Modeling motion processing in macaque area MT/V5: From single cells to population codes

Temujin Gautama & Marc M. Van Hulle

Abstract.
Theoretical and psychophysical studies have shown that the relative motion between observer and environment induces a retinal flow from which object edges and 3D surface shapes (shape-from-motion) can be reconstructed. The primate middle temporal area MT or V5, for which there is general agreement that it is implicated in motion processing, is a likely candidate for generating such descriptions. The great majority of the macaque MT neurons possess an antagonistic (suppressive) region surrounding the classical receptive field. It is believed that this "surround" is involved in locating image discontinuities induced by the edges of moving objects. On the other hand, experimentalists and modelers have also conjectured that the surround is involved in the reconstruction of 3D surface shape. We have adopted a modeling approach to investigate whether these two, seemingly dissonant views can be reconciled. The parameters of the model neurons were optimized using a training algorithm so that they reproduce the responses of a selection of 31 cells taken from the MT database of Orban and co-workers. Each model neuron was thereafter tested for its selectivity to moving 2D edges and 3D surface shapes. The observed selectivities were entirely explainable by the surround's spatial and speed characteristics. We also examined the responses of neural populations and showed that a linear coding scheme suffices for generating accurate surface slant and tilt estimates in populations as small as six neurons. Finally, we developed a plausible scheme for extracting 3D descriptions of a visual scene from the retinal flow.
Keywords: middle temporal area, MT/V5, modeling, center-surround antagonism, motion segmentation, object segmentation, slant/tilt coding, shape-from-motion

1. Introduction

Visual perception does not occur in isolation but, rather, in a system that interacts with its environment. As we move about, we can actively control the image flow that reaches our retinas by making purposeful body, head and eye movements that will greatly facilitate the recovery of the 3D structure of the environment. The relative motion between an observer and a spherical object that is stationary or that translates in the observer's visual field (Fig. 1), induces a retinal velocity field from which the relative depth and the 3D shape of the sphere can be reconstructed. There is a growing body of evidence that the middle temporal (MT, V5) cortical area, located in the posterior bank of the superior temporal sulcus of the macaque and its human homologue, hMT/V5+, play a prominent role in this process. The receptive fields of MT neurons in the macaque are distinguished from those of other early visual areas by their selectivity for the direction and speed of retinal motion [12, 34, 21, 1, 23, 18]. In addition, the majority of MT cells have a surround which envelopes the classical receptive field [3, 26, 7, 25, 30, 31, 32, 33, 8]. Moving stimuli in the surround alone do not cause MT cells to

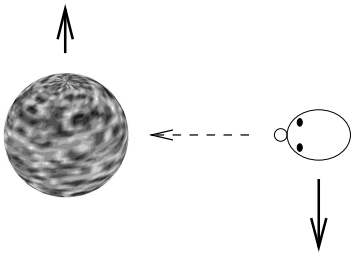


Figure 1: Schematic representation of the stimulus setup. A velocity field is induced by the relative motion between the observer (downwards) and the opaque sphere (upwards).

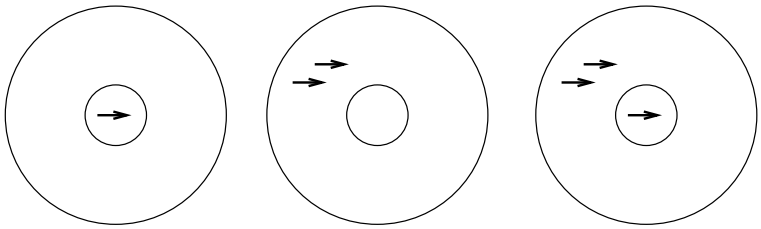


Figure 2: Schematic rendition of the antagonistic influence of the surround. The MT cell responds when its center is activated (left panel). Stimulation of the surround by itself does not evoke a response (middle panel), but modulates (mostly inhibits) the center response when both center and surround are stimulated simultaneously (right panel).

respond, but their presence modulates the response to stimuli in the classical receptive field (center). For the large majority of MT cells, the cell's response is suppressed when stimuli in the surround move in the same direction as those in the center, as is shown in Fig. 2. These findings suggest that these cells carry information about local motion differences and/or local gradients, and this, in turn, has fostered a number of speculations about the functional role of the antagonistic surround, of which the two most important ones are motion segmentation and surface shape description.

Motion Segmentation Theoretical studies of Gibson [14] and Barlow [5] have predicted that motion differences in the retinal flow are highly informative about the three-dimensional environment since large differences are observed at the edges of objects lying at different distances. Nakayama and Loomis [24] have further predicted the existence of motion-selective neurons in the visual system which possess antagonistic center/surround receptive field organizations for detecting motion discontinuities. In line with this, several physiological studies have interpreted the antagonistic MT surround in terms of a mechanism for locally performing motion segmentation and, more specifically, for locating image discontinuities [3, 26, 1, 2]. Furthermore, Bradley and Andersen [8] have recently found evidence for a center/surround disparity antagonism in macaque MT that operates on top of the velocity antagonism, and which has led them to conclude that the surround not only provides information about discontinuities based on motion but also based on binocular disparity. Furthermore, in addition to motion segmentation, several other roles for the MT surround have been suggested based on differential motion detection such as the computation of self-motion during eye movements [29], figure/ground segregation [3, 19], and the differentiation of object

motion from egomotion [4, 8]. What these views have in common, including the one concerned with motion segmentation, is that the MT surround is assumed to be radially symmetric and spatially homogeneous.

Surface Shape Description In a second line of theoretical studies, instead of looking for motion discontinuities and edges of moving objects, it has been shown that the reconstruction of the 3D object shape from retinal motion (shape from motion, SFM) is feasible provided that the first- and second-order directional derivatives of the 2D-velocity field are available [11, 17]. Furthermore, in these studies predictions have been formulated about the center/surround receptive field organizations needed to locally extract these directional derivatives, but without making clear statements whether the predictions applied to MT neurons. Furthermore, psychophysical evidence which indicates that human SFM perception involves surface interpolation [16, 27] has led to the hypothesis that the spatial structure of the receptive field is involved in this process. Hence, in this view, SFM is redefined as the problem of characterizing interpolating surfaces. These studies have inspired Buračas and Albright [9] to model the velocity selectivity curves of the MT neuron as sums and differences of two Gaussian-distributed velocity filters, one for the center and another for the surround (*i.e.*, a parametric model). They have used these models for predicting the role played by the MT surround in the extraction of the surface orientation-in-depth and the surface curvature. However, they did not model the spatial layout of the surround – they assumed it to be radially symmetrical – and, hence, only the magnitude of the surface orientation in depth, namely the slant, and the mean curvature could be extracted. On the other hand, they have also qualified their assumption by predicting the existence of asymmetrical surrounds (but, "possibly in areas FST or MST", p. 975). More recently, Orban and co-workers have found that the majority of the MT neurons in the macaque are *not* radially symmetrical [30, 31], contrary to what was originally assumed by the experimentalists [26], and they have conjectured that these surrounds enable MT neurons to locally compute the first- and second-order directional derivatives of the retinal flow. Furthermore, they have also quantified the selectivity of these neurons to the orientation-in-depth of a translating plane (thus, both the slant *and* tilt) and have reported that this selectivity depends on the presence of asymmetric surrounds and on the speed selectivity of the asymmetric surround interaction. In light of these findings, Buračas and Albright [10] have complemented their original model with asymmetrical surrounds, and have discussed the implication of their MT surround model on slant, tilt and curvature extraction. However, contrary to slant and tilt, no published experimental study to date has demonstrated selectivity for surface curvature in single MT neurons. Finally, Liu and Van Hulle [20] have adopted a non-parametric (*i.e.*, neural network) approach to modeling the role played by the symmetrical as well as the asymmetrical MT surround in the orientation in depth coding of translating planes.

In summary, there is supporting evidence for both of these seemingly dissonant views on the functional role of the antagonistic MT surround, namely the detection of an object's edge and the extraction of descriptors for its surface shape (SFM). However, they have not yet been combined in a single study. We will use a modeling approach to examine both functionalities, using a *single* model, the structure of which takes into account several experimental findings, such as the center/surround antagonism and the speed tuning. We re-examine the experimental data of Orban and co-workers [30, 31, 33], and develop a model neuron for every MT cell in their database. The parameters for every model neuron are adjusted such that the experimental data is replicated

(*training*). In the next step, the model neurons are used to predict MT's behavior to stimuli that have not been used in the experiments. In particular, we use stimuli to dissociate between motion segmentation and SFM in motion processing. We have found that our model neurons can perform both types of motion processing. The simulations are used to formulate new hypotheses concerning the function of the antagonistic MT surround and allow for the development of new experimental paradigms to validate these hypotheses. At the end of the chapter, we propose a possible mechanism for analyzing a visual scene based on retinal motion, using the developed MT model neurons.

2. Materials and Methods

This section first describes the experimental data and the stimulus paradigms. Next, the architecture and parameters of the model we have developed are explained, its stimuli and, finally, the training procedure.

2.1. Experimental Data Set

The electrophysiologically obtained data we have modeled, is a subset of the MT database of Orban and co-workers. The details concerning the recording procedure and visual stimulation are described in their publications [18, 22, 25, 30, 31]. We limit ourselves to a brief description.

Extracellular single unit recordings have been made in area MT of anesthetized and paralyzed adult macaque monkeys (*Macaca fascicularis*) prepared for acute recording. Cell activation has been recorded in a 200 ms window shifted 50 ms with respect to stimulus presentation, to correct for response latency. The data is represented as a normalized response, α, *i.e.*, the net response (after subtraction of the spontaneous firing rate) divided by the net response to the control condition of the surround asymmetry speed test (SA-speed test[1], see further).

The stimuli consist of random patterns of white dots (luminance 48 cd/m^2) with a diameter of 0.7° of visual angle, on a dark background (0.2 cd/m^2), covering an area with a diameter of 25.6° and moving in the cell's preferred direction. Three sets of stimulus conditions have been considered (Fig. 3): the first two measure the surround inhibition and the third the selectivity to the orientation-in-depth of translating planes.

SA-speed test The surround asymmetry speed test (SA-speed test) consists of three sets of eight stimuli, each of which is constructed by a central circular patch of dots moving in the cell's preferred direction at its optimal speed, v_{opt}, together with a second, peripheral patch located in one of eight positions in the antagonistic surround (see Fig. 3A). The central patch corresponds to the cell's classical receptive field. The dots of the peripheral patches move in the cell's preferred direction, at the same speed or at a speed four times slower or faster than the speed in the center. Hence, there are 24 stimuli containing a central and peripheral patch, and 1 stimulus containing the central patch only, *i.e.*, the control condition. The degree of inhibition is computed with respect to the response to the control condition. The SA-speed test results are represented in three polar plots (one for every speed in the peripheral patch). The angle corresponds to the position of the peripheral patch, and the length is the degree of inhibition for that condition. This is illustrated for cell m82 7 in Fig. 9A–C.

[1] For the sake of comparison, we will adopt the terminology used by Orban and co-workers for their stimuli.

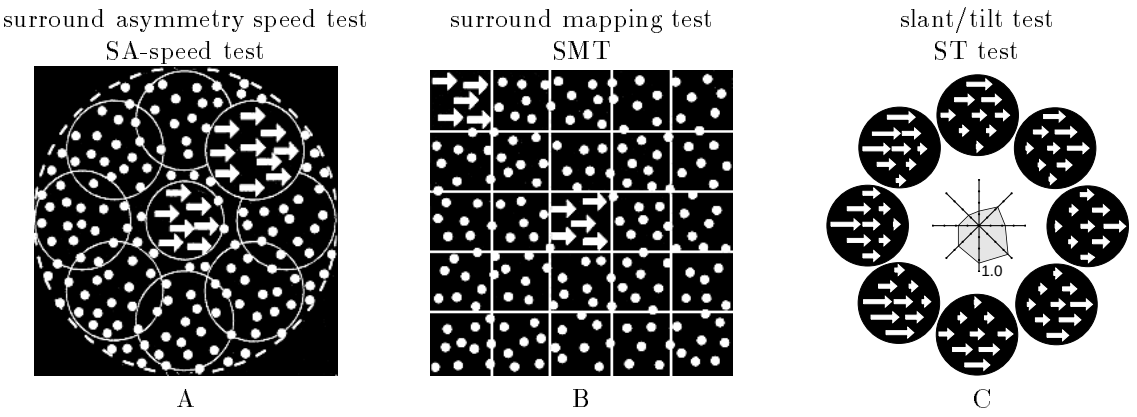


Figure 3: Experimental Stimuli. Diagram of the stimuli used in the surround asymmetry speed test (SA-speed test) (A), surround mapping test (SMT) (B) and slant/tilt test (ST test) (C). A) The SA-speed test stimulus consists of a central patch of moving dots (arrows), moving in the preferred direction (here, rightward) at the optimal speed, and a second peripheral patch of moving dots along one of eight possible angles. The dots in the peripheral patches move at a speed slower, equal or faster than the speed in the central patch, also in the preferred direction. The central and the peripheral patches are separated by a 1 cm gap. B) The SMT stimulus consists of dots moving at optimal speed and direction in a center square and in 1 of 24 peripheral squares on a 5×5 grid. Lines delineating stimulus borders are shown for clarity's sake and are not present in the actual stimuli. The SA-speed test and SMT stimuli are reprinted from Fig. 1A,C in [30]. C) Schematic rendition of the ST test. The eight stimuli are depicted, with longer arrows indicating higher speeds. At the center, the response of the cell is plotted: the magnitude corresponds to the response strength and the angle to the tilt angle of the stimulus (reprinted from [31] Fig. 2).

SMT The surround mapping test (SMT) probes the surround at v_{opt} in 24 of the 25 positions on a 5×5 grid, with the dots in the center square also moving at v_{opt} (Fig. 3B). In the control condition, only the central patch is stimulated. The degree of inhibition is computed with respect to the response to the control condition. Figures 10A and C show example SMTs after spatially interpolating the computed inhibition.

ST test The slant/tilt test (ST test) consists of a series of tests in which eight directions of a motion gradient are presented for each of three gradient magnitudes, thus, 24 stimuli in total (see Fig. 3C). The motion gradients correspond to oriented planes under a perspective projection from a point at a given viewing distance d to the video monitor. The orientation-in-depth is expressed in terms of slant and tilt. Slant σ is the angle between the inward surface normal and the line of sight (Z-axis), and tilt τ is the angle over which it is rotated around the line of sight, as illustrated in Fig. 4. Note that the plane moves in the cell's preferred direction. A cell's normalized response to tilt can again be represented in a polar plot (vector length = cell's response; angle = tilt). An example is shown in Fig. 9D, again for cell m82 7.

From the complete MT database of 145 cells [33], 31 cells are taken into consideration, namely those for which both the SA-speed and ST tests have been performed. Furthermore, only the motion antagonism in the cell's preferred direction will be considered, since the cell's responses in other directions has not been systematically recorded.

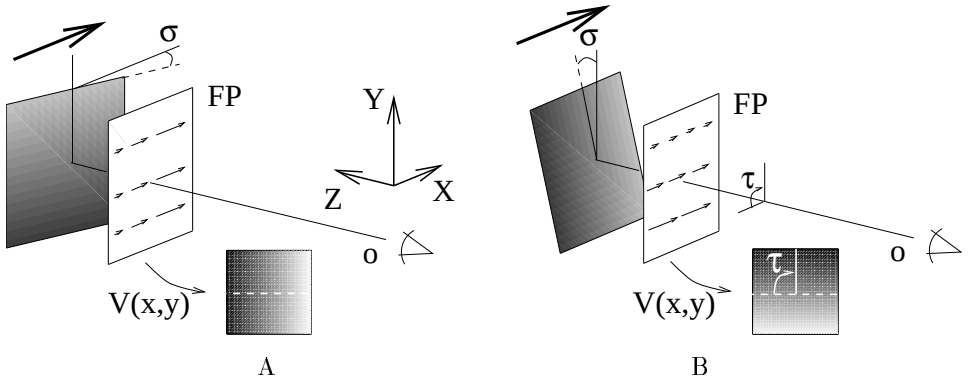


Figure 4: Stimulus setup for a translating plane under the perspective projection. The plane translates in the direction indicated by the straight thick arrow and induces a velocity field V(x,y) on the frontoparallel projection plane FP, which is seen by the observer o. Since the translation is always in the same direction, the velocity vector reduces to a scalar speed, the distribution of which is represented as a 2D grayscale image (see the icon pointed at by the curved arrow). The orientation in depth of the translating plane is denoted by slant σ, which is defined by the angle between the inward surface normal and the line of sight, and tilt τ, which corresponds to the angle over which the plane is rotated around the line of sight. Examples are shown for a positive slant and zero tilt (A), and for the same slant but -90° tilt (B).

2.2. MT Model

Our MT model adheres to the center-surround paradigm. It takes into account the possible heterogeneity and the speed dependency of the surround receptive field structure, that have been experimentally observed [30, 31, 32, 33]. For every cell in the MT database, a specific model neuron is developed by determining the model parameters by means of a training algorithm. The MT model neurons are validated by comparing their behavior to that of the corresponding MT cells in the database, for stimuli that have not been used for training.

2.2.1. Model Architecture and Training

The inputs to the model are velocity fields. However, since electrophysiological recordings have systematically been performed for translational motion in the cell's preferred direction only, we restrict our model to this case, and let the preferred direction coincide with the X-axis[2] following the axis conventions shown in Fig. 4A. Velocity then reduces to speed (a scalar, rather than a two-dimensional vector), and the stimuli can be represented by two-dimensional grayscale images $V(x, y)$. These images then form the inputs to the model (Fig. 5). Furthermore, the effect of stimulus contrast and luminance on the model neuron's response is not taken into account, since no electrophysiological recordings for contrast and luminance are available in our database.

The experimentally observed center/surround antagonism is implemented as follows. A model neuron's activation, α_{model}, is computed by subtracting the spatially pooled contribution of the surround from that of the center. The result is set to 0 if $\alpha_{\text{model}} < 0$

[2]This can be done without loss of generality: experimental evidence [31] indicates that there is no correlation between a cell's preferred direction and its tilt selectivity, and we will not compare between different model neurons.

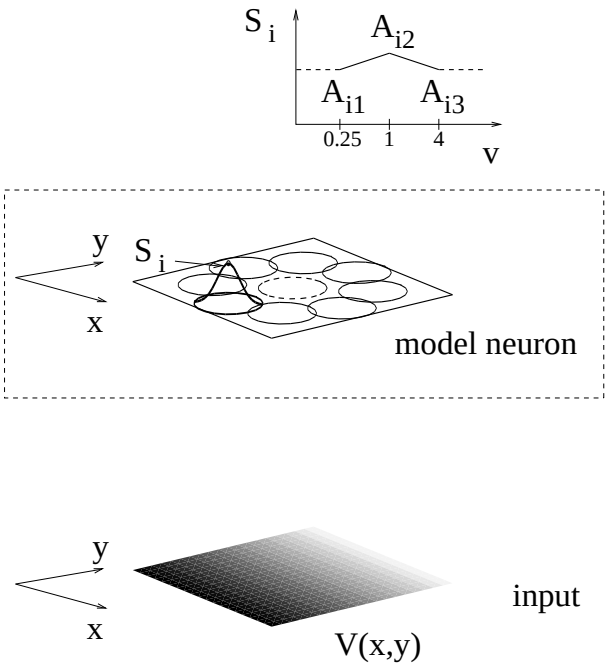


Figure 5: Model architecture. Spatial layout and local speed tunings of the classical receptive field and surround. The input (gray shaded plane) represents the speed distribution $V(x, y)$, with x and y the retinal coordinates. The spatial layout of the classical receptive field and surround results from the central Gaussian and the summation of the 8 surrounding Gaussians, with amplitudes S_i and C_i, respectively. The σ-borders of these Gaussians are depicted as dashed and full circles. The local speed tuning for Gaussian i, $S_i(v)$, is shown in the top curve. It is determined by three values (A_{i1}, A_{i2} and A_{i3}) using the inter- (solid line) and extrapolation strategy (dashed line) explained in the Model Architecture and Training Section.

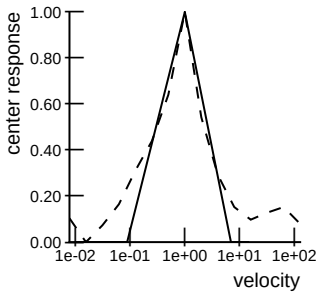


Figure 6: Normalized speed tuning for the center. The dashed curve is taken from [21] (Fig. 6B). The solid curve is generated by linearly interpolating the points at speeds 0.25, 1 and 4, and represents the center speed tuning that has been used in the model.

(note that the model's normalized activation of 0 is equivalent to the spontaneous firing rate in the corresponding biological cell). The model neuron's net response then equals the (linear) difference between the center and the surround contributions; the individual center and surround contributions are non-linear functions of the input velocity field $V(x, y)$.

Center The classical receptive field is modeled as a Gaussian, $\mathcal{G}(x, y)$, (see Fig. 5, dashed circle; the Gaussian width is 5 pixels) with an amplitude C, which is a function of speed. The center contribution $\mathcal{F}_{\text{center}}$ at spatial location (x, y) (in retinal coordinates with respect to the receptive field center) is equal to the local speed, $V(x, y)$, times the response characteristic which depends on $V(x, y)$:

$$\mathcal{F}_{\text{center}}(x, y) = V(x, y)\ C(V(x, y))\ \mathcal{G}(x, y). \tag{1}$$

The value of $C(v_{\text{opt}})$ is chosen in such a way that the center produces a unit response in the control condition of the SA-speed test and the center speed tuning characteristic is further modeled after Maunsell and Van Essen [21] (see Fig. 6). In this way, all model neurons possess identical classical receptive field layouts and center speed tunings, so that differences in the responses of the model neurons are attributable to differences in the surround characteristics only. Indeed, it has been shown that motion processing properties such as tilt selectivity are induced by the surround only, thus, not by the classical receptive field [31].

Surround In the experiments, the surround inhibition has been systematically verified for eight peripheral positions. Therefore, we model the surround as a summation of eight Gaussians, $\mathcal{G}_i(x, y)$ (see Fig. 5, solid circles; Gaussian widths are set to 5 pixels), with amplitudes S_i, which are functions of speed. Note that we do not claim that one MT cell is receiving input from eight neighboring MT cells (or other motion selective cells): rather, the spatial resolution reflects the experimental paradigm (SA-speed test). Similar to the center, the surround response characteristic depends on $V(x, y)$ and the surround contribution $\mathcal{F}_{\text{surround}}$ at location (x, y) is given by:

$$\mathcal{F}_{\text{surround}}(x, y) = \sum_{i=1}^{8} V(x, y)\ S_i(V(x, y))\ \mathcal{G}_i(x, y). \tag{2}$$

As a result, the surround's layout is allowed to be spatially heterogeneous, as described in [30]. Furthermore, the surround layout also varies with speed, since every spatial location has a *local* speed tuning curve. Indeed, the SA-speed test results of Orban and

co-workers [33] suggested that the local speed tuning characteristic changes over the surround. The local speed tuning curves will be obtained with a training algorithm (see Appendix A).

A model neuron is, thus, characterized by 8 local speed tuning curves for the surround (note that the center speed tuning is fixed and identical for all model neurons). The tuning curve at position i in the surround is defined by three parameters, A_{i1}, A_{i2} and A_{i3}, *i.e.*, the contributions for, respectively, $v = 0.25$, $v = 1$ and $v = 4$. Intermediate values are found by a piece-wise-linear (first-order) interpolation on a logarithmic scale for speeds in the interval $[0.25; 4]$. A zeroth-order extrapolation is used for speeds beyond that interval: *e.g.*, for speeds lower than $v = 0.25$, we take the S_i-value that corresponds to $v = 0.25$. The inter- and extrapolation schemes are exemplified in Fig. 5 (the interpolated parts of the speed tuning curve are drawn with solid lines, the extrapolated parts with dashed lines). In summary, there are 24 model parameters, A_{ij} ($i = 1, \ldots, 8$ and $j = 1, 2, 3$), which need to be determined. For reasons of convenience, we will use only one index for these parameters, *i.e.*, A_i ($i = 1, \ldots, 24$) in the remainder of the text.

For every biological cell considered in the MT database, a model neuron is constructed. The parameters, A_i, are adjusted using a training algorithm in order to replicate the experimental responses of that MT cell. Since there are very few training data available per biological cell (49 data points to estimate 24 parameters), a normal gradient descent training algorithm is likely to lead to an overfitting of the data. Therefore, we have added two refinements: cross-validation and a gradual increase of the number of free parameters. The complete training algorithm is described in detail and exemplified in Appendix A. The resulting 24 parameters define the local speed tuning curves of the model neuron, which is illustrated for model neuron m82 7 in the icons of Fig. 7. The surround of this model neuron has regions where the local speed tuning is lowpass (*e.g.*, at six o'clock) and where it is bandpass (*e.g.*, at one o'clock). Highpass speed tunings have also been observed, but are not shown here.

2.2.2. Model Stimuli

Three types of stimuli are used in this study: the first two are used for characterizing the surround properties of a model neuron, and the third is used for examining a model neuron's shape-from-motion capabilities.

SA-Speed Test The model stimuli for the SA-speed test (Fig. 8A) are similar to the ones for the biological experiments, and have been generated with the following specifications. The central and peripheral patches have diameters of 6.6 cm and 8.6 cm respectively[3]. Central and peripheral patches are separated by a 1 cm gap and the peripheral patches are allowed to overlap. The central patch has a normalized speed of 1 (v_{opt}), and the speed in the peripheral patches is 0.25, 1 or 4. A complete stimulus measures 25.9 cm across (33 pixels or 25.6°), so that it corresponds to the size of a model neuron's surround. The degree of inhibition is computed with respect to the response to the condition where only the center patch is stimulated. Again, the angle corresponds to that of the peripheral patch, and the length is the degree of inhibition for that condition. An example is given in Fig. 9E–F for the model neuron corresponding to cell m82 7.

[3] We will express distances in terms of centimeters, rather than in degrees of visual angle. They are not linearly related, since we are dealing with large spans.

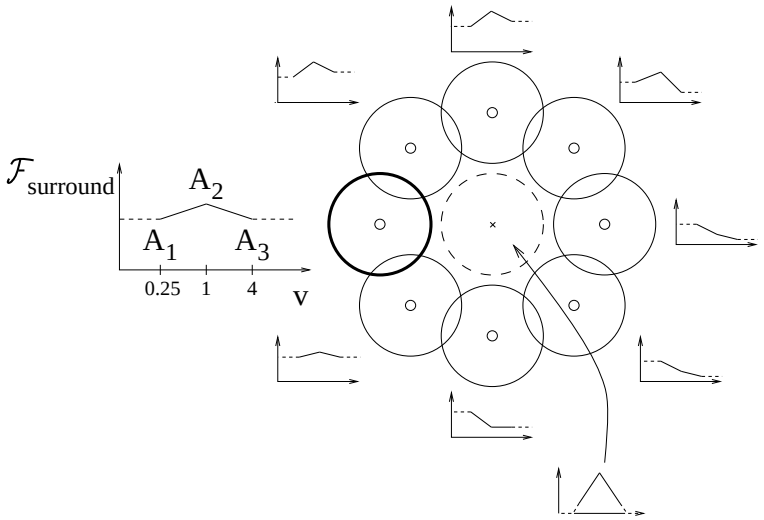


Figure 7: Local speed tuning curves for center and surround of model neuron m82 7. The spatial layout is plotted as circles corresponding to the Gaussian widths (dashed line and cross for the center and solid lines and open dots for the surround. For every Gaussian, the local speed tuning is plotted (see icons), one of which is shown in more detail (lefthand side).

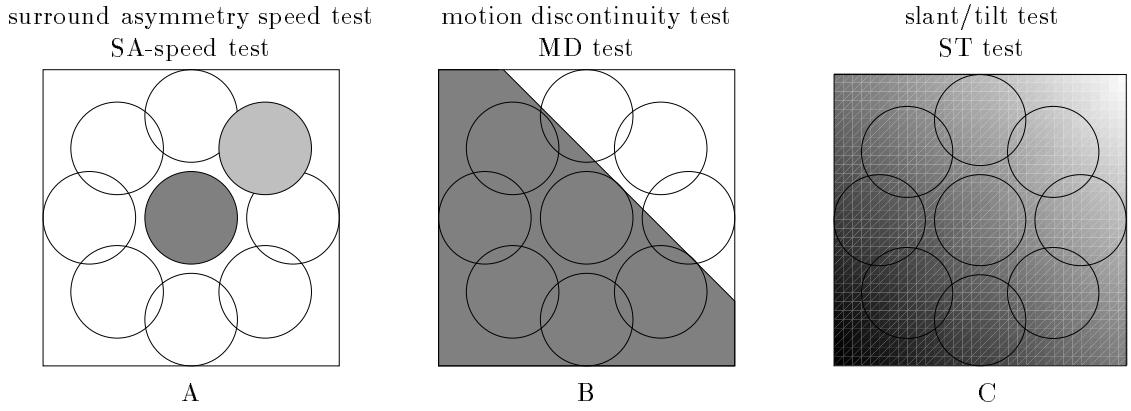


Figure 8: Model Stimuli. The velocity fields are represented by grayscale images (dark denotes low speeds). The circles represent the widths of the Gaussians that constitute center and surround receptive field. A) SA-speed test for the "fast" condition at a 45° angle. B) MD-test for the 45° condition. C) ST test for the condition of a positive slant and 45° tilt angle.

Motion Discontinuity Test A new stimulus set, the motion discontinuity test (MD test), is created in order to examine a model neuron's motion segmentation behavior. It consists of a number of uniform motion half planes (at v_{opt}) that are tangent to, and cover the center patch Fig. 5 (dashed circle). The cell responses are represented in a polar plot. The angle denotes the orientation of the motion discontinuity (the orientation in Fig. 8B equals 45°), and the vector length corresponds to the model neuron's response for that orientation. Examples are shown in Fig. 12A–D.

Slant/Tilt Test The third set of stimuli consists of velocity fields induced by translating planes of varying orientations-in-depth, expressed in terms of slant σ and tilt τ, as illustrated in Fig. 4A and B: slant σ is the angle between the inward surface normal and the line of sight (Z-axis), and tilt τ is the angle over which the slated plane is rotated around the line of sight. A moving plane induces a linear motion gradient on the frontoparallel projection plane (FP, located at $d = 57$ cm), which is related to the plane's slant and tilt. Let D be the distance between the observer o and the point where the Z-axis crosses the plane (object distance), and d the distance between the observer and the corresponding fixation point on FP (viewing distance). The velocity field induced by a plane, translating at speed T in a direction parallel to FP, is computed as:

$$V(x,y) = \frac{T}{D\ \cos\sigma}(\sin\sigma\ \cos\tau\ x + \sin\sigma\ \sin\tau\ y + d\ \cos\sigma). \tag{3}$$

We define the maximal slant $\sigma_{\max}$ as that for which all projected velocities V are nonnegative in the visual field. In our stimulus setup, the receptive field measures 25.9 cm across and $d = 57$ cm, so that $\sigma_{\max} = 72°$. A model neuron's slant/tilt selectivity plot is constructed by examining the responses to stimuli that correspond to slant/tilt combinations on a rectangular grid spanning the valid (σ, τ)-range, $\sigma \in [0°; 72°]$ and $\tau \in [0°; 360°]$. Examples are given in Fig. 14A and B.

2.2.3. Model Validation

As a result from training, there is a good correspondence between the SA-speed and ST response characteristics of the MT cells and the corresponding model neurons. This is exemplified for cell m82 7 (SA-speed response in Fig. 9A–C; ST response in Fig. 9D) and its model neuron (Fig. 9E–G and H, respectively). One should note that this correspondence is not *per se* guaranteed by the training procedure, since there are fewer parameters than there are stimulus conditions. Thus, it indicates that the model's structure is able to *explain* the surround's spatial heterogeneity at different speeds (*cf.* SA-speed test) with respect to the orientation-in-depth tuning (*cf.* ST test), and *vice versa*. More quantitatively, we can compare, for every stimulus condition, the mean squared model error with the variance on the biological data, taken over all runs. The medians over all conditions and all cells are 0.0073 for the squared model error (quartiles of 0.0006 and 0.0489), and 0.115 (quartiles of 0.0442 and 0.2278) for the variance on the biological data, respectively. Since the squared model error is much lower than the variance of the biological data, and, hence, according to statistical learning theory, it is unlikely that our models are biased estimates [6]. We can further validate our results by comparing the SMT data, which have not been used for training, to the model neuron's surround layout at v_{opt}. This is shown for two cells (Fig. 10A and C), and the corresponding model neurons (Fig. 10B and D).

Figure 11 shows surround layouts of four model neurons, evaluated at five different speeds within the [0.25; 4] speed range. It is clear that the surround layout does not

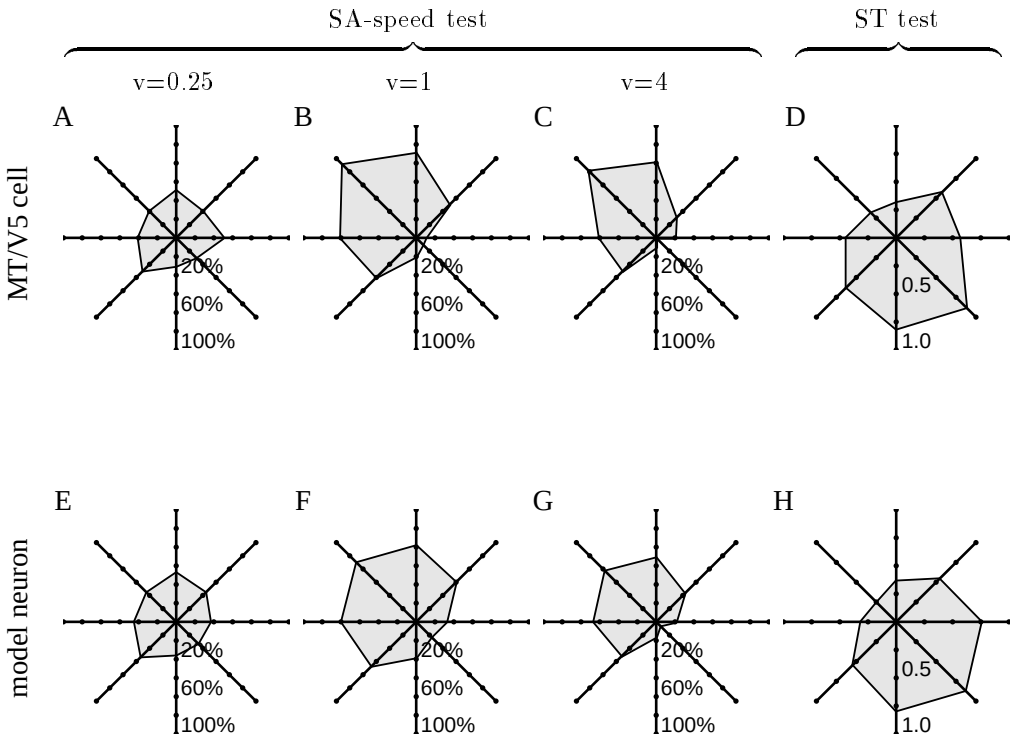


Figure 9: SA-speed and ST test results for cell m82 7 (first row; cell data taken from [38] (Fig. 4), and for the corresponding model neuron (second row). The first three columns represent the SA-speed test results (slow, medium and fast); the last column represents the ST test results for slant $\sigma = 70°$. The results are plotted in polar coordinates. In the SA-speed test results: the length of the vector denotes the degree of inhibition with respect to the response α_0 to the control condition, $(\alpha_0 - \alpha_i)/\alpha_0$, whereas the angle corresponds to that made by the surround patch. In the ST test results, the length of the vector represents the response strength, whereas its angle corresponds to the tilt angle.

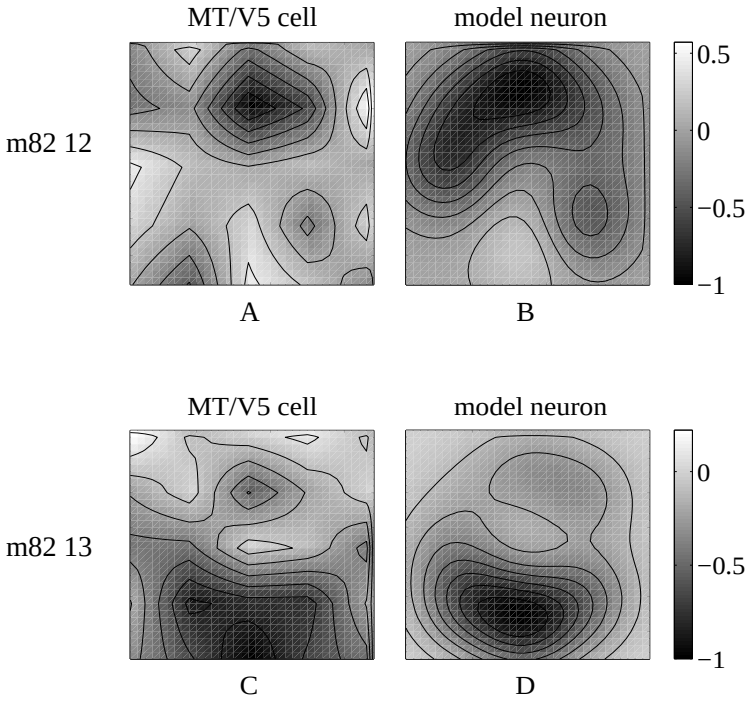


Figure 10: Spatially interpolated SMT results for cells m82 12 (A) and m82 13 (C); Surround layouts at v_{opt} for the model neurons of cells m82 12 (B) and m82 13 (D). All maps were normalized such that the maximal inhibition equals -1 (negative values indicate inhibition).

always remain of the same type (termed symmetric, asymmetric or bilateral symmetric in [27], which has also been observed in the experiments ("... the surround lost its position dependence [asymmetry], becoming circularly symmetric", p. 1323 in [30], and Fig. 2 therein).

3. Results

3.1. Motion Segmentation

Unlike the motion discontinuity detectors predicted by Nakayama and Loomis [24], our model neurons can have asymmetric surrounds. We will now examine the effect of the spatial layout of a model neuron with respect to its ability to detect motion discontinuities. More precisely, we will examine a model neuron's selectivity to the orientation of a motion discontinuity. For this purpose, the motion discontinuity test (MD test, see Materials and Methods Section) has been developed, in which an edge is defined by a motion half plane and is presented at different orientations. Figure 12 shows the MD test results (in this case implemented using 8 orientations) of the model neurons of which the surround layouts were already shown in Fig. 11. There is a close relation between the surround layout at $v = 1$ and the MD test results (also at $v = 1$): the angle of strongest inhibition in the surround layout corresponds to the preferred orientation of the motion discontinuity. This correspondence is due to the absence of stimulation in the region with the strongest inhibition: the uniform motion is present in the half-plane that is tangent to and that includes the classical receptive field (center), thus excluding the locus of strongest inhibition for the preferred MD test stimulus

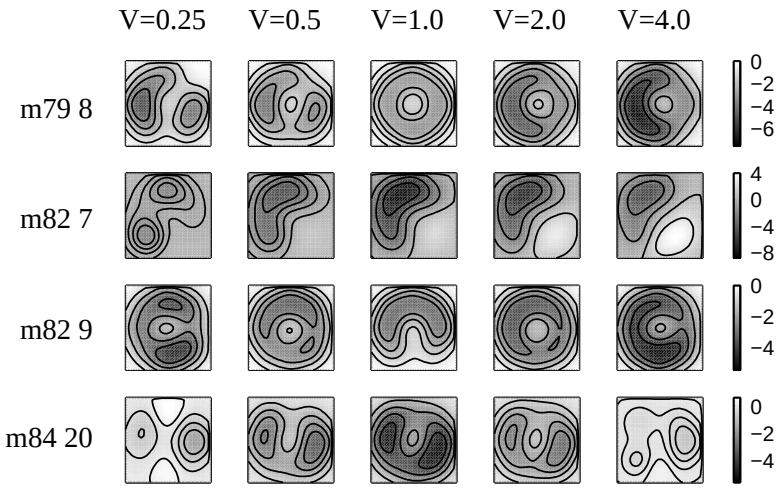


Figure 11: Surround layouts for four model neurons (rows), m79 8, m82 7, m82 9, and m84 20, and for five normalized speeds (columns), 0.25, 0.5, 1.0, 2.0 and 4.0. Surround strength is encoded in grayscales, scaled by a factor 10^{-3}; negative values indicate inhibition.

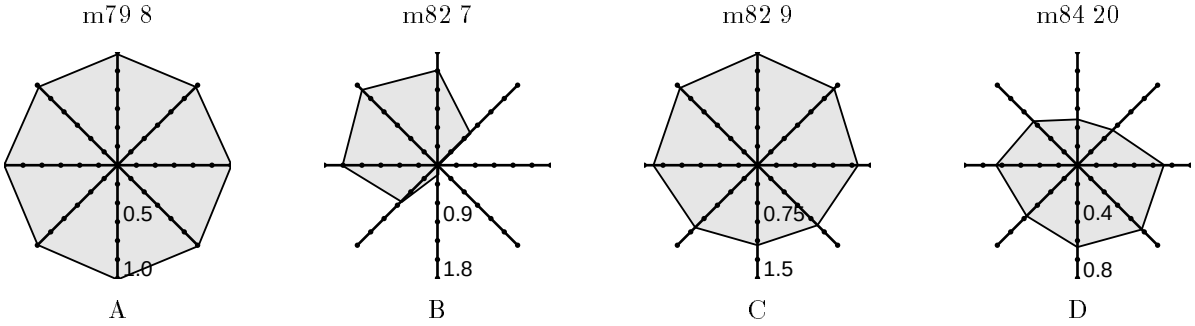
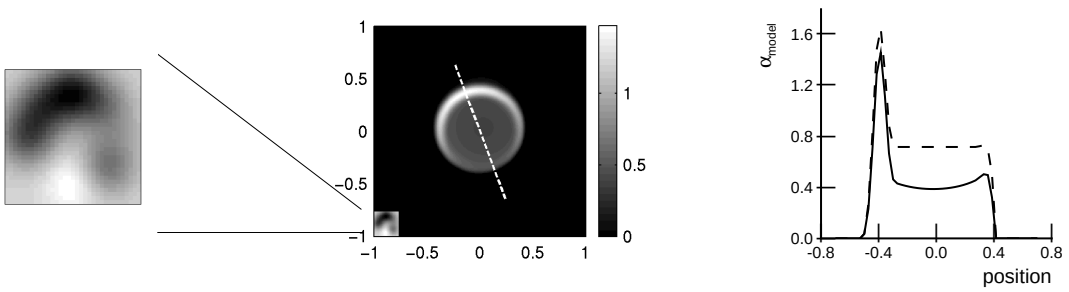


Figure 12: Polar response plots of the motion discontinuity test (MD test, in this case implemented for 8 orientations) for four model neurons m79 8, m82 7, m82 9 and m84 20. Note that these model neurons have also been used in Fig. 11. The angle denotes the orientation of the motion discontinuity and the length denotes the model neuron's response, α_{model}.

condition. Furthermore, the type of surround layout is reflected in the response polar plot: the circular symmetric surround (m79 8) corresponds to a circularly symmetric polar plot, the asymmetric surround (m82 7 and m82 9) to an asymmetric plot and the bilateral symmetric surround (m84 20) to a bimodal plot (two lobes, symmetrically extending from the origin on opposite sides of the same axis).

In order to further examine the surround's functional role in the coding of translating objects, consider the following setup. A sphere that extends well beyond the receptive field of the model neuron, translates to the right. It is positioned at a distance of 157 cm from the observer and has a diameter of 200 cm. The speed at its edge, v_{edge}, is equal to v_{opt}. The velocity field[4] is computed using the perspective projection for the complete visual field (130° or 244 cm). A model neuron's responses to different portions of the visual field, each one of which being equal in size to one surround

[4]Note that, since the input to our model consists of this velocity field, the exact definition of the object (textured or composed of dots) is not important.



A B

Figure 13: A) Right panel: Sphere response profile for the model neuron corresponding to cell m82 12 (black denotes a zero-response). As a reference, the surround is shown on the same scale in the bottom left corner, an expanded version is shown in the left panel. The maximal response lies on the top left corner of the sphere. B) A cross section along the dashed line in Fig. A. The horizontal axis denotes the position relative to the stimulus center (at position 0). The solid line corresponds to the sphere response and the dashed line to the disc response.

receptive field, are computed. They are visualized in a two-dimensional grayscale image by plotting the response at the location corresponding to the center of the stimulus portion under consideration. The result will be referred to as the sphere response profile, $\mathbf{P}_{\text{sphere}}$. Figure 13A shows the response profile for the model neuron corresponding to cell m82 12. Although the sphere is moving to the right, the model neuron's peak response lies on the top left corner of the sphere stimulus, and not orthogonal to the direction of translation, as one would intuitively expect. The lower left corner shows, on the same scale, the model neuron's surround layout at v_{opt}. The responses along the dashed line in Fig. 13A are plotted to illustrate the mechanism underlying the oriented edge selectivity. As is the case in the MD test, the model neuron responds maximally when its surround locus of strongest inhibition is not covered by the stimulus. The response builds up as the sphere "slides" over the center part of the receptive field. The peak value lies at the point where the center is completely stimulated, *without* stimulating the surround locus of highest inhibition, in a manner analogous to the situation in the MD test. The response decreases when the remaining part of the surround is covered completely. The dashed line in Fig. 13B shows the response along the same line, if we were to use a translating, flat disc, rather than a three-dimensional sphere, *i.e.*, a stimulus without first-order shape information and only the motion discontinuity present. Since the velocity fields corresponding to the inner region of the disc are uniform, the responses remain constant. Hence, it is clear that the complete sphere response cannot be explained solely by the presence of the edge: the varying part of the response in the inner region of the sphere cannot be determined by the response of the disc and originates from the sphere's shape. Therefore, the following question can be raised: is it possible to perform both edge segmentation and shape coding using the same MT cells?

3.2. Shape-from-motion

In order to examine an MT model neuron's slant/tilt-selectivity, we use velocity fields induced by translating planes with slants $\sigma \in [0°; 72°]$ and tilts $\tau \in [0°; 360°]$. The speed

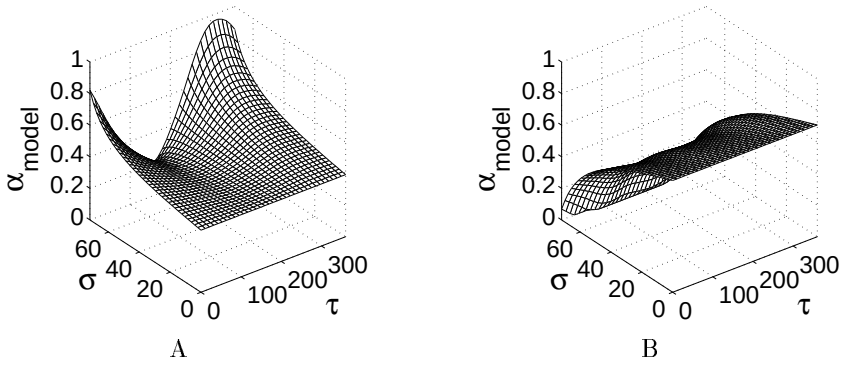


Figure 14: Slant/tilt plots of the model neurons of cells m82 7 (A) and m84 18 (B). The response α_{model} (vertical axis) is plotted as a function of slant σ and tilt τ (horizontal axes).

at the stimulus center is equal to v_{opt}. This analysis is performed on the complete set of 31 model neurons. The resulting output of the model neuron, α_{model}, is plotted as a function of σ and τ for two example model neurons in Fig. 14A and B. The model neuron in Fig. 14A responds maximally to a translating plane with a 72° slant, *i.e.*, the maximum slant in our simulations, $\sigma_{\max}$, and a 308° tilt. It is an example from a group of 20 model neurons that have their peak responses at $\sigma_{\max}$; the median of the 50% bandwidths of the tilt selectivity at $\sigma_{\max}$ for these neurons is 198°. Figure 14B shows an example of another group which consists of 9 model neurons: the activation is lower for higher slants and there is no strong selectivity for tilt. The remaining group of 2 model neurons is unresponsive to this stimulus.

In order to demonstrate the importance of the spatial heterogeneity of the surround with respect to the tilt selectivity, Orban and coworkers [31] have compared the preferred tilt in the ST test to the angle of the peripheral patch that evokes the strongest inhibition in the SA-speed test. They have found a clearly bimodal distribution (black bars in Fig. 15): the preferred tilt lies predominantly either in the same or exactly the opposite direction with respect to the locus of maximal surround inhibition. We have also performed this analysis on the subgroup of 20 model neurons that show a maximal response for $\sigma_{\max}$, since these have a clear tilt tuning. However, rather than the angle of the peripheral patch that evokes the strongest inhibition, we have used the angle made by the locus of strongest inhibition of the surround layout, since this yields a higher angular resolution. We have compared this angle to the preferred tilt for slant $\sigma = 72°$. It is evident from Fig. 15 (grey bars) that we also find a bimodal distribution, and we are, thus, well in line with the experimental results. We have also implemented the O_1^o velocity filter of Buračas and Albright (anti-symmetric first-order operator [10]), which is a model for the MT neuron with an asymmetric antagonistic surround, using the difference of two displaced Gaussians, one for the center and another for the surround. However, in order to make a fair comparison, we have modeled the surround with eight Gaussians of varying magnitude, corresponding to our model neurons' surround layouts at v_{opt}, but have further adopted their assumptions, such as a zero-response to uniform motion which is achieved by scaling the surround accordingly. We have performed the same analysis as for our model. The result is shown in Fig. 15 (white bars), indicating that the Buračas and Albright model does not yield a bimodal distribution and only predicts the opposition case. This is due to the linearity of the velocity filtering in the surround: a uniform (*i.e.*, flat) tuning characteristic is assumed and, hence, the SA-

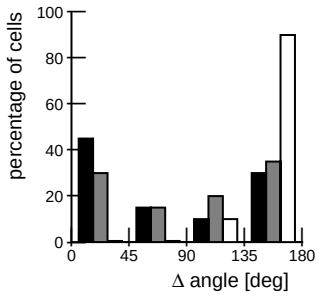


Figure 15: Distribution of the angle between the preferred tilt and the surround position evoking the strongest inhibition. Black bars denote the results from [31] (20 cells, Fig. 3D, p. 959), grey bars represent those of 20 model neurons, and white bars those of the same 20 model neurons but when adopting the assumptions of Buračas and Albright [10].

speed data and the speed dependency of the surround layout cannot be modeled. In our case, the velocity filtering is non-linear in the sense that the surround speed tuning characteristics may vary spatially and the result is pooled over the surround.

Slant and tilt cannot be extracted directly from a single neuron's response. Indeed, as pointed out by others [15, 13, 28], the response of a single model neuron is insufficient to unambiguously *code* for even a single object quality such as, in our case, the orientation-in-depth of a translating plane. Indeed, none of our model neurons has a slant/tilt selectivity that allows for an unambiguous coding of the orientation-in-depth: this would require a (narrow) selectivity to a specific slant/tilt combination for one model neuron, and a population of such neurons that covers the range of possible slant/tilt combinations. However, a population of model neurons might do so. The question is then: what could be the nature of the population and of the mechanism needed to *crack* the population code?

An object's orientation-in-depth can be represented as a vector in a polar space where the object's slant σ is denoted by the vector's length, and the tilt τ by the vector's angle. Such a representation allows for a vectorial coding scheme by a population of N model neurons with different preferred tilt angles τ_n. Note that the population should not display a net tilt preference (*null* preference). Experimental evidence [31] has shown that this is indeed the case in MT, since the preferred tilts in MT cover the entire range of possible values. We generated such a population by rotating the *same* model neuron's surround, and, thus, its tilt selectivity curve, $N-1$ times over $\frac{360°}{N}$ (see Fig. 16 for $N = 6$). In this way, the population has a null tilt preference, and we can easily control N, the number of constituting model neurons[5]. The response of model neuron n, $\mathbf{R}_n$, $n = 1, \ldots, N$, can be represented as a vector, with angle $\frac{(n-1)\times 360°}{N}$, and length equal to the model neuron's activation (Fig. 16). The neuronal population vector $\mathbf{P}$ is computed as the sum of these vectors:

$$\mathbf{P} \;=\; \sum_{n=1}^{N} \mathbf{R}_n, \tag{4}$$

and its length $\|\mathbf{P}\|$ and angle $\theta(\mathbf{P})$ linearly code for slant σ and tilt τ. Hence, a linear coding *mechanism* is assumed, the details of which are described in Appendix B. Note

[5] Note that if we were to use different model neurons, we would need a separate mechanism to ensure a null tilt preference.

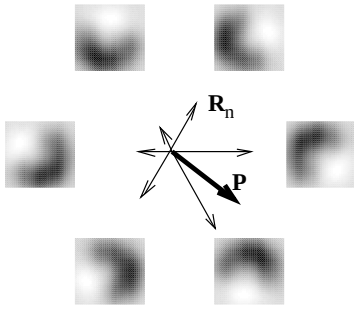


Figure 16: Schematic representation of the population coding scheme. The six constituting model neurons are rotated versions of the model neuron of cell m82 7 (black denotes maximal inhibition). The neuronal population vector $\mathbf{P}$ (thick arrow) is computed as the vector sum of the response vectors of the separate model neurons, $\mathbf{R}_i$ (thin arrows).

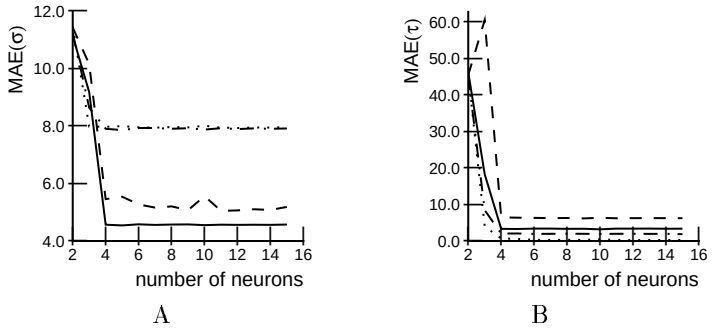


Figure 17: Mean Absolute Error (MAE) on slant σ (A) and tilt τ (B) for populations generated using the model neurons of cells m79 16 (solid line), m82 13 (dashed line), m84 19 (dot-dashed line) and m82 7 (dotted line). The first three model neurons correspond to the cells shown in Xiao *et al.*, ([33], Fig. 2D–F).

that this is not the same as assuming a linear *code*: since a single model neuron does not linearly code for, *e.g.*, slant, the code will also be non-linear at the population level.

We evaluate the Mean Absolute Error (MAE) between the true slant and tilt, and the corresponding estimates derived from the population, for slant and tilt values on a uniform 15×15 grid in (σ, τ) space, with $\sigma \in [10°; 72°]$ and $\tau \in [0°; 360°]$. In the set of 29 model neurons that respond to a translating plane, a relatively precise code for slant and tilt is found for a population of $N = 6$ model neurons (median values taken over all model neurons: $\mathrm{MAE}_{50}(\sigma) = 8.09°$ and $\mathrm{MAE}_{50}(\tau) = 6.33°$). When the number of model neurons in the population is further increased, performance does not improve. Figure 17 shows the MAE(σ) and MAE(τ) as a function of the number of model neurons, for four example populations of which the model neurons correspond to the cells displayed in Fig. 2D–F of [33] (p. 1325) and another example cell from the same database.

4. Discussion

Since its discovery, the antagonistic MT surround has been associated with motion segmentation and the detection of image discontinuities resulting from edges of moving

objects [3, 26, 1, 2]. In support of this view, Bradley and Andersen [8] recently found evidence for a center-surround antagonism based on disparity in macaque MT, which operates on top of the velocity antagonism, so that not only discontinuities based on motion could be detected, but also those based on binocular disparity. The motion discontinuity test used in this article (MD test) revealed that most of our model neurons display a selectivity to the orientation of a motion half plane, corresponding to a flat object moving against a static background. For other (non-flat) objects, such as the sphere used in this modeling study, we have shown that the model neurons display a preference for a portion of the sphere's edge spanning orientations within a particular range. The observed response is largely due to the motion discontinuity. The question is now: can this mechanism be applied to real-world situations, where the motion discontinuity originates from relative motion between the observer and a solid 3D object and/or from eye movements, due to which the background velocity will be non-zero?

In order to address this question, we have extended the motion discontinuity test by varying the speed in the "background", v_{back}, *i.e.*, in the part of the receptive field that is not covered by the motion half plane in Fig. 8B. For every such speed, we examine the model neuron's activation for different orientations ϕ of the motion discontinuity ("edge"). Examples are shown in Fig. 18A–D for the model neurons that have also been used in Fig. 11 and Fig. 12. Figures 18B and D show tuning curves that can be used to perform an oriented edge segmentation in the presence of background motion: higher activations correspond to situations where the discontinuity is caused by a slow background motion (Fig. 18B), or by a background motion different (slower or faster) than the plane's (Fig. 18D). The model neurons from Fig. 18A and C have their peak responses when stimulated with a frontoparallel plane ($v_{\text{back}} = v_{\text{plane}} = 1$). They correspond to the group of 9 model neurons that showed a lower activation for higher slants (see Results Section). At the end of this section, we will show that these types of responses can be used to analyze a visual scene based on motion information.

Orban and co-workers [31] have also investigated the correspondence between the spatial heterogeneity of the surround and the observed tilt selectivity (results are shown in Fig. 15, black bars). Since the preferred orientation of a motion discontinuity lies in the same direction as the surround locus with strongest inhibition (see Results Section), this analysis would predict that the *preferred* (linear) edge would lie in the same or the opposite orientation of the preferred tilt angle. However, this is not what we would expect, since in real-world visual scenes, where motion discontinuities often originate from occlusions with other objects, the occlusion border can lie in any orientation with respect to the object's surface orientation-in-depth. As an example, consider an observer viewing, through a window, an object translating to the right. The orientation of the motion discontinuity at the window's edge will be due to the window only, irrespective of the object's surface orientation-in-depth. The MT model now allows for a more detailed analysis of this behavior: rather than restricting ourselves to the orientation of the surround locus with strongest inhibition, we can use the MD test and correlate the resulting MD selectivity curve with the model neuron's tilt tuning curve for the highest slant, $\sigma_{\max}$. Both curves are generated for 50 values between 0° and 360° on the horizontal axis and they are DC-compensated and normalized prior to computing the correlation. Examples are shown in Fig. 19A and B for model neurons with a correlation of -0.989 and -0.335. Figure 19C shows that, rather than a bimodal distribution with peaks at -1 and 1, we find a continuum of correlation values. Thus, a model neuron's preferred (linear) edge will have both a preferred tilt *and* a preferred edge orientation, not necessarily the same or opposite.

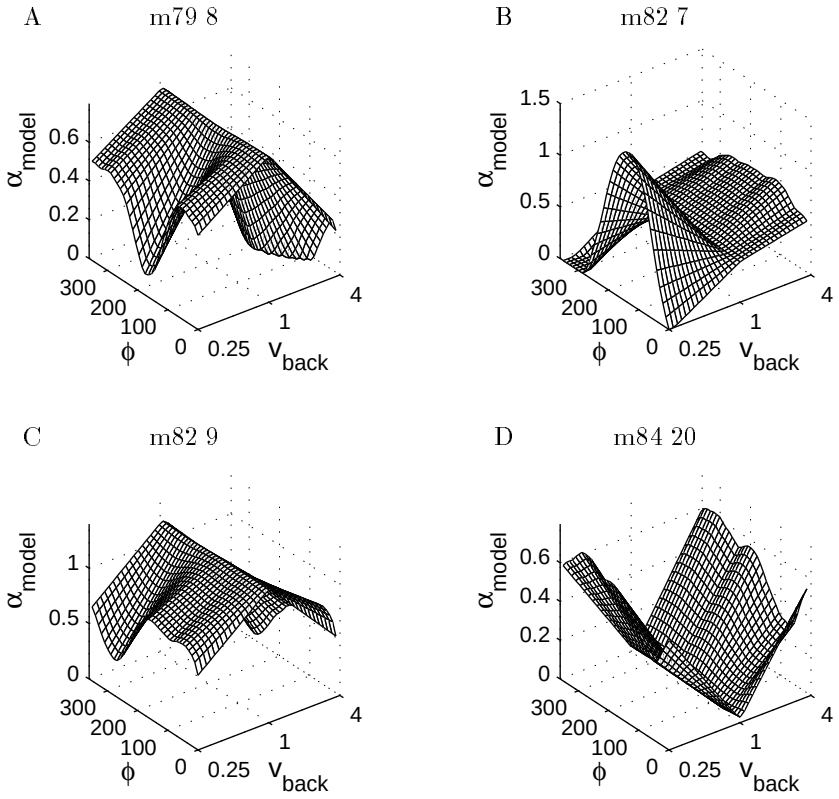


Figure 18: Motion discontinuity tuning for different speeds in the background. Results are shown for the model neurons corresponding to cells m79 8, m82 7, m82 9 and m84 20. Note that the orientation of the motion discontinuity, ϕ, does not influence the input when $v_{\text{back}} = 1$, since in that case, the stimulus consists of a uniform velocity field.

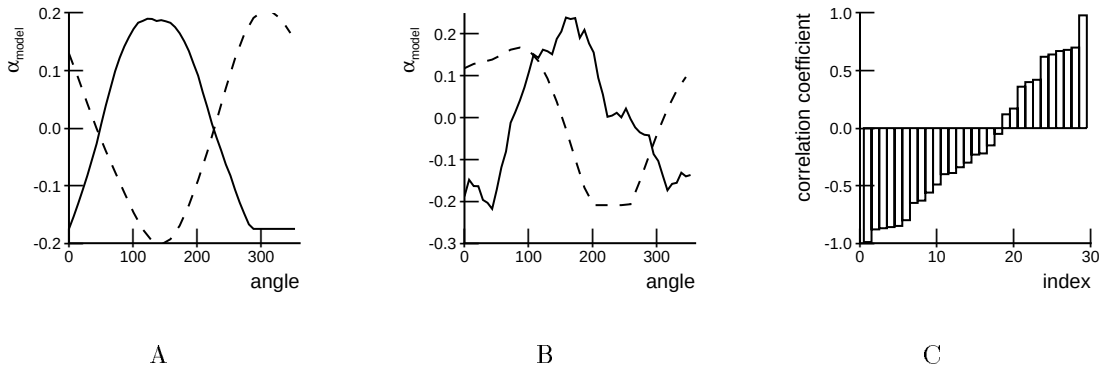


Figure 19: Correlation between normalized tilt tuning curves (dashed line) and MD test results (solid line); both test have been performed for 50 orientations. Examples are shown for model neurons A) m82 7 (correlation value of -0.989) and B) m84 19 (correlation value of -0.335). C) Correlation values of all model neurons (in ascending order).

We have also examined MT's ability to *code* for a translating plane's orientation-in-depth. Clearly, a single neuron is unable to perform such a task, since the orientation-in-depth consists of two components, namely slant and tilt. Also, the slant/tilt selectivities of the model neurons are such that a winner-takes-all (WTA) mechanism over a population of neurons will not be sufficient to code for slant. Indeed, the preferred slants are either maximal or zero, and for a WTA mechanism to work, the neurons in the population should display selectivity to different slant/tilt combinations and cover the entire slant/tilt space. Therefore, we have taken a different approach which takes into account the response of a single population, which is jointly representing slant and tilt. We introduce a simple linear mechanism, which takes into account the dependencies between the slant and tilt selectivities, for "reading" the code with reasonable precision (Fig. 17). The population consists of identical, but rotated model neurons that process the same part of the visual field. The precision of the slant/tilt population code of the corresponding model neurons is shown in Fig. 17, and it reveals that about six model neurons suffice to achieve a relatively precise slant/tilt code; using more model neurons did not lead to an improvement, possibly due to the assumed linearity of the coding mechanism. However, larger populations might be necessary if the cell outputs are noisy: using more neurons will make the coding mechanism more robust. These modeling results allow us to predict that populations of MT cells code for the first-order surface shape descriptors slant and tilt and that a linear mechanism suffices for reading the code. However, effects that would influence the response characteristics of our model, such as contrast and luminance, could also alter the requirements of the population coding mechanism.

We will now describe a possible setup for analyzing a visual scene based on motion information. We assume that the velocity field is caused by a relative motion between observer and object, as depicted in Fig. 1. Since the observer's motion is leftwards. the entire scene moves rightwards, but the object moves at a speed different from the (frontoparallel) background. In order to analyze this visual scene, we need to distinguish between the background motion, the object's edge and the object's "body". We will now illustrate how a distinction can be made between these regions by comparing the activations of different neurons. In order to do so, we need to consider neurons with two types of MD tuning curves.

Background The first type of neuron should prefer a full-field, frontoparallel plane to a full-field, slanted plane or a motion half-plane. Its edge and slant/tilt selectivities are of the type shown in Fig. 18A and C, and Fig. 14B. A group of such neurons, with different preferred speeds (of both center and surround) can signal background motion: there is only background motion present (uniform motion within the receptive field) if the most active neuron of the group reaches its maximal response.

Object Edge and Body The second type of neuron should respond maximally in the presence of a motion discontinuity under a certain orientation. Its edge and slant/tilt selectivities are of the type shown in Fig. 18B and D, and Fig. 14A. A group of such neurons, with different preferred speeds and edge orientations, would be able to detect an object's edge, namely if the most active neuron reaches its maximal response, assuming that there is no background motion present. In the presence of background motion, the response would be non-maximal, which could also be due to a full-field, slanted plane. Thus, such a response yields ambiguous information about the scene. However, if the surround speed tuning were narrow enough, causing the background speed to lay outside of the neuron's speed region, the response *would* be maximal. Therefore, such

"narrowband" cells would be able to signal the object's edge. The other neurons, the "broadband" cells, could code for the surface orientation-in-depth of the object using the population coding scheme explained earlier. They could construct a first-order representation of the object (surface interpolation) if there is neither background motion, nor a motion discontinuity present within their receptive field (which can be detected in the manner described in the previous two cases). This still leaves open the possibility that slant coding is supported by a population of MT cells with a mix of preferred speeds and spatial locations (Buračas, personal communication).

Appendix A: Training Algorithm

The MT model is characterized by 24 free parameters, A_i, $i = 1 \ldots 24$. These are adjusted in order to let the model replicate the biological data. We determine the parameters with a training algorithm which performs gradient descent on a cost function E, *i.e.*, the sum of the squared differences between the model output $\alpha_{\text{model},k}$ and the normalized experimental data α_k for all stimulus conditions k, $k = 1, \ldots, 49$ ("fitting"):

$$E = \sum_k E_k = \sum_k (\alpha_k - \alpha_{\text{model},k})^2. \tag{5}$$

For the training data, we take the median values of the SA-speed and ST test data, over all runs. The gradient vector for stimulus condition k is approximated by examining the changes ΔE_k when parameter A_i is incremented by a small fixed value ε:

$$\frac{\partial E}{\partial A_i} \approx \sum_k \frac{\Delta E_k(\mathbf{A})}{\varepsilon} \triangleq \sum_k \frac{E_k(\mathbf{A}_i) - E_k(\mathbf{A})}{\varepsilon}, \tag{6}$$

where $E_k(\mathbf{A})$ and $E_k(\mathbf{A}_i)$ are the squared differences with respect to stimulus condition k, for parameter sets $\mathbf{A} = \{A_1, \ldots, A_{24}\}$ and $\mathbf{A}_i = \{A_1, \ldots, A_i + \varepsilon, \ldots, A_{24}\}$, respectively. The batch training rule for parameter set A_i is as follows:

$$\Delta A_i = -\frac{\eta}{49} \sum_{k=1}^{49} \frac{E_k(\mathbf{A}_i) - E_k(\mathbf{A})}{\sqrt{\sum_{r=1}^{24} (E_k(\mathbf{A}_r) - E_k(\mathbf{A}))^2}}. \tag{7}$$

Note that the denominator corresponds to a normalization step, due to which every partial update has a fixed magnitude. However, since there are very few training data available (49 samples to estimate 24 parameters), such a training algorithm is likely to lead to an overfitting of the data. Therefore, we add two refinements: cross-validation and a gradual increase of the number of free parameters.

First, we include a cross-validation strategy in order to improve the model's generalization performance. The training algorithm is split into different runs (Fig. 20), with every run j consisting of T iterations. In every run j, P models are trained, starting from identical configurations and, for every model p, $p = 1 \ldots P$, a different subset C_p of the training set is excluded and reserved for measuring the generalization performance (cross-validation). During training, the mean squared error for C_p, $\text{MSE}_{\text{test},p}$, is monitored. The final configuration for every model p is the one for which $\text{MSE}_{\text{test},p}$ is minimal (early stopping). After every run j, we compute the "*gain*" in generalization performance for every model (Fig. 20, numbers in rectangles): the ratio between the $\text{MSE}_{\text{test},p}$ at the beginning of the run and the model's minimal one for this run. The parameter values of the model with the highest gain are used as the initial configuration for the next training run (Fig. 20, arrows). This process is iterated through until the

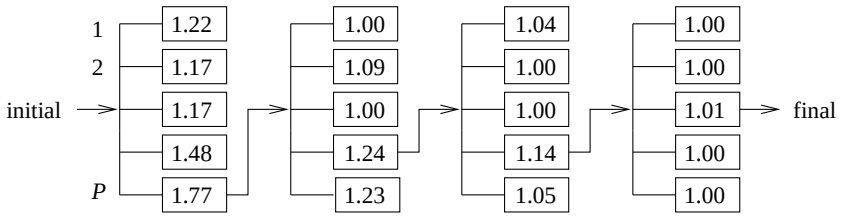


Figure 20: Typical training session consisting of four runs (model neuron m82 7, 24 parameters case). At every training run, $P = 5$ models are trained (5 boxes per column). The numbers displayed in the rectangles are the gains in generalization performances obtained for the corresponding models. The model with the highest gain is used as the initial configuration for the next run (indicated by arrows). The final model is found after the fourth run, since there, the highest gain does not reach the 1.05 threshold.

highest gain improvement over the P models is below a certain threshold. In our simulations we have set this value to 1.05. Figure 20 schematically shows a typical training session.

Second, we gradually increase the number of free parameters that specify the local speed tuning characteristics of the surround so as to control the model's complexity: the training algorithm is less likely to overfit the data if the number of free parameters is lower. More in particular, we start with the bandpass filter which corresponds to that used for the center (Fig. 6), but allow its magnitude to be scaled by a single free parameter. This leads to 8 model parameters to be estimated. Next, we use a symmetric bandpass filter where the values for the slow and fast speeds are identical so that it is specified by two free parameters. Hence, we now have 16 model parameters. Finally, we use the piece-wise linear filter explained earlier in the Materials and Methods Section (24 parameters). We start by training for the 8 parameters case until convergence. We then use these parameter values as a starting point for the 16 parameters case, which we in turn use as a starting point for the 24 parameters case. The mean performance ($\text{MSE}_{\text{train}}$) for the consecutive stages are: 0.0256 (0.0165–0.0450), 0.0249 (0.0135–0.0351) and 0.0187 (0.0095–0.0306), with the values between brackets the lower and upper quartiles. If we only apply the cross-validation strategy, thus without gradually increasing the number of parameters, $\text{MSE}_{\text{train}} = 0.0198$ (0.0112–0.0315).

Each run consists of $P = 5$ models and is trained for $T = 100$ epochs with a training rate of $\eta = 0.0003$. In the 24 parameters case, a complete session requires 3 runs on average to reach the 1.05 gain criterion.

Appendix B: Population Code

The neuronal population vector $\mathbf{P}$ is computed as the vector sum of the separate response vectors (Fig. 16). The length $\|\mathbf{P}\|$ and angle $\theta(\mathbf{P})$ are linear estimates of the slant σ and tilt τ:

$$\tau = \theta(\mathbf{P}) + \phi \tag{8}$$

$$\sigma = c_1 \, \|\mathbf{P}\| + c_2. \tag{9}$$

A reasonable value for ϕ can be obtained by taking the mean deviation over a dataset $\mathcal{V}$ covering the slant/tilt range of interest. We perform a linear regression to fit $(\|\mathbf{P}\|, \sigma)$ over $\mathcal{V}$ to determine the coefficients c_1 and c_2. We choose $\mathcal{V}$ as the slant/tilt values on a 10×10 (σ, τ) grid, with $\sigma \in [10°; 72°]$ and $\tau \in [0°; 360°]$.

Acknowledgments

The authors wish to express their gratitude to Giedrius Buračas (Salk Institute, USA), Cyril Goutte (DTU, Denmark), Reinoud Maex (University of Antwerp, Belgium), and Stefan Treue (University of Tübingen, Germany) for helpful comments on earlier versions of the manuscript. Special thanks to Guy Orban (K.U.Leuven, Belgium) for his continuous encouragement and for letting us use his MT/V5 database. T.G. is supported by a scholarship from the Flemish Institute for the promotion of Scientific Technological Research in Industry (I.W.T.). M.M.V.H. is a research associate of the Fund for Scientific Research – Flanders (Belgium) and is supported by research grants received from the Fund for Scientific Research (G.0185.96), the National Lottery (Belgium) (9.0185.96), the Flemish Regional Ministry of Education (Belgium) (GOA 95/99-06; 2000/11), and the Flemish Ministry for Science and Technology (VIS/98/012).

References

[1] Albright, T. D. (1984), Direction and orientation selectivity of neurons in visual area MT of the macaque, *J. Neurophysiol.*, **52**, 1106–1130.

[2] Albright, T. D (1993), Cortical processing of visual motion and its role in the stabilization of gaze. In: Visual motion and its role in the stabilization of gaze (Miles, F. A. & Wallman, J., eds.), Amsterdam: Elsevier Science Publishers, pp. 177–201.

[3] Allman, J., Miezin, F. & McGuinness, E. (1985a), Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal area (MT), *Perception*, **14**, 105–126.

[4] Allman, J., Miezin, F. & McGuinness, E. (1985b), Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local-global comparisons in visual neurons, *Annu. Rev. Neurosci.*, **8**, 407–430.

[5] Barlow, H. B. (1961), Possible principles underlying the transformations of sensory messages. In: Sensory Communication (Rosenblith, W. A., ed.), New York: Wiley, pp. 217–234.

[6] Bishop, C. M. (1995), Neural Networks for Pattern Recognition. New York: Oxford University Press.

[7] Born, R. T., & Tootell R. B. (1992), Segregation of global and local motion processing in primate middle temporal visual area, *Nature*, **357**, 497–499.

[8] Bradley, D. C. & Andersen, R. A. (1998), Center-surround antagonism based on disparity in primate area MT, *J. Neurosci.*, **18**, 7552–7565.

[9] Buračas, G. T & Albright, T. D (1994), The role of MT neuron receptive field surrounds in computing object shape from velocity fields, *Advances in Neural Information Processing Systems*, **6**, 969–976.

[10] Buračas, G. T & Albright, T. D (1996), Contribution of area MT to perception of three-dimensional shape: a computational study, *Vision Res.*, **36**, 869–888.

[11] Droulez, J. & Cornilleau-Pérès, V. (1990), Visual perception of surface curvature. The spin variation and its physiological implementation, *Biol. Cybern.*, **62**, 211–224.

[12] Dubner, R. & Zeki, S. M. (1971), Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey, *Brain Res.*, **35**, 528–532.

[13] Georgopoulos, A. P., Taira, M. & Lukashin, A. (1993), Cognitive Neurophysiology of the Motor Cortex, *Science*, **260**, 47–52.

[14] Gibson, J. J. (1950), The perception of the visual world. Boston: Houghton Mifflin.

[15] Hinton, G. E., McClelland, J. L. & Rumelhart, D. E. (1986), Distributed representations. In: Parallel Distributed Processing (Rumelhart, D. E., McClelland, J. L., eds.), vol. 1, Cambridge, MA: MIT Press, pp. 77-109.

[16] Hussain, M., Treue,S. & Anderson, R.A. (1989). Surface Interpolation in Three-dimensional Structure-From-Motion Perception. *Neural Computation*, **1**, 324–333.

[17] Koenderink, J. J. & van Doorn, A. J. (1992), Second order optic flow, *J. Opt. Soc. Am. A.*, **9**, 530–538.

[18] Lagae, L., Raiguel, S. & Orban, G. A. (1993), Speed and direction selectivity of macaque middle temporal neurons, *J. Neurophysiol.*, **69**, 19–39.

[19] Lamme, V. A. F. (1995), The neurophysiology of figure-ground segregation in primary visual cortex, *J. Neurosci.*, **15**, 1605–1615.

[20] Liu, L. & Van Hulle, M.M. (1998). Modeling the Surround of MT Cells and their Selectivity for Surface Orientation In Depth Specified by Motion. *Neural Computation*, **10**, 295–312.

[21] Maunsell, J. H. R. & Van Essen, D. C. (1983), Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, *J. Neurophysiol.*, **49**, 1127–1147.

[22] Marcar, V. L., Xiao, D.-K., Raiguel, S. E., Maes, H. & Orban, G. A. (1995), Processing of kinetically defined boundaries in the cortical motion area MT of the macaque monkey, *J. Neurophysiol.*, **74**, 1258–1270.

[23] Movshon, J. A., Adelson, E. H., Gizzi, M. S. & Newsome, W. T. (1985), "The analysis of moving visual patterns. In: Pattern Recognition Mechanisms (Chagass, C., Gattas, R. & Gross, C., eds.), Vatican City: Pontificia Academia Scientiarium, pp. 117–151.

[24] Nakayama, K. & Loomis J. M. (1974), Optical velocity patterns, velocity-sensitive neurons, and space perception: A hypothesis, *Perception*, **3**, 63–80.

[25] Raiguel, S., Van Hulle, M. M., Xiao, D.-K., Marcar, V. L. & Orban, G. A. (1995), Shape and spatial distribution of receptive fields and antagonistic motion surrounds in the middle temporal area (V5), of the macaque, *Eur. J. Neurosci.*, **7**, 2064–2082.

[26] Tanaka, K., Hikosaka, H., Saito, H., Yukie, Y., Fukada, Y. & Iwai, E. (1986), Analysis of local wide-field movements in the superior temporal visual area of the macaque monkey, *J. Neurosci.*, **6**, 134–144.

[27] Treue, S., Andersen, R.A., Ando, H. & Hildreth E.C. (1995). Structure from motion: Perceptual evidence for surface interpolation. *Vision Res.*, **35**, 139–148.

[28] Treue, S., Hol, K. & Rauber, H.-J. (2000), Seeing multiple directions of motion – physiological and psychophysics, *Nature Neurosci.*, **3**, 270-276.

[29] Warren, W. H. (1995), Self-motion: Visual perception and visual control. In: Handbook of perception and cognition Vol 5: Perception of space and motion, pp. 263–325. Los Angeles: Academic Press.

[30] Xiao, D.-K., Raiguel, S., Marcar, V., Koenderink, J. & Orban, G. A. (1995), Spatial heterogeneity of inhibitory surrounds in the middle temporal visual area, *Proc. Natl. Acad. Sci. USA*, **92**, 11303–11306

[31] Xiao, D.-K., Marcar, V. L. , Raiguel, S. E. & Orban G. A. (1997a), Selectivity of macaque MT/V5 neurones for surface orientation in depth specified by motion, *Eur. J. Neurosci.*, **9**, 956–964.

[32] Xiao, D.-K., Raiguel, S. E. & Orban, G. A. (1997b), The Spatial Distribution of the Antagonistic Surround of MT/V5 Neurons, *Cereb. Cortex*, **7**, 662–677.

[33] Xiao, D.-K., Raiguel, S. E., Marcar, V. & Orban, G. A. (1998), "Influence of stimulus speed upon the antagonistic surrounds of area MT/V5 neurons, *NeuroReport*, **9**, 1321–1326.

[34] Zeki, S. M. (1974), Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey, *J. Physiol.*, **236**, 549–573.