

Neural Mechanisms for Mid-Level Optical Flow Pattern Detection

Stefan Ringbauer, Pierre Bayerl, Heiko Neumann

Ulm University, Inst. for Neural Information Processing, Ulm, Germany

Abstract. This paper describes a new model for extracting large-field optical flow patterns to generate distributed representations of neural activation to control complex visual tasks such as 3D egomotion. The neural mechanisms draw upon experimental findings about the response properties and specificities of cells in areas V1, MT and MSTd along the dorsal pathway. Model V1 cells detect local motion estimates. Model MT cells in different pools are suggested to be selective to motion patterns integrating from V1 as well as to velocity gradients. Model MSTd cells considered here integrate MT gradient cells over a much larger spatial neighborhood to generate the observed pattern selectivity for expansion/contraction, rotation and spiral motion, providing the necessary input for spatial navigation mechanisms. Our model also incorporates feedback processing between areas V1-MT and MT-MSTd. We demonstrate that such a re-entry of context-related information helps to disambiguate and stabilize more localized processing along the primary motion pathway.

Keywords: Motion perception, optical flow, motion integration, motion gradient, feedback, navigation, area MT, area MSTd.

1 Introduction

The robust analysis of optical flow patterns is one of the basic computational tasks for steering egomotion of human and higher animals [9]. Self-motion induces typical patterns of optical flow on the retina which are analyzed by the visual system over several stages of hierarchically organized visual areas in the dorsal cortical pathway. The two cortical areas MT (medial temporal) and MSTd (dorsal medial superior temporal) are primarily concerned with the processing of optical flow. Cells in area MT represent proper features of optical flow information, e.g., motion direction and speed, in medium-size regions of the visual field [1]. MT cells project to area MSTd where cells have huge receptive fields and are tuned to specific patterns of optical flow. While [8] found evidence for a dominance of cells selective to radial and rotational flow (defining an orthogonal basis), the investigation of [10] found support for the existence of a continuum of flow sensitive cells with equally likely preference to spiral motion patterns.

Previous approaches to modeling optical flow extraction have been suggested which consider MSTd cells sensitive to large-field optical flow. For example, [4] investigated the development of optical flow sensitive cells using a two-layer

backpropagation network. [11] studied optical flow processing by MSTd cells using a spatial retino-cortical mapping with log-polar magnification. In this model the output from direction selective MT cells were spatially integrated by MSTd cells with different selectivities that are fed forward to a heading map. The model by [17] was the first incorporating the functionality of speed gradient cells (at MT). Different pools of MSTd neurons spatially integrate motion responses and speed gradient responses separately to generate cells tuned to complex motion patterns.

Our model extends previous models drawing upon our own previous investigations on recurrent feedforward-feedback processing for integration and disambiguation of motion signals. Here we propose a new model mechanism for extracting changes in velocity by employing cells with asymmetric receptive field profiles oriented along the motion direction. This extends previously proposed speed gradient mechanisms by computing velocity changes. MSTd in turn integrates information signaled by MT gradient cells to form large-field motion cells selective to complex motion patterns. We demonstrate that MSTd gradient cells show position-independence properties with respect to variations of the location of centers of motion. Furthermore, we also incorporate predictive feedback from MSTd to MT that stabilizes the extraction of noisy complex motion patterns even in the presence of moving objects.

2 Mid-level optical flow pattern detection

2.1 Input sequences and initial motion detection

Our approach draws upon experimental evidence about the structure and function of the primary stages of the dorsal pathway in visual cortex. Initial motion detection is realized at the stage of the primary cortical area V1, while visual area MT focuses on the robust detection of salient motion components, tracking of localized features, and the disambiguation of locally ambiguous motion patterns that were caused, e.g., by the aperture problem [13]. Visual area MSTd is primarily concerned with the detection and representation of large whole field flow patterns, such as those that were generated by observer self-motion.

Local motion detection is accomplished by a Reichardt-like correlation scheme with shunting inhibition between half-detectors tuned to opposite movement directions. These activities feed into a feedforward-feedback loop modeling V1-MT interaction [2]. In a nutshell, motion sensitive cells in model MT integrate activities from V1 direction selective cells by pooling same velocities over a spatial as well as directional neighborhood, essentially low-pass filtering the noisy motion estimates. The resulting activities are represented in a pool of motion cells in area MT. Pooled activities $a_{\mathbf{u}}^{MT}$ are subsequently fed back to enhance V1 cell activations that are compatible with the velocity represented in model MT. The resulting activations were also fed forward into a population of MT model cells that compute *changes* in the MT velocity field representation. This processing and subsequent integration of these activities by large-field model MSTd cells are at the primary focus of this contribution and are described in subsection 2.2.

The dynamics of each model area is described by a cascade model of linear-non-linear-non-linear (LNN) processing [12] derived from single-compartment neuron models with firing-rate activation dynamics. We utilize three-stages of (a) input filtering, (b) non-linear signal amplification (via feedback), and (c) shunting inhibition to realize activity normalization in a local neighborhood. This basic architecture has been utilized in previous models for static form as well as dynamic motion perception [14][2].

2.2 Detection of motion patterns and gradients

The V1-MT motion detection scheme is extended in several ways, namely (a) a neural mechanism to compute changes in the dense velocity representation (motion gradients) at the level of the MT/MSTl complex, (b) a stage of integrating gradient signals over a large spatial neighborhood to generate cell responses in model area MSTd, and (c) incorporating feedback to deliver a re-entry signal for stabilizing the processing of motion gradients. Similar to the approach developed by Tsotsos and colleagues [17] we employ mechanisms to measure local changes in velocity generating a gradient representation. Experimental evidence for the existence of cells sensitive to velocity gradients has been presented by, e.g., [16], and the spatial pattern of surround weighting was studied in [18].

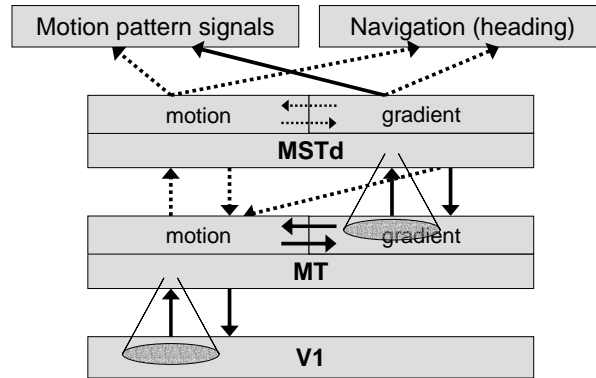


Fig. 1. General outline of processing stages and communication pathways of the model. Bold arrows denote information flow currently implemented in the model, dashed arrows depict those connectivities that are planned to be included in further model extensions. Representation of *motion* in MT and MSTd denotes a location-velocity space, $\langle x, y, \theta, s \rangle$, while the *gradient* representations denote a location-velocity-gradient space, $\langle x, y, \Delta\theta, \Delta s \rangle$.

The primary input to area MSTd is delivered via area MT where it is integrated over a large spatial neighborhood to generate the observed pattern specificity to large-field motion patterns. Physiological investigations have demonstrated that cells in area MSTd respond primarily to large field motion patterns having different selectivities, e.g., expanding or contracting radial motion, clockwise or counter-clockwise rotation, or linear superpositions that lead to spiral motion patterns [10]. A

sub-population of cells also responds to pure translatory motion. MSTd activation in turn is fed back to cells in area MT to incorporate prediction and enhancement of noisy activation distributions. In all, we suggest a coherent architecture in which the bidirectional signal flow defines a key component of functionality to achieve robust cortical motion representation (Fig. 1). Size ratios between cells in the different model cortical areas are defined in accordance to experimental data [6, 10] and were set to V1:MT:MSTd = 1:5:25.

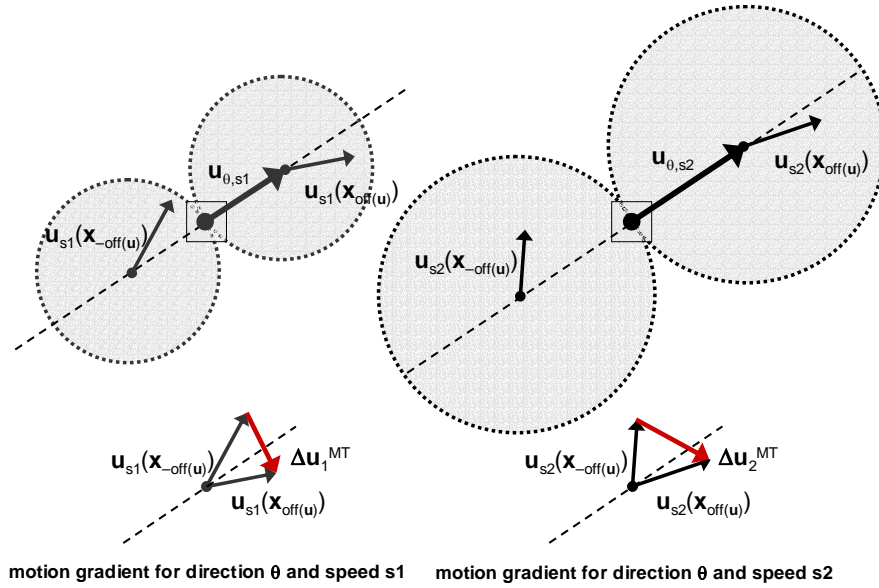


Fig. 2. Mechanism of computing differences in velocity (gradient) in model area MT for two different speed amplitudes in the same direction (left: slow speed, right: fast speed). The size of sub-field integration scales as a function of the speed amplitude (see text).

Detection of changes in MT motion fields (motion gradients). Motion is encoded by populations of cells sensitive to direction and speed, i.e., $\mathbf{u} = (\theta, s)$. In order to keep the computational efforts in reasonable bounds we employ a rank-order coding approach [15] in which spike sequences are generated algorithmically and represented in ordered lists. This allows highly efficient on-demand representation of motion in various directions and allowing the detection of arbitrary speed amplitudes (see [3] for details).

We propose a scheme to measure changes in the velocity field along local motion directions in MT, which are represented in a second pool of MT neurons, $a_{\Delta\mathbf{u}_\theta}^{MT}$, where $\Delta\mathbf{u} = (\Delta\theta, \Delta s)$ symbolizes changes in velocity. The population of motion sensitive cells is sampled at each location by two sub-fields that are spatially offset along the movement direction at the target cell. For a given motion along direction θ and speed $s = |\mathbf{u}|$, sampling locations of cells are at $\mathbf{x}_{\text{-off}(\mathbf{u})} = \mathbf{x} - \mathbf{u}_\theta$ and $\mathbf{x}_{\text{off}(\mathbf{u})} = \mathbf{x} + \mathbf{u}_\theta$, respectively (\mathbf{x} denoting the spatial target location). The radius of the sampling

kernels varies with the speed of the current velocity, i.e. $r = |\mathbf{u}|$, such that we get an increased spatial uncertainty with increasing speed (see Fig. 2). In each sub-field the velocities are weighted and summed to derive a population response for the velocity. The differences between velocities of the populations from both offset locations are defined by the quantity¹

$$\Delta \mathbf{u}^{MT}(\mathbf{x}) = \mathbf{u}(\mathbf{x}_{off(\mathbf{u})}) - \mathbf{u}(\mathbf{x}_{-off(\mathbf{u})}). \quad (1)$$

We specify cells to represent certain $\Delta \mathbf{u}$ which compute their activity utilizing the following non-linear mechanism²,

$$a_{\Delta \mathbf{u}}^{MT}(\mathbf{x}) = a_{\mathbf{u}}^{MT}(\mathbf{x}) \cdot a_{\mathbf{u}}^{MT}(\mathbf{x}_{off(\mathbf{u})}) \cdot a_{\mathbf{u}}^{MT}(\mathbf{x}_{-off(\mathbf{u})}) \quad (2)$$

to generate a field of motion gradients. These activities are efficiently approximated similar as in [3]. Each local difference in activity is represented with respect to the direction of motion at the respective target location that defines a local gauge coordinate system (compare [17]). Activities of motion cells as well as of motion gradient cells are subsequently normalized by a process of divisive (shunting) center-surround competition in the space-feature domain. For simplicity, at the moment, we employ normalization over gradient cell activities at single locations, namely

$$a_{\Delta \mathbf{u}_\theta}^{MT}(\mathbf{x}) = a_{\Delta \mathbf{u}_\theta}^{MT}(\mathbf{x}) / \sum_{all\ a_{\Delta \mathbf{u}}(\mathbf{x})} a_{\Delta \mathbf{u}}^{MT}(\mathbf{x}). \quad (3)$$

The resulting activations are then fed forward to cells in model area MSTd.

MSTd gradient motion integration. Cells in model MSTd sum up the responses of corresponding MT gradient units over a large spatial neighborhood similar as in previous models [4][11][17]. The mechanism utilizes a convolution by a suitable kernel to weight activities in the spatial-feature domain

$$a_{\Delta \bar{\mathbf{u}}_\theta}^{MSTd}(\mathbf{x}) = \sum_{\mathbf{x}', \phi} a_{\Delta \mathbf{u}_\phi}^{MT}(\mathbf{x}') \cdot \Lambda_{\mathbf{x}\mathbf{x}'} \cdot \Psi_{\theta\phi} \quad (4)$$

with separable kernels Λ and Ψ for weighting in the spatial and the direction domain, respectively.

The spatial resolution of model area MSTd is down-sampled by a factor of 1:5. In the current version of our model we integrate responses of the MT gradient cells to be represented in a pool of motion cells, $a_{\Delta \bar{\mathbf{u}}_\theta}^{MSTd}$. Those cell responses resulting from non-zero directional differences, $\Delta \theta^{MT}(\mathbf{x}) \neq 0$, encode large-field motion patterns such as radial expansion/contraction, rotation, and spiral motion. For zero directional

¹ To simplify the necessary calculations we employed a vector notation for computing the velocity differences. In order to represent the signals in a neurally plausible scheme direction and speed components could be represented in separate sub-populations of cells with proper sampling resolution. These quantities could then enter into competitive interactions.

² We suggest taking the product of the three measures to calculate a response that denotes a likelihood of the presence of the particular change in velocity. The independence of the three measures is assumed.

differences, $\Delta\theta^{MT}(\mathbf{x}) = 0$, and vanishing speed gradients, $\Delta s^{MT}(\mathbf{x}) = 0$, one gets cells with pure large-field translatory motion pattern selectivity (center of motion shifted to infinity).

Motion pattern prediction and modulatory MSTd-MT feedback. Integrated measures of velocity changes, in turn, provide a context for more localized measures at the earlier computational stages (model area MT in our case). In other words, MSTd activation serves as a predictor signal that is fed back to gradient responses in area MT. This feedback is proposed to be modulatory and excitatory such that activities of cells in MT selective for a particular gradient direction can be amplified by MSTd gradient activation for the same (or similar) gradient direction. The modulation is denoted by the scheme (with constant coefficient C)

$$a_{\Delta\mathbf{u}_\theta}^{MT} \cdot (1 + C \cdot a_{\Delta\bar{\mathbf{u}}_\theta}^{MSTd}). \quad (5)$$

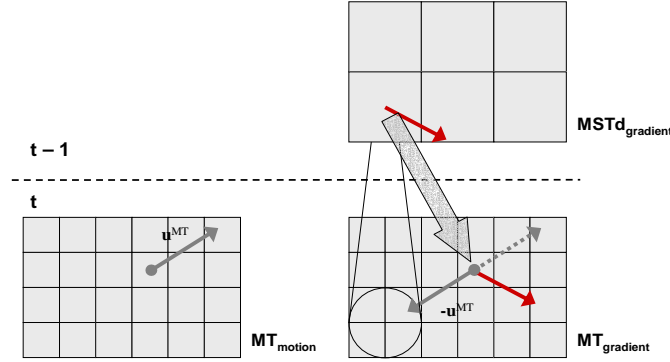


Fig. 3. Sketch of mechanism employed for model MSTd-MT feedback of velocity measures.

The rationale is that MT activations gate the modulation signal so that existing gradient representations will be amplified by MSTd feedback or left unchanged in the case when no feedback signal exists. In case no feedforward activation has been computed at a given location feedback alone cannot generate new activities. This property stabilizes the network behavior in accordance with the no-strong loop hypothesis [7]. The feedback signal amplifies filtered feedforward activations at the second stage of the three-level cascade model briefly sketched above. The amplified responses subsequently undergo center-surround shunting competition (stage three of the cascade model). This realizes a biased competition since those activities in the competitive pool that were amplified now have a stronger bias and consequently reduce the activities of cells which have not received any feedback.

The feedback mechanism is predictive in the sense that the amplification is shifted spatially to a location that coheres with the target motion direction, $\mathbf{u}^{MT}(\mathbf{x})$, in the next frame of the sequence. In a nutshell, the prediction of a gradient measure (with amplitude and direction) utilized “votings” that were spatially shifted according to the velocity \mathbf{u}^{MT} at the target location (see Fig. 3).

3 Simulation results

The neural model has been tested on a variety of input motion sequences. In order to demonstrate its selectivity in the processing of large-field motion patterns and the response distribution of flow sensitive MSTd cells over time we first probe the network with a motion sequence of a simulated flight through a corridor with synthetically generated wall texture patterns (Fig. 4). Here the observer maneuvers such that a continuous sequence of steering commands lead to forward motion followed by left and right turns as well as reversing the spatial movement in between.

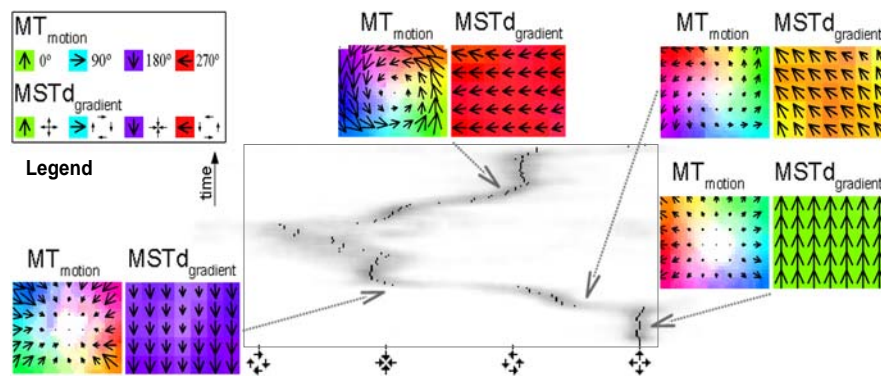


Fig. 4. Results of recurrent V1-MT-MSTd processing for an image sequence from a flight through a tunnel and respective maneuvers over time. The four different pairs of flow field patterns represent the equilibrated motion estimates at different times each showing MT motion responses and MSTd gradient cell responses, respectively (color coded directions represent a mapping of corresponding radial/rotation/spiral patterns shown in the legend top left), response trace of pattern selective MSTd gradient cells over time shown in center (see text).

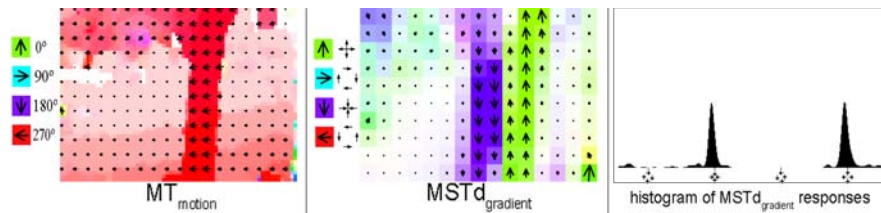


Fig. 5. Results of recurrent V1-MT-MSTd processing for the “Flower garden” sequence. Left: MT motion responses, center: MSTd gradient cell responses, right: histogram of MSTd gradient responses (see text for discussion).

Equilibrated MT motion responses are shown that were generated by V1-MT feedforward/feedback interaction at different times of the temporal sequence and the corresponding results of integrated velocity gradients at the stage of area MSTd (equilibrated response of MT-MSTd feedforward/feedback interaction). The response tuning of large-field motion pattern responses of a continuum of MSTd cells (selective to radial expansion/contraction, clockwise/counter-clockwise rotation, and spiral motion) is shown in the panel at the center. The display of responses of the

motion pattern selective cells with different tunings demonstrates that the temporal course helps to infer the observer motion.

We also tested the model using the “Flower garden” sequence (Fig. 5). This specifically investigates the computational results for scenes containing mutually occluding objects thus generating locations of temporally disappearing structure as well as structure reappearing at locations that were uncovered. The key observations are that (1) dense motion is computed with speeds corresponding to different depths of the scenic objects (including speed gradient in vertical direction of the ground plane), and (2) that motion gradient cells detect occlusion boundaries by signaling opposing motion directions. The gradient directions indicated might be surprising at a first glance but can be explained after a closer look at the mechanisms involved. The signal for “expansion” motion along the right tree boundary is generated by the different velocities (increasing speed) in motion direction from background to the tree region in front which is interpreted as acceleration. At the opposite tree boundary the speed difference indicates a decrease in the velocity (indicative for deceleration) which is signaled by cells tuned to “contraction” motion. A prediction of the model is that MSTd cells when probed by stimulus patterns containing significant depth structure and thus motion parallax should signal similar local pattern motion.

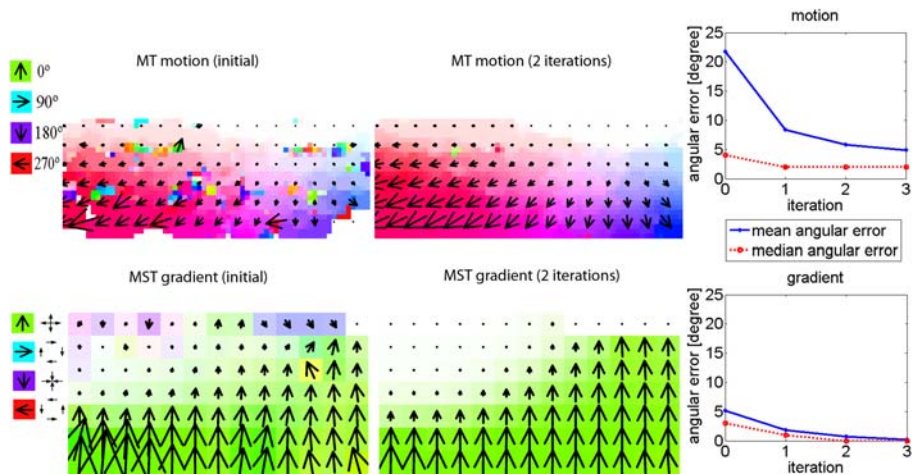


Fig. 6. Results of recurrent V1-MT-MSTd processing for ground plane motion with view direction offset from motion direction. Top: MT responses (initial response, after 2nd iteration), bottom: MSTd gradient cell responses, right panels: improvement of motion and gradient estimates (mean and median angular error) over time (see text).

Finally, we probed the model with the flow pattern generated by forward observer motion over a ground plane with a view direction 15 deg. offset to the left from the translation axis (Fig. 6). This investigates the model properties in navigation tasks with different centers of motion due to varying view directions. The perspective effect of flows on the ground plane poses a problem for the model proposed by Tsotsos and colleagues which employs speed gradient measures for flow pattern extraction [17]. As a consequence, the depth gradient for the ground plane leads to patches of different optical flow field interpretations, namely rotations to both sides of the center

of motion, expansion below and above the center, and spiral motion segments in between. Our approach leads to a response pattern that is invariant against the surface projection. In addition, the study demonstrates how MT flow estimates as well as MSTd pattern representations were improved by the iterative feedforward-feedback interaction. MT motion responses are shown for the (noisy) initial estimates and after 2 iterations as well as the corresponding responses for MSTd gradient cells. The temporal course of angular error reduction is shown over 3 iterations (right panels).

4 Discussion and conclusions

We propose a novel neural architecture of motion detection, integration and the extraction of large field optical flow patterns building upon evidence about cells at different stages along the dorsal pathway of primate visual cortex. Our model makes several new contributions in comparison with previous computational models. In particular, a previous model of early motion detection and integration was extended using the same type of basic mechanisms, namely feedforward integration, modulatory feedback, and shunting competition. Unlike many other models, with the notable exception of the architecture proposed by Tsotsos and coworkers [17], we propose a stage of making explicit velocity changes. Unlike the Tsotsos' model, we propose a speed sensitive scheme of difference filtering between sub-fields along the direction of motion at a target location. Our scheme can be considered as a more generalized approach that also allows detecting direction gradients (in addition to speed gradients) and measuring such velocity gradients in a gauge coordinate frame along the local direction of motion. This idea is reminiscent of the approach for dense texture flow field extraction that has been proposed by Zucker and coworkers [5]. Whereas, Zucker focuses on the long-range lateral connectivities for integration of oriented patterns in static form processing, we are concerned with measuring the direction changes in flow fields. For that reason, we suggested to employ a scheme that utilizes oriented receptive fields with excitatory and inhibitory sub-fields whose sizes scale with the speed of the motion patterns. Unlike the model proposed by Grossberg and colleagues [11] which utilizes feedforward integration only, we employ feedback and integrate velocity gradient information as well. The proposed scheme does not at the moment incorporate learning to automatically develop cells being selective for mid-level motion patterns, such as, e.g., [4].

As we have indicated in Fig. 1 that displays an overview of the computational architecture the extraction of large-field motion patterns serves as an input representation for further computations supporting different behavioral tasks. For example, the estimation of heading is useful for navigation in the spatial environment. In order to reliably extract the heading direction the complex motion field must be somehow decomposed into translatory and rotational flow-field components of different relative amounts. A mixture of expansion and rotation component in flow patterns occurs routinely during fixations of a target object while an observer is moving in a particular direction. Also scenic objects can move in certain directions independently of the observer and must be detected and segmented from the global flow field. We suggest that motion and motion gradient information provide

complementary information since the extraction of global motion pattern signals requires invariance against, e.g., the focus of expansion. On the other hand, estimating the heading direction needs to gain information about the localization of the focus-of-expansion. Both type of information is robustly encoded in the motion and gradient signals at the level of model MSTd. Their proper combination remains a topic for further investigation.

Acknowledgments. This research has been supported in part by a grant from the European Union (EU FP6 IST Cognitive Systems Integrated project: Neural Decision-Making in Motion; project no. 027 198).

References

1. Albright, T.D.: Direction and orientation selectivity of neurons in visual area MT of the macaque. *J. Neurophysiol.* 52 (1984) 1106–1130
2. Bayerl, P., Neumann, H.: Disambiguating visual motion through contextual feedback modulation. *Neural Comp.* 16 (2004) 2041–2066
3. Bayerl, P., Neumann, H.: A fast biologically inspired algorithm for recurrent motion estimation. *IEEE Trans. on PAMI* 29 (2007) 246–260
4. Beardsley, S.A., Vaina, L.M.: Computational modeling of optical flow sensitivity in MSTd neurons. *Network: Comp. Neural Syst.* 9 (1998) 467–493
5. Ben-Shahar, O., Zucker, S.: Geometrical computations explain projection patterns of long-range horizontal connections in visual cortex. *Neural Comp.* 16 (2004) 445–476
6. Born, R.T., Bradley, D.C.: Structure and function of visual area MT. *Ann. Rev. Neurosci.* 28 (2005) 157–189
7. Crick, F., Koch, C.: Constraints on cortical and thalamic projections: The no-strong loops hypothesis. *Nature* 391 (1998) 245–250
8. Duffy, C.J., Wurtz, R.H.: Sensitivity of MST neurons to optic flow stimuli I. A continuum of response selectivity to large-field stimuli. *J. Neurophysiol.* 65 (1991) 1329–1345
9. Gibson, J.J.: *The Ecological Approach to Visual Perception.* LEA, Hillsdale, NJ (1986)
10. Graziano, M.S.A., Anderson, R.A., Snowden, R.: Tuning of MST neurons to spiral motions. *J. Neurosci.* 14 (1994) 54–67
11. Grossberg, S., Mingolla, E., Pack, C.: A neural model of motion processing and visual navigation by cortical area MST. *Cerebral Cortex* 9 (1999) 878–895
12. Herz, A.V.M., Gollisch, T., Machens, C.K., Jaeger, D.: Modeling single-neuron dynamics and computations: A balance of detail and abstraction. *Science* 314 (2006) 80–85
13. Pack C.C., Born, R.T.: Temporal dynamics of a neural solution to the aperture problem in cortical area MT. *Nature* 409 (2001) 1040–1042
14. Thielscher, A., Neumann, H.: Neural mechanisms of cortico-cortical interaction in texture boundary detection: A modeling approach. *Neuroscience* 122 (2003) 921–939
15. Thorpe, S., Delorme, A., VanRullen, R.: Spike-based strategies for rapid processing. *Neural Networks* 14 (2001) 715–726
16. Treue, S., Anderson, R.A.: Neural responses to velocity gradients in macaque cortical area MT. *Vis. Neurosci.* 13 (1996) 797–804
17. Tsotsos, J.K., Liu, Y., Martinez-Trujillo, J.C., Pomplun, M., Simine, E., Zhou, K.: Attending to visual motion. *Computer Vision and Image Understanding* 100 (2005) 3–40
18. Xiao, D.-K., Raiguel, S., Marcar, V., Orban, G.: The spatial distribution of the antagonistic surround of MT/V5 neurons. *Cerebral Cortex* 7 (1997) 662–677