96, 2002.
[22] N. C. Rust and V. Mante and E. P. Simoncelli and J.A. Movshon, How MT cells analyze the motion of visual patterns, *Nature Neuroscience*, 9(11):1421-1431, 2006.
[23] R. Shapley and C. Enroth-Cugell, Visual adaptation retinal gain control, *Prog Ret Res*, 3:263-346, 1984
[24] A. Kohn and J. A. Movshon, Neuronal Adaptation to Visual Motion in Area MT of the Macaque, *Neuron*, 39:681-691, 2003
[25] D. Attwell and S.B. Laughlin, An Energy Budget for Signaling in the Grey Matter of the Brain, *Journal of Cerebral Blood Flow and Metabolism*, 21:1133-1145, 2001.
[26] H. B. Barlow, Possible principles underlying the the transformations of sensory messages, in *Sensory Communication* (ed. Rosenblith, W. A.), 217-234, 1961
[27] D.A. Gutnisky amd V. Dragoi, Adaptive coding of visual information in neural populations, *Nature*, 452:220-225, 2008.
[28] A. Kohn, Visual Adaptation: Physiology, Mechanisms, and Functional Benefits, *Journal of Neurophysiology*, 97:3155-3164, 2007.
[29] N.J. Priebe, C.R. Cassanello and S.G. Lisberger, The Neural Representation of Speed in Macaque Area MT/V5, *The Journal of Neuroscience*, 23(13):5650-5661, 2003.
[30] N.J. Priebe, M.M. Churchland and S.G. Lisberger, Constraints on the Source of Short-Term Motion Adaptation in Macaque Area MT. I: The Role of Input and Intrinsic Mechanisms, *Journal of Neurophysiology*, 88:354-369, 2002.
[31] N.J. Priebe and S.G. Lisberger, Constraints on the Source of Short-Term Motion Adaptation in Macaque Area MT. II: Tuning of Neural Circuit Mechanisms, *Journal of Neurophysiology*, 88:370-382, 2002.
[32] B. Krekelberg, R.J.A. van Wezel and T.D. Albright, Adaptation in Macaque MT Reduces Perceived Speed and Improves Speed Discrimination, *Journal of Neurophysiology*, 95:255-270, 2006.