Interactions of motion and form in visual cortex – A neural model

Cornelia Beck*, Heiko Neumann

Institute for Neural Information Processing, University of Ulm, Oberer Eselsberg, 89069 Ulm, Germany

A R T I C L E   I N F O

Keywords:
Neural model
Vision
Motion
Form
Feedback
Chopstick illusion

A B S T R A C T

In this work, we present a neural model simulating parts of the motion and the form pathway of the visual cortex. It is shown how the visual features motion, disparity, and form that are represented in a distributed way in areas V1, V2, and MT mutually interact at several levels. Thus, their information is shared without the need of explicit neural representation for each combination of features. In particular, we address the issue of 2D extrinsic motion cues generated at occlusions that have to be treated differently than 2D intrinsic motion features of the same object. We suggest that here information of the form channel, namely the indication of a junction, is necessary to achieve a correct percept in the motion pathway. Furthermore, we investigated the question of how a percept of either pattern or component motion is generated in a scenario of moving bars that only differs in the presence or absence of occlusions, like in the chopstick and the barberpole display. We propose different roles for various kinds of MT cells that are involved in the interactions with the form pathway, simulating purely integrative cells tuned to motion and to motion and stereo, but also contrast cells responding strongly when motion in the surround is in the opposite direction. The model simulations reproduce psychophysical and neurophysiological results of the chopstick as well as of the barberpole illusion. The temporal course of the dominant motion percept generated by the iterative interplay between motion and form pathway is in line with data of ocular following responses in primates and humans.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Surfaces moving in space are characterized by feature conjunctions of, amongst others, lightness, velocity, and relative depth. The related feature detection processes activate certain representations in the brain. The corresponding structure as well as the possible interaction of the underlying neural processes is a topic of intense investigation. Evidence suggests that early and mid-level processing in primary visual area (V1), secondary visual area (V2), and middle temporal area (MT) of static and dynamic visual information is organized along parallel streams with mutual interactions at different levels (Mishkin et al., 1983). One potential advantage of such an architecture is that features can be coded in a basically distributed fashion while mutual connections between processing stages allow the exchange of information. In MT for example, cells have been reported that are tuned to both motion and disparity (Born and Bradley, 2005). Ponce et al. (2008) showed that when areas in the form pathway like V2 are deactivated, the disparity tuning of these MT cells is largely reduced, which is a clear hint to the interactions between the two pathways. The corresponding roles that different cell types found in MT play in this context are not yet clear. For the perceptual mechanisms involved, there are several computational problems to solve, like the segmentation of the objects of a scene while at the same time integrating regions that belong together. The representation in lower areas should remain spatially localized despite feedback of higher areas with larger receptive fields (RFs). Furthermore, ambiguous motion cues that appear along a 1D structure have to be resolved (“aperture problem”). At 2D structures, where unambiguous motion estimates can be computed, another difficulty appears: If the structure is intrinsic, i.e. belonging to the object, the motion estimated there is reliable and helps to solve the aperture problem. However, if the termination is extrinsic and thus belonging to a different object surface because of an occlusion, the feature response needs to be suppressed, as measured by Shimojo et al. (1989). We suggest that these issues can be solved by appropriate interactions between the two pathways, as presented in the following chapters. Also, the contribution of different kinds of MT cells within these mechanisms is addressed. We suggest how different percepts of global motion are generated for stimulus configurations in changing contexts, like in the chopstick (Anstis, 1990) and the barberpole illusion (Wallach, 1935). In model simulations we show how feedback from MT to V1 cell populations and mutual form-motion interactions between V1/V2 and MT influence the generation of global percepts. In contrast to our model where the computation is relying on a feedforward-feedback processing loop, also pure
feedforward models for motion processing along the V1-MT hierarchy have been proposed (Rust et al., 2006; Simoncelli and Heeger, 1998; Nowlan and Sejnowski, 1994; Qian et al., 1994). These models focus on the neural properties of initial motion detection and partial integration, but do not consider temporal aspects of motion disambiguation.

The question of how motion estimates for stimuli formed by several components are computed is also connected to the generation of ocular following responses (OFR). Evidence exists that they are directly influenced by cortical motion processing in areas MT/MST (Kawano et al., 1994, 1997). Our model proposed here represents a possible basis for motion segmentation to explain the time course of human and primate OFR for moving stimuli perceived as pattern motion like in the barberpole experiment (Masson and Stone, 2002; Masson and Castet, 2002; Born et al., 2002).

2. Model

Our model consists of the first stages along the dorsal and the ventral pathway, namely areas V1, V2, and MT (Fig. 1). Subpopulations of neurons within the model areas compute either form or motion information of the input sequence. In this section, we will explain the different model stages and the interactions that exist between the subpopulations.

The dynamics of the individual stages is formally defined by first-order ordinary differential equations. We utilized firing-rate models to denote the dynamics of the average gradual activation of a group of neurons in a sheet of visual cortex at individual processing stages. Each model area is specified by a cascade of such stages, including a filtering stage for driving input, a feedback signal path to modulate the input, and a stage of center-surround shunting competition. In particular, we have

\[ \frac{\partial}{\partial t} V^{(1)} = -V^{(1)} + S^{\text{FF}} + 4\xi_{d1} + \Psi_{d2} \quad (1) \]

\[ \frac{\partial}{\partial t} V^{(2)} = -V^{(2)} + \frac{V^{(1)}}{1 + C \cdot 2^B} \quad (2) \]

\[ \frac{\partial}{\partial t} V^{(3)} = -D \cdot V^{(3)} + \left( \Psi_{d2} \right)^2 - (E + \Psi_{d3}) \cdot \sum_{\text{velocities}} \left( \Psi_{d2} \right)^2. \quad (3) \]

The terms \( V^{(1)} \), \( V^{(2)} \), and \( V^{(3)} \) denote the activity within the three stages of the particular model area, the term \( S^{\text{FF}} \) in Eq. (1) denotes the driving input signal, while \( z^{\text{FF}} \) in Eq. (2) is the modulatory feedback signal; the functions \( \xi \) and \( \Psi \) in Eq. (1) denote weighting kernels in the spatial and the velocity domain, respectively, \( \ast \) denotes the convolution operator. The constants \( C \) and \( E \) in Eqs. (2) and (3) adjust the strength of feedback and lateral inhibition, constant \( D \) adjusts the decay term. In the results presented here, the steady-state-solutions of Eqs. (1)–(3) are used to compute the neural activity. For this reason, the delay of neural responses between the different model areas is not taken into account. A detailed description of the model equations for motion and form processing is included in the Appendix A. In addition, a detailed description of the form processing can be found in (Thielscher and Neumann, 2008).

2.1. Motion processing

The model areas for motion detection that we propose are derived from a previously developed model of motion processing using feedforward and feedback interaction between areas V1 and MT (Bayerl and Neumann, 2004). This model suggests a possible explanation how the motion aperture problem is solved (Born and Pack, 2001), e.g., for a bar aligned diagonally moving in horizontal direction. In their approach, initial motion estimates were computed using an extended version of the Reichardt detector (Van Santen and Sperling, 1985). In the architecture proposed here, two subpopulations of neurons in V1 compute the initial motion information (Beck and Neumann, 2009). In a nutshell, normal flow responses are provided by model V1 complex cells that show a coarse speed and direction tuning. In addition, inhibitory interactions between cells at different V1 laminae with different scales generate local endstop responses. These responses are integrated for a short temporal episode, direction selective filters are then used to generate direction selective responses. The three-level processing cascade as explained in the previous subsection is applied in both subpopulations of V1 independently. Only at the level of MT their activity is combined. The V1 input activities are added up at each position using a slight multiplicative enhancement for the 2D motion cells. In addition to MT Motion cells that integrate V1 input, MT Motion contrast cells are included using a center-surround mechanism with opposite direction tuning. These cells respond stronger if the velocities in the surround differ from the movement detected in the center. A third neural subpopulation in model area MT are the MT Motion + Stereo cells tuned to both motion and disparity (Born and Bradley, 2005). They receive driving input from MT Motion cells and modulatory input from V2 Stereo cells. The main integrative fashion of the forward processing in the motion path is indicated by increasingly larger receptive field sizes, with a ratio of approximately 1:5 for V1:MT Motion and MT Motion + Stereo and 1:1.25 for MT Motion: MT Motion contrast cells.

2.2. Form processing

The stages along the form pathway, namely model areas V1 and V2, are modelled on the basis of a former model of long-range grouping and texture boundary formation (Thielscher and Neumann, 2003, 2008; Weidenbacher and Neumann, 2009). Both form and disparity tuned cells are simulated. V1 Form cells detect local luminance gradients of eight different orientations, V1 Stereo cells are activated by shifts in horizontal direction between the left and the right input image based on the responses of a correlation-based disparity detector (similar to the Reichardt detector) that is tuned to different horizontal disparities. V2 Form and Stereo cells then integrate the corresponding V1 inputs weighted with spatial bipole filters. Multiplicative combinations of V1 and V2 Form cells allow the detection of end-stop positions and possible X- or T-junctions (see Fig. 1 right, ratio of RF size V1:Endstop and V1:Bipole approximately 1:4).
2.3. Form-motion interactions

In this paper, we focus on the question how the form pathway influences the motion detection processing, in particular for complex stimuli with occluders. We suggest that the mutual cross-pathway interactions between selective representations of motion and form/disparity, namely bidirectional V1/V2-MT and MT-V1/V2 processing, are based on modulatory interactions. The bidirectional mutual interactions between V1/V2 Stereo and MT Motion + Stereo, the feedback from MT Motion contrast cells such as denoted in Eq. (2), and the input from endstop activity to MT Motion all follow the linking principle proposed by Eckhorn et al. (1990). Further interactions between form and motion pathway are the input at possible X- and T-junctions. A tonic input that is not velocity selective is added to V1 Motion cells after the recurrent step. This means that the overall activity at each position is increased strongly. In the subsequent normalization process this will shut down the activities at these positions considerably compared to other positions where only few neurons are active. Thus, the net effect of this tonic input is a suppression at the corresponding positions (see Bayerl and Neumann, 2007). The influence of this interaction is still modulatory as the activity is only reduced, but not completely inhibited. V2 Form information also influences the integration process in MT Motion. This means, that the feedforward activity of V1 is not simply added up in an isotropic fashion, but is weighted according to the factor \((1 + \upsilon_{V2, Form})\), where the activity of V2 Form \(\upsilon_{V2, Form}\) is proportional to the strength of activation in the form sensitive cells of area V2.

3. Results

The complementary interactions of the motion and form processing pathways in the model were tested with two stimuli consisting of moving bars, namely the chopstick (Anstis, 1990) and the barberpole illusion (Wallach, 1935). These stimuli are well-known psychophysical experiments as they lead to different percepts of motion when occluders cover the bar endings. This change of perception is a clear indication of the exchange of information between form and motion coding neurons. In addition, we tested the chopstick scenario using different disparities for the bars to show how stereo and form information influence the generation of disparity-tuned motion detection in MT.

3.1. Experiment 1: chopstick illusion

In the chopstick illusion two bars arranged like an “X” move horizontally in opposite directions. When two horizontal occluders are added at the top and the bottom of the bar endings the stimulus leads to a different percept switching from two components moving independently to one coherent pattern motion along upward or downward direction (depending on the movement of the two bars). This effect can be explained by our model based on mutual interactions between the motion and the form pathway. In Fig. 2, we show the result for the chopstick illusion without occluders, where two bars are moving in the same depth plane in opponent directions.

Initially, the motion estimates in area MT represent the true bar movement only at the bar endings. The 2D intrinsic terminators there detect the correct motion, mainly due to the input of V1 Motion endstop neurons. In contrast, along the edges at the beginning normal flow is indicated (input of V1 Motion complex), in the center the locally measured pattern motion is represented, which corresponds to an upward motion of the crossing itself if the movement of the two bars is divergent. The mutual interactions between cells in different model areas for motion detection now help to disambiguate these initially incoherent representations. During several iterations, the strong activity at the bar endings can propagate along the individual bars (Fig. 2 center). This effect is facilitated by two mechanisms: First, MT Motion contrast cells respond strongly at the bar endings. In the center of the RF high input activity is present due to the unambiguous activity at the bar ending. In the large surround, the opponent motion of the other bar further strengthens the activity. The strong activity of the contrast cells is fed back to MT integration cells and enhances the component motion of each bar. Second, the activity in the central part indicating motion in upward direction is weakened due to input from V2 grouping cells indicating that at this position an X-junction appears.

We now add static occluders in the input images to investigate the influence of the form channel when occlusions appear in the image which change the appearance of the moving components. The human perception for this scenario is that the whole pattern formed by the two bars is moving vertically in a coherent way. This phenomenon is replicated by our model results: As represented in Fig. 3 the overall motion after seven iterations (right image) shows a uniform upward movement. What are the mechanisms that contribute to this effect leading to a completely different percept in comparison to the scenario without occluders? After the first iteration (Fig. 3, left) the motion at the bar endings is already less pronounced than in the first scenario. The T-junction formed by the bar endings and the occluders leads to strong activity in the form channel which has an additive input in MT Motion for all velocities at this position. In other words, the overall activity at this position is enhanced which means that in the following divisive inhibition all these neurons will be weakened considerably. After the normalization in MT Motion, this result in reduced activity for the motion at the bar endings. At the same time, MT Motion contrast cells receive less input from these bar endings and, hereby, the feedback supporting their motion is only small. These changes allow the pattern motion to propagate along the edges and hence to dominate the whole percept.

The suggested model utilizes independent pathways and complex interactions, both within and across the pathways sensitive to motion and form information. In order to investigate the individual contributions of the components and their interactions in more detail we selectively added lesions in the network model. Such lesions involve: (a) the silencing of particular areas or subsystems (by eliminating their computations and the resulting representations) and (b) the selective cutting of connections or reducing their impact in the simulations. To show the contributions of the modulatory feedback connections, we will here show the results after systematic impairment of the computation by cutting or reducing connections in the model simulations. The results can be compared with those achieved by the intact model. In Fig. 4a, the results for the chopstick scenario without occluders are shown with feedback from MT Motion contrast cells having been eliminated. These cells respond strongly if opposite motion directions are found in area MT in their RF center and surround, respectively. For the chopstick experiment with occluders the effect is negligible (not shown), again global movement for the whole stimulus is indicated. However, without occluders the results look rather different: When the additional strengthening by modulatory input for the two bars moving in different directions is missing, the motion segregation is no longer successful. The pattern motion can partly propagate, the true motion direction is only partly achieved indicating a functional role of the contrast cells in the model. Furthermore, we tested the influence of V2 Form cells (results not shown) guiding the motion integration in MT for the scenario without occluders. When the connection is cut, instead of the integration weighted with V2 Form activity the input is only integrated in a spatially isotropic way, weighted with a Gaussian
As a consequence, the propagation of the motion along the bar was changed leading to a situation where neither the responses to the center nor to the line endings were able to spread along the contour to disambiguate the motion appearance. We also tested the influence of the additive input from V2 Form cells for the scenario with occluders (Fig. 4b). The additive input was weakened by a factor of 0.2. As a consequence, the motion at line endings was less reduced thus that the center motion was not strong enough to propagate. The results look similar to those without occluders in the intact model as depicted in Fig. 2.

3.2. Experiment 2: barberpole illusion

Similar to the chopstick illusion, also for a display where a diagonal grating moves in normal flow direction the motion percept changes when occluders are added. When the stimulus is shown behind an invisible square aperture, the normal flow direction indicating diagonal movement is perceived. Wallach (1935) has observed that changes of the shape of the (invisible) aperture lead to changes in the perceived motion of the grating pattern with a bias towards the direction of the elongation of the aperture. When either the horizontal or vertical line endings are hidden by (visible) extrinsic occluders, the percept is shifted towards the motion direction that is indicated by the visible not occluded line endings. Our results (Fig. 5) show that the model response is coherent with the perceptual phenomenon measured in humans and macaques (Wallach, 1935; Lidén and Mingolla, 1998; Pack et al., 2004): For a square aperture, the main activity in MT corresponds to the normal flow direction of the bar grating. In contrast, for a square aperture and occluders attached as flankers to the grating along the
vertical edges of the aperture, the response is dominated by the responses to the line endings along the horizontal direction. Also in this scenario, the interplay of local and global mechanisms of both motion and form is crucial for the two different percepts. Unlike the chopstick illusion, in the scenario without occluders no segmentation of single objects is necessary, but the integration of the overall response to the grating determines the human percept.

3.3. Experiment 3: chopstick scenario in two depth layers

Finally, in this experiment we demonstrate the effects of complementary interactions between disparity and motion coding. The MT Motion + Stereo cells receive input from both V1 Motion and modulatory enhancement from V2 Stereo cells. In Fig. 6, the activity of MT Motion + Stereo cells is shown for two bars aligned in different depth layers. A segregation of the bars is achieved after some iterations, in this example each bar clearly generates the main activity in its corresponding depth layer. Of particular interest is the region where the two bars cross. As the bars are presented in two different disparities, the bar in front should determine the motion at the crossing and it should be represented as continuous object in MT without an interception at the crossing. The simulations showed that after only few iterations the representation of the near bar can be completed in MT Motion + Stereo cells due to interactions with model area V2, whereas the second bar in the background is not completed as depicted in Fig. 6. After the first iteration, mainly the line endings of the bars are indicated in the two depth layers with very weak activity for the pattern motion at the crossing. After seven iterations, the bar in front is completed, responses to pattern motion are completely inhibited. This is also visible in the distribution of mean velocities in MT Motion + Stereo (depicted in Fig. 6). While at the beginning neurons tuned to various directions respond, after seven iterations a clear peak exists for the neurons indicating leftward and rightward motion like for the chopstick illusion without disparity information.

4. Discussion

4.1. New contributions

We have developed a model for the early and intermediate stages of motion and form processing in visual cortex. The model aims at revealing the function of cortical areas with neurons of different computational competences and their interactions to generate coherent task related representations of the input stimuli. The

Fig. 5. Barberpole illusion. Left: The distribution of the motion directions encoded in MT (weighted with its activity) for a square aperture is shown in a polar plot and the corresponding histogram. The highest activation is indicated in normal flow direction. Right: When occluders are added by juxtaposition of left and right flankings on line endings moving vertically, the peak shifts towards neurons tuned to horizontal motion. This effect is caused by the stronger responses at horizontal line endings that are due to the suppressive effect of the occluders not competing with neurons tuned to vertical motion.

Fig. 6. Chopstick scenario in two depth layers. Top: Activity of MT Motion + Stereo neurons tuned to upward, left and right motion for the far (top row) and the near (bottom row) plane. After the first iteration, both the near and the far bar are not continuously represented, some neurons tuned to the pattern motion respond in the center. After some iterations, the representation of the near bar moving rightwards is correctly completed. Neurons tuned to the near plane and movement to the right reach high activity all along the outline of the near bar. Also the representation of the bar in the background is completed, but correctly the central part is left as it is covered by the near bar. Bottom left: Mean velocity of MT Motion + Stereo neurons. In comparison to the chopstick illusion without different disparities (Fig. 2), the central part is here clearly dominated by the leftward motion provided by the near bar. Bottom right: The direction distribution in MT is shown after one (light bars) and seven iterations (black bars). At the beginning, neurons tuned to all directions respond, after seven iterations a clear segmentation of the two independently moving bars is achieved. As the bar moving rightward is in front, more neurons tuned to this direction are activated compared to the leftward movement.
model is used to replicate perceptual results of psychophysical and neurophysiological experiments.

The new contributions of this work are in particular: (i) the development of a model for interactions between the form and motion channel including form, motion, and stereo features that attributes different roles for different kinds of MT cells, (ii) an explanation of different perceptual phenomena in one single model that has generic mechanisms in all its areas leading to the perceptual competences from distributed neural computations, (iii) the specification of a new way of integrating motion estimates in MT guided by form information, and (iv) a suggestion of how the interaction between the different areas allows complementary coding keeping a partially distributed representation of the features.

4.2. Related work

Other work addressing the issue of coding complementary aspects of motion and form is the model of Berzhansky et al. (2007), building on previous developments in Grossberg (2000). Their approach utilizes a detailed description of signal pathways in the dorsal and the ventral stream of visual cortex as well as interactions between. The development is mainly driven by the knowledge of the (laminar) architecture and cell types that occur at different levels in the visual hierarchy and in the different processing pathways. In their model the final stage resembles large-scale MST cells which are sensitive to opposite motion direction. MST cells send feedback signals to earlier model stages of MT (oriented motion grouping mechanisms) and, in turn, influence complex cell responses in model V1. For the chopstick illusion, the model describes the various mechanisms in the processing areas up to the different layers within, amongst others, V1 and MT, and uses very large-spanning RF sizes in the higher levels like MT and MST. These MST cells act like a decision mechanism to bias motion integration along a particular direction in an all-or-none fashion. Our model, on the other hand, investigates the particular mechanisms needed to explain the main observations in perceptual decisions due to context-sensitive interpretation of the input stimuli. In particular, we focus here on the response amplification at localized features (line ends) and reduced, yet broadly tuned responses for locally ambiguous motion along extended boundaries (bars). The modulation of such localized feature responses in the case of occlusion patterns (T-junctions in our case) and the response to X-junctions at crossing form components is further analyzed and their concerted interaction is investigated computationally. Unlike the model of Berzhansky et al. (2007), we suggest that V1/V2 boundary cell responses signalling mutual occlusions send excitatory signals to MT that are unspecific with respect to velocity. Through this enhancement and subsequent competition the net effect leads to a reduction in response amplitude and, at the same time, an increase of uncertainty at occlusions (as measured in Pack et al. (2004)). Our model is based on a minimal level of description concerning the details necessary to replicate relevant data, namely using the three-level processing cascade as presented in the Model section. We have already shown that a number of computations for different feature dimensions, like form (Thielscher and Neumann, 2008), texture (Thielscher and Neumann, 2003), and opaque and transparent motion detection (Bayerl and Neumann, 2004; Raudies and Neumann, 2010) can be explained using this generic architecture to describe the neural computation on an abstract level. In the context of the application presented here, we want to stress the combination of a coarser, but present here, we want to stress the combination of a coarser, but

Another model for form-motion interaction was proposed by Lidén and Pack (1999) where T-junctions explicitly inhibit motion signals. In our contribution, an excitatory (additive) signal of V2 form enhances all motion sensitive cells at locations of potential occlusions. This tonic input in velocity space leads to decreases of salient responses at this position: First, the activity of all motion sensitive cells is increased by the input such that individual velocity sensitive responses are driven towards their upper saturation limit. The following normalization stage, in turn, shrinks down the entire activity pattern since the activation distribution lost its selectivity by injecting more energy. This mechanism provides the possibility for a highly activated neuron to keep its saliency even within the normalization process. It is hence consistent with the idea of incorporating modulatory interactions between the two pathways only.

Tlapale and coworkers (2008) proposed an extension of the motion model by Bayerl and Neumann (2004) to simulate the chopstick illusion as well as the barberpole illusion. One of the major differences to the original model is the integration in MT motion cells. The integration does not simply use spatial integration weighted with a gaussian function, but it is also depending on the local luminance similarity. The more similar the pixel within the integration field is compared to the central one, the stronger it is weighted. For example for the integration at a corner position of a unicolored box moving in front of a background, this would mean that the input from the box would be weighted stronger than the surround due to the identical color within the object, but not the background. Form and disparity information are not used. In our model, we picked up the idea of not purely isotropic integration in MT Motion. Unlike the approach of Tlapale and colleagues, we use the response of bipole filters in V2 Form to steer the integration process in MT Motion. This has the advantage that the integration is not based on simple, and possibly erroneous luminance differences, but on more reliable cues of the form pathway. This kind of integration is an advantage, for instance, in the case of the chopstick experiment with occluders: Due to the different orientations of the occluders and the bar endings, the integration will basically stop at the junction of the different objects. Hence, the motion integration for the different objects is kept more separately.

4.3. Structure and function of form-motion interaction

The model mechanisms presented in this work incorporate reentrant processing in the motion and form pathway and various cross-pathway interactions. These contribute to build up a distributed representation of segregated or coherently moving forms. To show the functional contribution of the different areas and connections, we chose the chopstick and the barberpole experiment. Both stimuli represent a challenging task for correct segregation as the movement of the whole stimulus is interpreted in different ways depending on the context (occluders/ no occluders).

For the chopstick illusion, without occluders the motion of the line endings indicating the correct component motion has to be propagated and it has to suppress the central pattern motion. We suggest that the segregation of localized objects is supported by MT Motion contrast cells that respond particularly strong for movements surrounded by opponent movement. Feedback from this area enhances MT Motion and Motion + Stereo cells which further feed back to V1 Motion cells. This mechanism supports the segmentation of different objects in the entire motion processing loop. The importance of the contribution of the contrast cells gets
visible if we cut the feedback connection to V1/MT Motion cells (Fig. 4). Without their feedback, the two bars moving separately can no longer be segmented correctly. Support for their opponent movement is necessary to propagate the true movement from the endings all along the bars and also into the central part that has initially only motion feature tracking signals in the vertical direction. In experiment 3 the same scenario was used, but the bars were represented in different depth planes. The results show that MT Motion + Stereo cells correctly achieve a segregation of the two bars based on the V2 Stereo information. Also here, the correct propagation along the bars depends on the MT Motion contrast cells. The additional depth information allows the correct assignment of the central region: After some iterations it only represents the movement of the bar in front. Without depth information in this region the motion of both bars is indicated as the two bars cannot be separated.

We recently proposed an algorithmic version of our neural model for motion processing also including motion contrast cells (Beck et al., 2008). Unlike to the model presented here, these investigations were aiming to arrive at an explicit object segmentation. In the experiments performed with the algorithmic version we showed another functionality of motion contrast cells. Since these cells indicate motion discontinuities and thus the border of a moving object they cannot only contribute to object segregation, as shown here, but are also useful for scene segmentation.

Concerning the robustness of the parameterization of the model, we tested stimuli of different sizes all leading to correct results. Nevertheless, the size of the receptive fields of the contrast cells is critical for the model in a way that for very large stimuli the motion contrast will not be detected at the bar endings. A possible solution for this issue is the integration of processing on multiple spatial scales. In addition, one has to keep in mind that also for a human observer, the percept generated by the chopsticks will disappear when the stimulus gets very big.

In the scenario with occluders, the disambiguation of the motion is driven by the central junction of the two bars with pattern motion, along the occluders the activities are suppressed. Here, the output of the contrast cells is less important and they respond only weakly. The pattern motion does not depend on the segmentation of the scene, but on integration mechanisms and thus also works without feedback from the contrast cells. However, the suppression of extrinsic motion cues along the occluders is crucial to avoid that the strong 2D features generated at the junction of the occluders and the bars are propagated along the bars. This is enabled via the modulatory connections from V2 Form. Also, the barberpole illusion depends on this effect: If occluders are covering the horizontal or vertical line endings, the intrinsic features at these positions become extrinsic features and are as such partially suppressed by the V2 Form input. As a consequence, the line endings that are not occluded will determine the perceived movement. Here, also the influence of MT integrating cells is important. When the aperture is quadratic, the line endings contribute likewise to a vertical and horizontal percept, the sum of the overall activity results in the normal flow direction. A difficulty that has to be taken into account when simulating the chopstick and the barberpole illusion in one common model is the different percept in the scenario without occluders. For the chopstick scenario, this leads to a segregation of two objects where the motion at the 2D positions dominates the result. In contrast, for the barberpole stimulus the normal flow dominates the perceived motion. This shows again that depending on the stimulus rather a strategy of integration or segregation into different objects is used. Our model is able to show the various results due to the interplay of the different pathways and the modulatory feedback connections by contrast ("segregation") and integrating cells.

In the context of the results of the psychophysical and neurophysiological experiments not only the final percept is of interest, but also the temporal change of neural activity leading to different percepts at different times. This temporal change is connected to the question of how the pattern selectivity in MT is computed (Pack and Born, 2008). Some evidence exists, that both 1D and 2D motion estimates of V1 contribute to this computation. Our model includes a suggestion how this could be achieved. The change of activity in MT during time was measured for stimuli where the aperture problem has to be solved. For both chopstick and barberpole illusion after the first iteration along the edges of the bars only normal flow is estimated, the true motion is then propagated along the stimulus. The time necessary to complete the propagation of the correct movement increases with bar length, as shown by neurophysiological data of Born and Pack (2001), that can be replicated with model areas V1 and MT Motion (compare Bayerl and Neumann, 2004). In the case of the chopstick illusion the process of propagation is more difficult: Depending on the configuration, either the activity from line endings or the central pattern motion should propagate (compare Figs. 2 and 3). This means that context information has to be included in the motion processing path in a way that the propagation is steered by the global configuration. In our model, the modulatory input of V2 Form cells and feedback of MT contrast cells to MT Motion represent a "soft switch" that biases the motion areas and allows the correct propagation step by step.

The iterative generation of the correct percept is also related to experiments concerning OFRs. The temporal shift of activity from normal flow to pattern flow (as in experiments 1 and 2) is in line with behaviour found for OFR. Kawano et al. (1994, 1997) measured that the activity in monkey MT/MST is preceding OFR by approximately 10 ms. For experiments showing stimuli first dominated by component and then by pattern motion as in the barberpole illusion or in moving plaids, experiments both with humans (Masson and Stone, 2002; Masson and Castet, 2002) and monkeys (Born et al., 2002) revealed that the initial direction of pursuit eye movements follows the component motion direction. Over time, the direction changes towards the direction of pattern motion. The interaction of form and motion pathway in our model represents a plausible explanation for the generation of different eye movement directions measured in these experiments. The interaction of the model areas provides robust segmentations and disambiguated representations of the scene at a low level of the cortical hierarchy and could generate the necessary representations underlying subsequent processes of sensory-initiated perceptual decision-making.

5. Conclusion

In conclusion, we propose a neural model of early visual processing areas for motion detection, segmentation, and integration that suggests interactions between the motion (V1, MT) and the form (V1, V2) pathway to (i) guide the integration of motion responses in MT, (ii) avoid the erroneous development of salient activity of motion cells at positions with occlusions and (iii) achieve response properties in MT cells tuned to both motion and disparity. We also explain the role of different MT cells, in particular for MT Motion contrast cells that support the segmentation of objects moving in different directions. The different model areas allow a distributed representation of the information which interacts via modulatory connections. We claim that without these connections, the individual areas fail to compute the correct estimate as demonstrated by lesion experiments. The presented model successfully reproduces experimental data of the chopstick and the barberpole illusion which have been used to investigate the effects of figural context in the computation of coherent scene motion.
Acknowledgements

This research has been supported in part by a Grant from the EU to STREP-project no. 027198-Decisions-in-Motion and to ICT-project no. 215886-SEARISE as well as the Transregional Collaborative Research Center SFB/TRR62 “Companion Technology for Cognitive Technical Systems” funded by the German Research Foundation (DFG). The authors would like to thank the four reviewers for their insightful comments that helped to substantially improve the manuscript.

Appendix A

The model was tested with input images of a maximal size of 120 × 90 pixels for the chopstick illusion. In the following, we will first depict the model equations to compute the motion detection as applied for the chopstick (Experiment 1) and the barberpole illusion (Experiment 2). Second, a short description of the model equations for form processing will be given. A more detailed description of the equations to compute V1 and V2 Form activity and responses of endstop neurons as well as T- and X-junctions can be found in Thielscher and Neumann (2008) and Weidenbacher and Neumann (2009).

A.1. Model equations for motion and form processing

In the model simulations, we used the steady-state responses of the generic processing mechanisms as described in the model section. In the following equations, the terms $v^{1D}$, $v^{2D}$, and $v^{3D}$ denote the activity within the three stages of the model area. Weighting kernels in the spatial and the velocity domain are denoted as $A$ and $\mathcal{V}$, respectively, $\ast$ denotes the convolution operator, $x$ the spatial position, $\phi$ denotes the direction of motion, $s$ the corresponding speed. The constants $C_{\text{ModelArea}}, G, F$ and $H$ adjust the strengths of feedback, multiplicative enhancement and lateral inhibition.

A.1.1. V1 Motion

The initial motion estimation in V1 can be distinguished in neurons with activity $v^{1D}$ and $v^{2D}$ that provide 1D (V1 Motion complex) and 2D motion detection (V1 Motion endstop).

V1 Motion complex

Initial motion estimation $I_{1D}$ is computed by a Marr-Ullman like motion detector (Marr and Ullman, 1981) using speeds ranging from 1 to 6 pixels in eight different directions. The detector is based on two succeeding input images $I_{m}$ and $I_{m-1}$ convolved in the spatial domain with a Laplacian of Gaussian filter ($\text{LoG}_m$ and $\text{LoG}_{m-1}$, respectively (cmp. Eq. (A1)) followed by a temporal derivation by computing their difference $\text{DiffLoG}_m$ (Eq. (A2)). These activities are separated in positive and negative responses ($\text{DiffLoG}^{\text{pos}}_m$, $\text{DiffLoG}^{\text{neg}}_m$, etc.). To enhance the responses for edges, the responses are multiplied with the spatially adjacent positions orthogonal to the movement direction computed. The direction of motion is detected by selectively multiplying the activities of the spatial and the temporal derivatives, depending on the contrast polarity. Here, the corresponding spatial derivatives have to be shifted spatially (by speed/2 pixel along and against the direction of motion indicated as function left() and right() in Eq. (A3)). As an example, rightward motion ($\phi = 0^\circ$) is both indicated by a) a high responses of $\text{LoG}^{\text{pos}}_m$ on the left side, of $\text{LoG}^{\text{neg}}_m$ on the right side and of $\text{DiffLoG}^{\text{pos}}_m$ in the center and b) a high responses of $\text{LoG}^{\text{neg}}_m$ on the left side, of $\text{LoG}^{\text{pos}}_m$ on the right side and of $\text{DiffLoG}^{\text{neg}}_m$ in the center (Eq. (A3)). Speed selectivity is achieved by using Laplacian of Gaussian filters of different size (cmp. Eq. (A1)). Zero motion is not represented explicitly. The motion detection was restricted to the regions where zero crossings had been detected after filtering with an off-center off-surround RF.

$$\text{LoG}_m = I_m \ast A_{\text{LoG}(\phi, \text{speed})}$$

(A1)

$$\text{DiffLoG}_m = \text{LoG}_m - \text{LoG}_{m-1}$$

(A2)

$$v^{\phi = 0^\circ}_{1D} = \max \left( \begin{array}{l} \text{left} (\text{LoG}^{\text{pos}}_m) \times \text{right} (\text{LoG}^{\text{neg}}_m) \times (\text{DiffLoG}^{\text{pos}}_m) \\ \text{right} (\text{LoG}^{\text{neg}}_m) \times \text{left} (\text{LoG}^{\text{pos}}_m) \times (\text{DiffLoG}^{\text{neg}}_m) \end{array} \right)$$

(A3)

where $\text{DiffLoG}^{\text{pos}}_m = [\text{DiffLoG}_m]^+$ and $\text{DiffLoG}^{\text{neg}}_m = [-\text{DiffLoG}_m]^+$. etc.

The responses of this motion detector and a weak background activity $\text{Noise}$ (normally distributed) represent the input for the V1 Motion complex cells (Eq. (A4)). In the second processing step, feedback $\text{rmt}$ from model area MT (only direction selective) and input $\phi_{\text{V3}}$ from endstop cells of the form channel enhance the neural activity. A small input $v_{\text{joint}}^{\text{V2}}$ from V2 indicating X- (small activity) or T-junctions (high activity) is added before the normalization step (Eq. (A5)). Due to the summation over all directions $\phi$ and speeds $s$ the net effect of this additive input is an inhibition of activity after the normalization (Eq. (A6)).

$$v^{\phi = 0^\circ}_{1D} = I_{1D} + \text{Noise}$$

(A4)

$$v^{\phi = 0^\circ}_{1D} = v^{\phi = 0^\circ}_{1D} \times \left( 1 + C_{\text{V1,1D}} \times \sum_s v^{(3)}_{\text{MT}} + v_{\text{V3}} \right) = 0.05 \cdot v_{\text{joint}}^{\text{V2}}$$

(A5)

$$v^{\phi = 0^\circ}_{1D} = \left(v^{\phi = 0^\circ}_{1D} \times 0.001 + \sum_{\phi, s} v^{(2)}_{\text{V1,1D}} \cdot 2 \right)$$

(A6)

V1 Motion endstop

Initial motion estimation $I_{2D}$ is like for $I_{1D}$ computed for a speed range from 1 to 6 pixels in eight directions. The responses of endstop cells of four succeeding image frames are used to generate a temporally blurred response. Motion direction is achieved using direction selective filters. To enhance the results at positions with high end-stops activity, the filter responses are multiplied with the endstop responses. In the following, the same processing steps are applied like in model area V1 Motion complex (Eqs. (A7)–(A9)). Very small neural activity is suppressed by a small subtractive inhibition followed by a rectification (indicated as $[\cdot]^+$ in Eq. (A9)).

$$v^{\phi = 0^\circ}_{2D} = I_{2D}$$

(A7)

$$v^{\phi = 0^\circ}_{2D} = \left[ 10 \cdot v^{\phi = 0^\circ}_{1D} \times \left( 1 + C_{\text{V1,2D}} \times \sum_s v^{(3)}_{\text{MT}} \right) \right] \times \left( v^{\phi = 0^\circ}_{2D} \times 0.001 \right) + C_{\text{V1,2D,joint}} \times \left( v_{\text{V2}}^{\phi = 0^\circ} \right)^2$$

(A8)

$$v^{\phi = 0^\circ}_{2D} = \left(v^{\phi = 0^\circ}_{2D} \times 0.001 + \sum_{\phi, s} v^{(2)}_{\text{V1,2D}} \right) \times 0.001$$

(A9)

A.1.2. MT Motion

In the first processing step of area MT the input of the two V1 subpopulations is integrated using a spatial gaussian filter that is modulated by the form information of V2 bipolar filters (Eq. (A10)). In addition to the feedback, a slight blur of directions is included in the second processing step (Eq. (A11)). Finally, the squared activity is normalized (Eq. (A12)).

$$v^{\phi = 0^\circ}_{\text{MT}} = G \cdot v^{\phi = 0^\circ}_{\text{V1,2D}} + v^{\phi = 0^\circ}_{\text{rmt}} \ast \left( 1 + v^{(3)}_{\text{V1,2D, bipolar}} \right) \ast A_{\text{V1,MT}}^{\phi = 0^\circ}$$

(A10)

$$v^{\phi = 0^\circ}_{\text{MT}} = \left( v^{\phi = 0^\circ}_{\text{MT}} \times \sum_s v^{(3)}_{\text{MT}} + \Psi (\phi, s, \text{MT} = 0, 2) \right)$$

(A11)

$$v^{\phi = 0^\circ}_{\text{MT}} = \left( v^{\phi = 0^\circ}_{\text{MT}} \times 0.1 + \sum_{\phi, s} v^{(2)}_{\text{MT}} \right)^2$$

(A12)
A.1.3. MT Motion contrast

In this model area, the neural response is increased if motion in the opponent direction is found in the surround. After spatial integration of the squared MT input (Eq. (A13)), the responses of the center and the surround of the receptive field are computed. The center activities are an integration of the neurons’ activities tuned to similar directions weighted with a gaussian filter applied in the spatial domain. The surround activity is based on the activities of neurons tuned to the opponent direction, again weighted with a spatial gaussian filter. The center response is modulatory enhanced to similar directions weighted with a gaussian filter applied in the center and the surround of the receptive field are computed. The input to MT Motion contrast cells was cut, parameter \( C_{\text{MT}} \) was set to zero.

\[
\begin{align*}
I_{\text{MT}}^{(1)} &= 10 \cdot \left( I_{\text{MT}}^{(2)} \right)^2 \cdot A_{\sigma_1, \text{MT}} \\
I_{\text{MT}}^{(2)} &= \left( \left( I_{\text{MT}, \text{on}} \right) \cdot \left( 1 + C_{\text{MT}} \cdot I_{\text{MT}, \text{off}} \right) \right) \\
I_{\text{MT}}^{(3)} &= \left( I_{\text{MT}, \text{on}} \right) \cdot \left( 0.08 + H \cdot \left( \sum I_{\text{MT}}^{(2)} \right) \right) + I_{\text{MT}, \text{off}} \cdot A_{\sigma_2, \text{MT}} \\
\end{align*}
\]

In Eq. (A14) the activations \( I_{\text{MT}, \text{on}} \) and \( I_{\text{MT}, \text{off}} \) are determined by the following convolutions:

\[
I_{\text{MT}}^{(1)} = I_{\text{V1}, \text{Form}}^{(1)} \\
I_{\text{MT}}^{(2)} = I_{\text{V1}, \text{Form}}^{(2)} \\
I_{\text{MT}}^{(3)} = \left( I_{\text{V1}, \text{Form}}^{(3)} \right)^2 \cdot \left( 0.01 + \sum I_{\text{V1}, \text{Form}}^{(2)} \right).
\]

A.1.5. V2 Form

In V2 Form the activity of V1 Form \( I_{\text{V1}, \text{Form}}^{(3)} \) is integrated by model bipole cells (Eq. (A19)). These bipole cells consist of two prolated subfields corresponding to two elongated gaussian filters \((\sigma_1 = 6 \text{ and } \sigma_2 = 1.25, 8 \text{ orientations})\) which are shifted parallel to the main axis of the cell. The two subfields are multiplicatively combined. At each position, the bipole filter in the orientation corresponding to the orientation tuning of the V1 input cells is applied. No feedback is applied at this stage in the current version (Eq. (A20)) without stereo cells. In Eq. (A21) the normalization of the activity is described. Form activity \( I_{\text{V2}, \text{Junct}} \) indicating junctions is computed as shown in Eq. (A22). This activity is used as additive input in Eqs. (A4) and (A8).

\[
\begin{align*}
I_{\text{V2}, \text{Form}}^{(1)} &= I_{\text{V1}, \text{Form}}^{(3)} \cdot A_{\text{bipole}} \\
I_{\text{V2}, \text{Form}}^{(2)} &= I_{\text{V2}, \text{Form}}^{(1)} \\
I_{\text{V2}, \text{Form}}^{(3)} &= I_{\text{V2}, \text{Form}}^{(2)} \cdot \left( 4 + \sum I_{\text{V2}, \text{Form}}^{(2)} \right) \\
I_{\text{V2}, \text{Junct}} &= I_{\text{V2}, \text{X}} + I_{\text{V2}, \text{Junct}} \\
\end{align*}
\]

where \( I_{\text{V2}, \text{X}} = 0.01 \cdot \left( I_{\text{V2}, \text{Form}}^{(2)} + I_{\text{V2}, \text{Form}}^{(3)} \right) \cdot A_{\sigma_1, \text{X, Junct}} + 0.04 \text{ and } I_{\text{V2}, \text{Junct}} = \left( I_{\text{V2}, \text{X}} + I_{\text{V2}, \text{Form}}^{(2)} \right) \cdot A_{\sigma_1, \text{Junct}}.

Parameters:

<table>
<thead>
<tr>
<th>( \sigma_1, MT )</th>
<th>( \sigma_2, MT_x )</th>
<th>( \sigma_2, MT_y )</th>
<th>( \sigma_1, \text{MT center} )</th>
<th>( \sigma_1, \text{MT surround} )</th>
<th>( \sigma_2, \text{MT}_{x,y} )</th>
<th>( \sigma_1, \text{T, Junct} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chopstick</td>
<td>5</td>
<td>0.5</td>
<td>0.75</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Barberpole</td>
<td>5</td>
<td>0.5</td>
<td>0.75</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \sigma_1, \text{X, Junct} )</th>
<th>( C_{1, \text{V1, 1D}} )</th>
<th>( C_{1, \text{V1, 2D}} )</th>
<th>( C_{\text{MT}} )</th>
<th>( C_{\text{MT}} )</th>
<th>( C_{\text{V1, 2D, Junct}} )</th>
<th>( G )</th>
<th>( F )</th>
<th>( H )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chopstick</td>
<td>7</td>
<td>1</td>
<td>0.1</td>
<td>100</td>
<td>100</td>
<td>0.01</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>Barberpole</td>
<td>7</td>
<td>1</td>
<td>0.1</td>
<td>100</td>
<td>100</td>
<td>0.01</td>
<td>50</td>
<td>10</td>
</tr>
</tbody>
</table>

References


