Research Article **Open Access**

Fernanda da C. e C. Faria*, Jorge Batista, and Helder Araújo

Biologically inspired computational modeling of motion based on middle temporal area

https://doi.org/10.1515/pjbr-2018-0005
Received November 30, 2017; accepted March 3, 2018

Abstract: This paper describes a bio-inspired algorithm for motion computation based on V1 (Primary Visual Cortex) and MT (Middle Temporal Area) cells. The behavior of neurons in V1 and MT areas contain significant information to understand the perception of motion. From a computational perspective, the neurons are treated as two dimensional filters to represent the receptive fields of simple cells that compose the complex cells. A modified elaborated Reichardt detector, adding an output exponent before the last stage followed by a re-entry stage of modulating feedback from MT, (reciprocal connections of V1 and MT) in a hierarchical framework, is proposed. The end-stopped units, where the receptive fields of cells are surrounded by suppressive regions, are modeled as a divisive operation. MT cells play an important role for integrating and interpreting inputs from earlier-level (V1). We fit a normalization and a pooling to find the most active neurons for motion detection. All steps employed are physiologically inspired processing schemes and need some degree of simplification and abstraction. The results suggest that our proposed algorithm can achieve better performance than recent state-of-the-art bio-inspired approaches for real world images.

Keywords: Motion Direction, Neural Computational Model, Area MT

1 Introduction

A biologically motivated computational model can be derived from neurophysiological studies. They give us essential information to develop biologically inspired approaches. Understanding the principles that underlie the basic known functional brain structures reveals information about how the concepts link with the general mathematical models.

The neural processing of perception of motion starts in the eye. The observed scene has, at least, two spatially separate images projected onto the 2D retinal surface. Several cortical areas are involved to give an interpretation of the data. To study this phenomenon based on two cortical areas, information of physiologically plausible characteristics that integrates visual motion signals are used to guide an appropriate algorithmic model to simulate the logic of motion computation.

Understanding how biological neural motion detection system is organized for processing of optic flow, we propose a model to estimate the direction and speed of image sequences. We describe a model that is not spatially restricted to the classical receptive field. Instead, we consider the area surrounding the receptive field center. The neural motion information perception is studied in this work using an elaborated Reichardt detector (original Reichardt detector plus spatiotemporal filters). As in our previous work, the V1 cells are modeled through a set of log-Gabor filters in the Fourier domain [1]. We explore a model for V1 cells that includes a linear stage (log-Gabor filters) and a multiplication stage that is followed by an additional nonlinear output (exponent). Then, the suppressive effect of the surround in V1 is computed using end-stopped units which are subtypes of the simple and complex types. Afterwards, the integration of the inputs from V1 is realized by MT cells through excitatory and suppressive components calculated as a divisive normalization. The mechanisms, algorithmic properties and implementations that induce direction selectivity will be characterized in detail.

From a computational perspective, we propose mathematical models that relate the findings from neuroscience to the processing of visual information. We focus on linking the representation of the algorithm structure to the computations performed by neurons in the brain. It might be possible that future realizations of these ideas provide an important source of inspiration for developing cogni-

***Corresponding Author: Fernanda da C. e C. Faria:** Institute of Systems and Robotics, University of Coimbra, Portugal, E-mail: cunhaecastro@gmail.com
Jorge Batista: Institute of Systems and Robotics, University of Coimbra, Portugal, E-mail: batista@isr.uc.pt
Helder Araújo: Institute of Systems and Robotics, University of Coimbra, Portugal, E-mail: helder@isr.uc.pt

tive abilities on humanoid robots [2] towards human-robot interaction, and others fields such as cybernetics, computer vision, computational neuroscience, etc.

In this work, we will describe the main theoretical considerations (second section) of how to represent a model of movement computation based on V1 and MT cells. In the third section of this study, a biologically inspired processing scheme to compute visual motion is explored. Then, the results (fourth section) are presented and compared with simulations without the proposed improvements and some state-of-the-art bio-inspired approaches. Finally, we discuss detailed comparisons between the tests and we finish with conclusions in the last section.

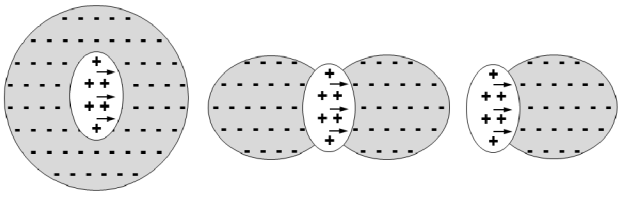

Figure 1: A coarse approximation of possible RFs examples that address center-surround antagonism. MT receptive field models consisting of concentric excitatory center (light shading and "+" signs) and inhibitory surround regions (dark shading and "-" signs). Arrows indicate the cell's preferred direction of motion. Three examples of different kinds of surround organization are presented. First, the RF has inhibitory regions that completely surround the excitatory area (left). Second, the RF has bilateral asymmetric inhibitory regions (middle). Third, the RF has a unilateral asymmetric inhibitory region (right).

2 A bio-inspired perception of motion

Motion perception is a basic quality of vision process that examines a sequence of different but related images over a brief span of time. The cerebral cortex of primates can be divided in anatomically and functionally distinct areas so-called cortical areas. This section addresses the study of perception of visual motion biologically pertinent with the small visual area known as Middle Temporal Area (MT or V5), whose neurons are particularly responsive to motion perception.

Motion computations and stereoscopic depth perception are closely related. The computation of motion direction begins in Primary Visual Cortex (V1), because, cortical feedforward inputs to MT come from several areas, including V1 and some other areas, such as Second Visual Area (V2). Inputs from V1 are the most potent [3].

MT neurons are selective for stimulus orientation, speed, and direction of movement. The majority of MT neurons are selective for the direction of motion stimulus. V1 neurons that project to MT are directionally selective, i.e., the cells respond strongly to motion in one (preferred) direction and very little or not at all to motion in the opposite or null direction. Basically, MT analysis may contribute to an important role in the processing of visual information related to motion perception [4, 5].

Disparity selectivity is already established for horizontal stimulus disparity in the V1 and V2 cells that project to MT neurons [6]. The large receptive fields (RFs) in MT are also sensitive to non-horizontal disparity, i.e., such cells are not specialized for detecting horizontal disparities, they are selective for both vertical and horizontal stimulus disparities [6].

Middle temporal area RFs are much more complex than previously described in [1], because they use a broader range of information. The properties of neuronal RFs in cortical area MT can roughly be divided in two main parts: its center region, so-called Classical Receptive Field (CRF), and its surround region (Fig. 1). MT receptive fields have interesting properties such as: i) surround area much larger than CRF; ii) center-surround RF relationships with antagonistic zones; iii) RF organized in a circular symmetric and asymmetric representation of the surround areas; and iv) many elongated RF along the preferred axis of motion [7–11].

Receptive fields of MT cells are complex structures, and they cannot be accurately modeled on the basis of linear mechanisms because direction selectivity is inherently nonlinear. Therefore, nonlinear mechanisms are essential to the generation of direction selectivity [9, 12, 13].

There have been many studies of motion detection mechanisms to account for direction selectivity [14–18]. They are inspired on many animals, including humans. These models carry the idea of three requirements which any motion detection algorithm has to compute. The first particular aspect of motion computation is the input stimulus, since we need to analyse at least two spatially separate input lines because motion detection requires comparing signals from two neighboring points in space. Second, the temporal filtering of the two input lines has to be asymmetrical in some way, otherwise motion detection mechanisms could be interchanged without affecting the output. Third, since not all directional computations in visual cortical neurons can be explained by linear interactions alone, the two input lines must be combined by a non-linear mechanism [19–22].

One of the first models of movement computation was proposed by Hassenstein and Reichardt [23], the Re-

ichardt detector (also called Hassenstein-Reichardt detector). Elaborated versions of the Reichardt detectors, i.e., the basic Reichardt model supplemented by spatial receptive fields, can explain psychophysical responses to motion in humans. Santen and Sperling [14] analyze the relationship between the Elaborated Reichardt Detector (ERD) and the motion energy detector of Adelson and Bergen [16] and Watson and Ahumada detector [17]. Adelson and Bergen detector is a reformulation of ERD with a multiplicative constant. They combine the filtered inputs with $\pi/2$ spatial phase shift and $\pi/2$ temporal phase delay. The output signal of the full opponent form of the ERD is formally identical to the spatiotemporal energy model. And, an elaborated Watson and Ahumada detector that includes squaring followed by temporal integration and subtraction of outputs of two mirror-image subunit is fully equivalent to an ERD with $\pi/2$ for both, temporal delay and spatial phase-shift filters. Both detectors [16, 17], despite having different internal structures, are a more complicated way used to construct methods of the direction selectivity computation than the standard formulation of the ERD.

Neurons interactions taking place in the nervous system have been inspiring algorithmic models. Those models are described by formal operations (convolutions, subtractions, multiplications etc.) to implement motion detection. Many studies of mechanisms using biologically inspired motion models adapt some of the existent motion detector [16, 17, 23] to analyse the direction-sensitive computation that are applied for several activities (action recognition, motion transparency, contour detection etc.). Some examples of these kind of computer vision applications for a variety of real world image sequences and synthetic image sequences can be found in [24–28].

Our work explores an alternative approach that computes motion based on the not yet fully understood information, about how the biological motion detection mechanisms realize neurons synaptic interactions, and how they can perform a multiplication. We suggest a model (Fig. 2) for analysis of the nonlinear mathematical operation, the exponent, in order to improve the response of cells. We are extending the existing elaborated Reichardt detector (ERD) [14, 15], simply adding an output exponent before the last stage, where the output of a leftward selective motion detector is subtracted from that of a rightward detector. This particular adaptation in motion processing can substantially alter neuronal selectivity. We incorporate these nonlinear component features to produce more accurate direction-selective predictions [12, 21, 29–31].

After the modified ERD, later responses take into account the larger stimulus context, and we compute the final response through two stages (Fig. 2). One, a re-entry stage of modulating feedback (FB) from MT followed by a nonlinear suppressive component is calculated, giving

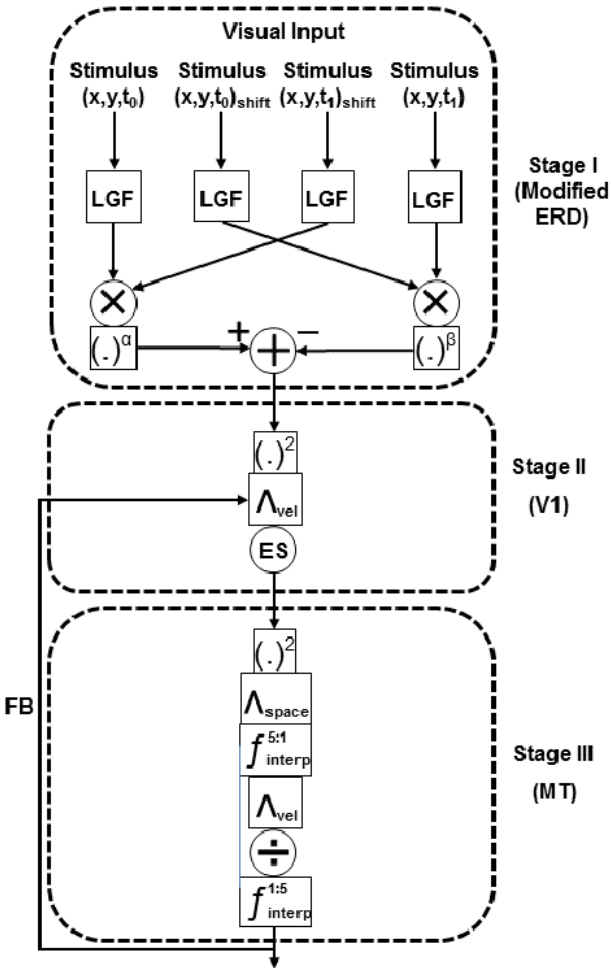

Figure 2: A model of a modified elaborated Reichardt detector (ERD) with end-stopped units (stage II) and a normalization (stage III). The model consists of a modified ERD where the input stimulus with and without a spatial shift (horizontal and vertical) and a temporally delayed (time t_0 to t_1) are filtered by log-Gabor functions (LGF). The symbol × represents a multiplication operation of the filtered images. Then the exponentiation $((.)^{\alpha}$ and $(.)^{\beta})$ of the correlation outputs are calculated before the subtraction step (+ represents summation with opposite signs). Stage II starts with a squaring operation followed by a velocity blur (Λ_{vel}) realized by Gaussian filters for motion speeds and directions. After, a signal enhanced by FB (MT feedback) is calculated. End-stopped units (ES) are modelled as a divisive operation of the response of the center cell by the surround suppression response. Stage III starts with a squaring operation followed by a smoothing with a Gaussian function (Λ_{space}) and the computation of the spatial size RFs difference between V1 and MT ($f^{5:1}_{interp}$ is a spatial interpolation for resolution reduction). Then, again, a velocity blur (Λ_{vel}) is realized. The symbol ÷ represents a normalization process. And, the original image size is reestablished ($f^{1:5}_{interp}$).

rise to their large RFs. This suppression has been found to be largely nonspecific, and, it is modeled as end-stopped units, which are subtypes of the simple and complex types. Neurons with this property (end-stopped) are found in the V1 layers and provide input to MT cells. A substantial portion of the input to MT cells is end-stopped. The other, a stage where cells are explained by standard normalization in order to connect the various results on motion integration in MT. In both stages, RF of cells are surrounded by suppressive regions, and the influence from the surround should be modeled as a divisive or subtractive suppression. Mathematically in our model, the interaction of center and surround mechanisms (suppression) is divisive. This type of interaction generated outside of the CRF serves to improve motion detection results [32–36].

In this work, we describe how to detect motion using a combination of a modified ERD, end-stopped units and normalization. All those methods are bio-inspired processing schemes and we evaluate our algorithm in synthetic and real-world image sequences.

3 Modeling optical motion estimation

To advance the understanding of how neurons of visual area V1 and MT contribute to perception of motion, we developed a hierarchical modeling framework. Fig. 2 shows the basic computational structure of the proposed motion detection algorithm. Optic flow processing is computed in a series of steps and our architecture starts with a modified ERDs which has two input stimulus (images) with a simple temporal delay ($I(x, y, t_0)$ and $I(x, y, t_1)$), i.e., the 2D image intensity coordinates (x, y) with a time delay (t_0 and t_1). Thus, for estimating spatio-temporal correlations we use a spatial shift (horizontal and vertical) of the two input stimulus defined by:

$$I(x, y, t_0)^{\Delta_{l,n}} = I(x - \Delta x_{l,n}, y - \Delta y_{l,n}, t_0), \tag{1}$$

$$I(x, y, t_1)^{\Delta_{l,n}} = I(x - \Delta x_{l,n}, y - \Delta y_{l,n}, t_1), \tag{2}$$

where $\Delta x_{l,n} = s_l \times cos(\delta_n)$, $\Delta y_{l,n} = s_l \times sin(\delta_n)$, δ_n is the direction angle, s_l represents speeds, $n = 1 \ldots N_\delta$ indexes motion directions, N_δ is the number of directions, $l = 1 \ldots N_s$ indexes motion speeds and N_s is the number of speeds.

Consistent with neurophysiological data, many neurons in the visual cortex have tuned direction response profiles. As in our previous work [1], each V1 RF that corresponds to direction selective neurons is modelled through linear band-pass filters, i.e., a set of two-dimensional log-Gabor functions in the Fourier domain (the temporal dimension is not Fourier transformed), as follows:

$$g(r, \phi; \theta_m, f_h) = \exp\left[\frac{-(\log(r/f_h))^2}{2(\log(k/f_h))^2}\right] \times \exp\left[-\frac{(\phi - \theta_m)^2}{2\sigma^2}\right], \tag{3}$$

where $(\phi - \theta_m)$ is the absolute angular distance, i.e., $(\phi - \theta_m) = |\tan^{-1}(\sin(\theta_m - \phi)/\cos(\theta_m - \phi))|$ where the coordinate system is rotated according to the orientation angle θ_m, m corresponds to motion orientations index and $\phi = \tan^{-1}(-y/x)$ is the anticlockwise angular coordinate. The standard deviation σ of the Gaussian function in the angular direction has a constant value of $\pi/9$ (this value is based on empirical data). The central radial frequency is f_h, h corresponds to spatial frequencies index, k is the standard deviation used to determine the bandwidth of the filter in the radial direction and r is the radial coordinate ($r = \sqrt{x^2 + y^2}$). The term k/f_h has a fixed value of 0.55 to achieve constant shape ratio filters [37, 38].

The Fourier transform of the input images is denoted by the hat-symbol and the change of arguments to angular frequencies (the temporal dimension is not Fourier transformed), i.e, $I(x, y, t_0)$ is Fourier transformed into $\hat{I}(\omega_x, \omega_y, t_0)$. Here, each input stimulus is filtered by log-Gabor functions and the computed signal is then given as follows (the four separable combinations outputs):

$$A(x, y, t_0; \theta_m, f_h) = F^{-1}\{\hat{I}(\omega_x, \omega_y, t_0) \times g(r, \phi; \theta_m, f_h)\} \tag{4}$$

$$A'(x, y, t_0; \theta_m, f_h)^{\Delta_{l,n}} = F^{-1}\{\hat{I}(\omega_x, \omega_y, t_0)^{\Delta_{l,n}} \times g(r, \phi; \theta_m, f_h)\} \tag{5}$$

$$B(x, y, t_1; \theta_m, f_h) = F^{-1}\{\hat{I}(\omega_x, \omega_y, t_1) \times g(r, \phi; \theta_m, f_h)\} \tag{6}$$

$$B'(x, y, t_1; \theta_m, f_h)^{\Delta_{l,n}} = F^{-1}\{\hat{I}(\omega_x, \omega_y, t_1)^{\Delta_{l,n}} \times g(r, \phi; \theta_m, f_h)\} \tag{7}$$

where the symbol F^{-1} corresponds to the inverse Fourier transform.

The filtered stimulus output, in spatial domain is calculated by inverse Fourier transform from time t_0 (4) is correlated (8) with the direct input from a time delay (t_1) with spatial shift (7). This can be modelled by a local image phase ($\psi^A_{m,h} = arg\{A(x, y; \theta_m, f_h)\}$ and $\psi^{B',\Delta_{l,n}}_{m,h} =$

$arg\{B^{'}(x,y;\theta_m,f_h)^{\Delta_{l,n}}\}$). We consider the variation of the temporal order (9) where (5) is correlated with (6) employing local image phase ($\psi_{m,h}^{A^{'},\Delta_{l,n}} = arg\{A^{'}(x,y;\theta_m,f_h)^{\Delta_{l,n}}\}$ and $\psi_{m,h}^{B} = arg\{B(x,y;\theta_m,f_h)\}$). Thus, we defined correlation equations as follows:

$$C_{l,n}^{AB^{'}} = \frac{1}{N_\theta}\sum_{m=1}^{N_\theta}[cos(\psi_{m,h}^{B^{'},\Delta_{l,n}} - \psi_{m,h}^{A})]^{+} \tag{8}$$

$$C_{l,n}^{BA^{'}} = \frac{1}{N_\theta}\sum_{m=1}^{N_\theta}[cos(\psi_{m,h}^{B} - \psi_{m,h}^{A^{'},\Delta_{l,n}})]^{+} \tag{9}$$

where N_θ is the number of orientations, $[.]^{+} = max(.,0)$ denoting a positive half-wave rectification, because cells cannot fire negative values. Afterwards, correlation results of different orientations are summed.

The basic idea behind the equations (8) and (9) to account for the responses of complex cells in V1 involves operations with complex numbers, i.e., sum of the responses (simple cells) corresponding to the real (Re) and imaginary (Im) parts of a complex-valued log-Gabor filters (orthogonal in phase). We consider the real and imaginary parts of each of 4 responses (4)-(7), therefore, we have 8 measurements available on the complex plane. Under this conditions, if we rewrite the cosine of the angular difference between image phases (quadrature) (8) and (9), we see that:

$$\begin{aligned} C_{l,n}^{AB^{'}} = 1/N_\theta \sum_{m=1}^{N_\theta}\{&Re[A(x,y,t_0;\theta_m,f_h)] \\ &\times Re[B^{'}(x,y,t_1;\theta_m,f_h)^{\Delta_{l,n}}] \\ &+ Im[A(x,y,t_0;\theta_m,f_h)] \\ &\times Im[B^{'}(x,y,t_1;\theta_m,f_h)^{\Delta_{l,n}}] \\ &/|A(x,y,t_0;\theta_m,f_h)| \\ &\times |B^{'}(x,y,t_1;\theta_m,f_h)^{\Delta_{l,n}}|\}^{+}, \end{aligned} \tag{10}$$

$$\begin{aligned} C_{l,n}^{BA^{'}} = 1/N_\theta \sum_{m=1}^{N_\theta}\{&Re[A^{'}(x,y,t_0;\theta_m,f_h)^{\Delta_{l,n}}] \\ &\times Re[B(x,y,t_1;\theta_m,f_h)] \\ &+ Im[A^{'}(x,y,t_0;\theta_m,f_h)^{\Delta_{l,n}}] \\ &\times Im[B(x,y,t_1;\theta_m,f_h)] \\ &/|A^{'}(x,y,t_0;\theta_m,f_h)^{\Delta_{l,n}}| \\ &\times |B(x,y,t_1;\theta_m,f_h)|\}^{+}, \end{aligned} \tag{11}$$

where × and / denote a multiplication and division operation for each matrix element. This formulation is consistent with the assumptions of [39, 40].

The modified ERDs model is completed by incorporating expansive static output nonlinearities (α and β), before the subtraction operation in the correlation asymmetric outputs as follows:

$$C(x,y,\delta_n,s_l) = [(C_{l,n}^{AB^{'}})^{\alpha} - (C_{l,n}^{BA^{'}})^{\beta}]^{+}. \tag{12}$$

In the next level (stage II – Fig. 2), corresponding yet to cortical area V1, a cascade architecture based on three steps involving spatial integration, signal enhancement, feedback (FB) signals re-enter and normalization (center-surround shunting inhibition) is used.

The integration of motion signals is a gradual process and the basic idea consists of nonlinear ordinary differential equations. Those equations consider membrane properties and normalizing gain control. The electrical properties of nerve cells can be characterized like electrical circuits. The cell membrane is modeled as a resistance and a capacitance, i.e., an RC circuit. Cortical cells behave like an RC circuit and the input to a cell is a current driven by the synaptic conductances that vary over time [41–43]. According to this model, the dynamics of the membrane circuit obeys the following equation:

$$\tau\dot{x} = -D \times x + (x - E_{ex}) \times g_{ex} - (x - E_{in}) \times g_{in}, \tag{13}$$

assuming a zero-level resting state with constant leak conductance $g_{leak} = D$; τ denotes the membrane constant and x is the cell's membrane potential. The terms g_{ex} and g_{in} represent the total inputs from excitatory and inhibitory neurons synapses on the cell, respectively. Parameters E_{ex} and E_{in} define reversal potentials for excitatory and inhibitory conductances, respectively.

Neuronal models implementation integrates these mechanisms in a generic three successive computational steps (14)-(16), which are variants of the membrane (13) [44–48]:

$$\dot{x}^{(1)} = -x^{(1)} + (x^{FF})^2 \star \Lambda_{vel}, \tag{14}$$

$$\dot{x}^{(2)} = -x^{(2)} + x^{(1)} \times (1 + F \times x^{FB}), \tag{15}$$

$$\dot{x}^{(3)} = -x^{(3)} + x^{(2)} - (G + H \times x^{(3)}) \times \sum x^{(2)}, \tag{16}$$

where the terms $x^{(1)}$, $x^{(2)}$ and $x^{(3)}$ denote the activity within the three stages of the particular model area. The parameter x^{FF} represents the driving input signal (forward stream). Λ_{vel} denotes weighting kernels for filtering operations (velocity blur). The term x^{FB} is the modulatory feedback signal and the constant F adjusts the strength of feedback. The constants G and H control the strength of normalization.

The mechanisms that underlie V1 neural dynamics implementation of feature extraction, exhibit end-stopped

units to some degree. We adapt several indications in the literature about the properties of end-stopped units to estimate motion [32, 35, 36]. We follow the mechanisms of neural three step processing of (14)-(16). It starts with a nonlinear signal enhancement, i.e., it is squared in order to sharpen the distribution. Then a velocity blur is processed by different Gaussian kernels (Λ_{vel}) for motion speeds and for motion directions [39, 44, 47, 49]:

$$\dot{x}_{v1}^{(1)} = -x_{v1}^{(1)} + (x_{v1}^{FF})^2 \star \Lambda_{vel}, \tag{17}$$

where $x_{v1}^{FF} = C(x, y, \delta_n, s_l)$, $x_{v1}^{(1)}$ ranges between zero and one, and the blur is calculated by a convolution (⋆) in the domain of velocities with a Gaussian defined by the standard deviations $\sigma_s = 0.2$ pixel per frame for motion speeds and $\sigma_\delta = 0.75$ for motion directions. Speeds are encoded on a logarithmic scale.

Motion signals integration provides important cues to the processes that are used by the visual system. Recurrent connectivity is a mechanism that is known to exist in cortical areas and the surround suppression in V1 is affected by feedback from MT. Feedback connections for the treatment of visual information can enhance the coherent inputs consistent with the motion direction. This helps the process of information extraction to other cortical visual areas. MT plays an important role in the integration of motion signals in area V1 [44, 47, 50–53]. Our model includes the reciprocal connections of V1 and MT. The feedback (FB) in V1 is computed as follows:

$$\dot{x}_{v1}^{(2)} = -x_{v1}^{(2)} + x_{v1}^{(1)} \times (1 + F \times x_{v1}^{FB}), \tag{18}$$

where F has a fixed value of 100 to amplify the FB signal (x_{v1}^{FB}) and $x_{v1}^{(2)}$ ranges between zero and one.

End-stopped model performs a suitable front end for our MT model. We examine the end-stopped properties without considering that their neurons form a separated class in V1, because many V1 neurons exhibit some end-stopped characteristics, i.e., inhibitory influences from the surround. Here we show another version of the end-stopped units, which consists of extending the analysis of the non-classical RF organization to the output of the initial motion detector. Thus, the response of V1 cells could be characterized as a divisive normalization (shunting inhibition) revealed by analysis of the center-surround interaction [32, 35, 36, 39]. After the FB step, based on the steady-state solution of the (16), the encoding of motion information using mechanisms that are known to exist in V1 can be approximated as follows:

$$x_{v1}^{(3)} = \frac{r_{center}}{\epsilon + r_{center} + k r_{surround}}, \tag{19}$$

where r_{center} is the response of the center cell and whereby it is approximated by filtering with a Dirac pulse ($r_{center} = x_{v1}^{(2)}$ – no kernel is applied). The $r_{surround}$ is the total surround response calculated by convolving the signal $x_{v1}^{(2)}$ with a Gaussian filter with parameter values of $\sigma_s = 0.5$ pixel per frame for motion speeds and $\sigma_\delta = 2.0$ for motion directions. The model is based on the assumption that the RF centers and surrounds overlap spatially. Therefore, the surround interactions overlaps the center and inhibits it. The constant k is a gain parameter (set to 5) and ϵ has a fixed value of 1.

Motion is a rich source of various types of biologically relevant signals. Activity in V1 influences the perceptual interpretation of the directional information and MT inherits considerable directional information from V1 area. We propose that the next level (stage III – Fig. 2) is performed by cortical area MT. We follow the three-step processing of the (14)-(16) again. First, the signal from V1 is squared for a nonlinear signal enhancement. Next a spatial integration (Λ_{space}), where the signal is convolved by a Gaussian kernel, and subsequently sampled (for a different spatial resolution) using a linear interpolation ($f_{interp}^{5:1}$) . This combination of poor spatial resolution (less ambiguous signals) is consistent with the visual inputs to MT and their larger RF size compared to that in V1 [5, 44, 47]. Then, as in stage II, motion signals are smoothed in the velocity domain (Λ_{vel}) with the same parameters of V1 area. We then calculated the integration as follows:

$$\dot{x}_{MT}^{(1)} = -x_{MT}^{(1)} + f_{interp}^{5:1}((x_{MT}^{FF})^2 \star \Lambda_{space}) \star \Lambda_{vel}, \tag{20}$$

where $x_{MT}^{FF} = x_{v1}^{(3)}$ and $x_{MT}^{(1)}$ ranges between zero and one. The spatial resolution is reduced by a factor of 5 (linear interpolation) and the standard deviation $\sigma_{space} = 5$ for Λ_{space}.

The feedback in MT area could be included if our model had influences of other cortical areas, such as the medial superior temporal area (MST). However, in the current model that involves connections of V1 and MT, the FB in MT is computed as follows:

$$\dot{x}_{MT}^{(2)} = -x_{MT}^{(2)} + x_{v1}^{(1)} \times (1 + F \times x_{MT}^{FB}), \tag{21}$$

where the FB signal x_{MT}^{FB} does not receive any input and $x_{MT}^{(2)}$ ranges between zero and one. Thus, the result response is denoted by the indetity $x_{MT}^{(2)} = x_{MT}^{(1)}$ in its steady state.

As demonstrated in Fig. 2, detailed picture of the activities, visual system mechanisms are represented as a sequential elaboration of processing steps through a hierarchy of visual areas. MT cells play an important role in integrating and interpreting inputs from earlier-level

end-stopped cells (V1). To analyze visual motion, MT neurons opponency is an essential aspect. We consider that suppression contributes to the speed and direction tuning of MT neurons. We fit a divisive suppression by applying a center-surround interaction, whereby the MT neuron is constructed by excitatory and suppressive components which shows the normalization property. More importantly, cells model generates nonspecific suppression which is divisive. The Center and surround differ in size and the excitatory component is defined as the outputs of each of the $x_{MT}^{(2)}$, and the surround calculates the sum over all velocities [33, 34, 54]:

$$\dot{x}_{MT}^{(3)} = \frac{x_{MT}^{(2)}}{\sigma_{MT}^2 + \sum x_{MT}^{(2)}}, \tag{22}$$

where $\sigma_{MT} = 0.01$. Equation (22) is based on the steady-state solution of the (16).

Finally, as demonstrated in Fig. 2, a linear interpolation ($f_{interp}^{1:5}$) is computed to reconstruct the original image size.

The estimation of velocity is based on similar spatial RFs with a range of directions (δ_n) and speeds (s_l), i.e., a set of hypotheses at each location. Those sets of possibilities can be interpreted as local neural population code which is biologically plausible with cells of MT. To find the best fit, we can pool local motion information composed of a population of units, using a divisive normalization to determining the most active neurons. Therefore, to interpret the population of motion for a given spatial location, we simply normalize to present a single activation signal [55–57]:

$$R_{norm} = \frac{\sum M_{vel} \times x_{MT}^{(3)}}{\sum x_{MT}^{(3)}}, \tag{23}$$

where M_{vel} represents the velocities of cells that are tuned ($+\Delta x_{l,n}$ and $-\Delta y_{l,n}$).

The specific parameters used to estimate motion for real and synthetic image sequences are given in the next section according to the tests realized.

4 Experimental results

The technique described for optic flow has been implemented and applied to a variety of real image sequences and synthetic image sequences. The experimental evaluation considers greyscale synthetic and real-world images from the Middlebury flow repository (vision.middlebury.edu/flow/) [58, 59] and from [60] as input. We consider real world images, scenes captured with a camera, composed of real objects and natural texture. The synthetic images are artificial sequences generated using computer graphics and some use a real image to create surface texture.

We use an angular measure of error for quantitative comparisons and we also compute the absolute error. Quantitative computational theories can be an effective tool for summarizing existing data and help us for testing the consistency of our computational model. For evaluation (see Appendix A) we used the average angular error (AE) and the absolute flow endpoint error (EE)[58–62].

We have used almost all the same parameter values of [47]. For log-Gabor filters, the orientations are $\theta_m = m\pi/N_\theta$, where $N_\theta = 8$ is the number of orientations ($m = 1 \dots N_\theta$). The spatial frequencies are defined for $f_n = 1/(2\pi/(s_w s_1 s_b^{N_f - h}))$, where $s_w = 3.0$ relates an angular frequency to a speed, $s_1 = 0.8$ is the minimal speed, s_b is a factor that influences the overlap between filters ($h = 1 \dots N_f$, $N_f = N_s - 1$, $s_b = 1.5$). The number of speeds is $N_s = 7$ for image sequences with low speeds (Army, Mequon, Schefflera, Wooden, Yosemite, Rubber-Whale, Hydrangea, Dimetrodon, Grove2, Grove3, Translating tree and Diverging tree). For the image sequences with high speeds $N_s = 9$ (Grove, Urban, Teddy, Urban3) and $N_s = 10$ (Urban2). Speeds are multiples of minimal speed defined for $s_l = (s_w s_1 s_b^{l-1})/s_w$, where $l = 1 \dots N_s$. The directions are $\delta_n = n2\pi/N_\delta$ where the number of directions is $N_\delta = 16$ ($n = 1 \dots N_\delta$). Moreover, the first speed used is zero as well as the first frequency. Motion has been computed with 5 frames. For Middlebury flow repository we use frames 7 to 11 (except when the sequence provides only two images) and for [60] we use frames 17 to 21.

The Gaussian kernels for motion directions used for velocity blur and for surround interaction are computed in a region of 3x1 and 9x1 pixels, respectively. The Gaussian kernels for motion speeds used for velocity blur and for surround interaction are encoded on a logarithmic scale and they are computed in a region of 1x5 for both filters. The Gaussian kernel for spatial integration (Λ_{space}) is computed in a region of 21x21 pixels. For Gaussian filtering, we use two boundary conditions: a circular boundary condition for the domain of motion directions and a replicate boundary condition for the domain of motion speed and space.

Tables 1 and Table 2 present the performance of our algorithm for a collection of optical flow datasets with hidden (Dataset 1) and public (Dataset 2) ground truth flow, respectively. In our experiments, we compare two alternative of configurations besides the original configuration for the same input sequences (Table 2). First, to examine the importance of the exponents (α and β – Fig. 2), we execute

Table 1: Error statistics for motion perception estimation for a variety of synthetic and real-world image sequences (Dataset 1).

	Our Method	
	AE	*EE*
Army	8.97	0.24
Mequon	8.89	0.88
Schefflera	10.9	1.08
Wooden	8.30	0.78
Grove	8.16	2.0
Urban	22.9	3.09
Teddy	19.9	3.81

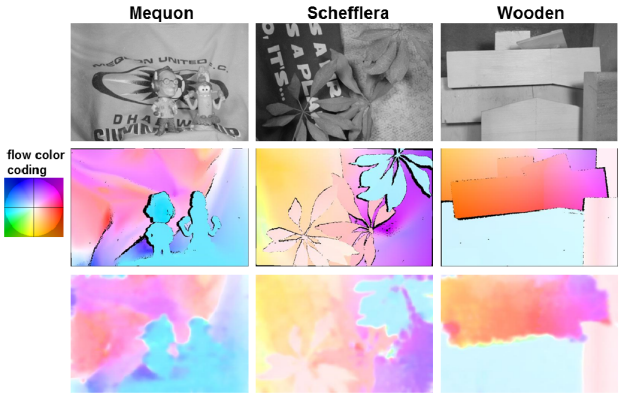

Figure 3: The top row shows the first frame of the input sequences used to our study. The middle row shows the color coded ground truth flow of each image. The bottom row shows the estimated optic flow of our algorithm. In the left-most column we visualize the color coded ground truth flow that indicate speed and direction.

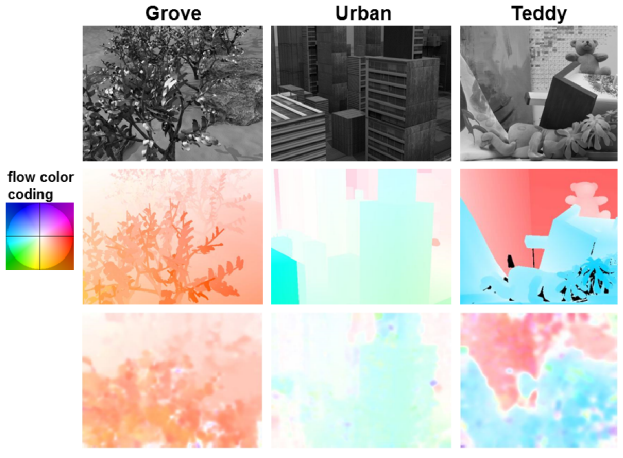

Figure 4: The top row shows the first frame of the input sequences used to our study. The middle row shows the color coded ground truth flow of each image. The bottom row shows the estimated optic flow of our algorithm. In the left-most column we visualize the color coded ground truth flow that indicate speed and direction.

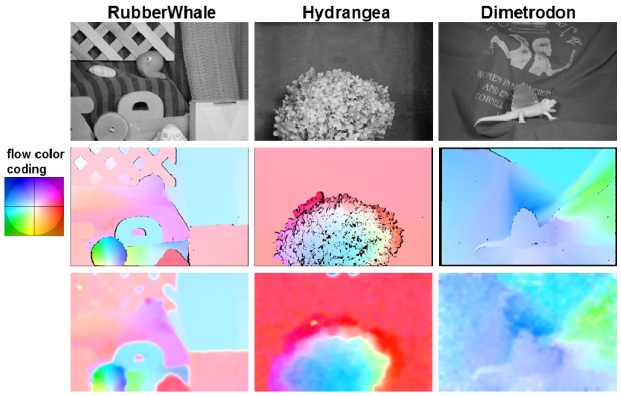

Figure 5: The top row shows the first frame of the input sequences used to our study. The middle row shows the color coded ground truth flow of each image. The bottom row shows the estimated optic flow of our algorithm. In the left-most column we visualize the color coded ground truth flow that indicate speed and direction.

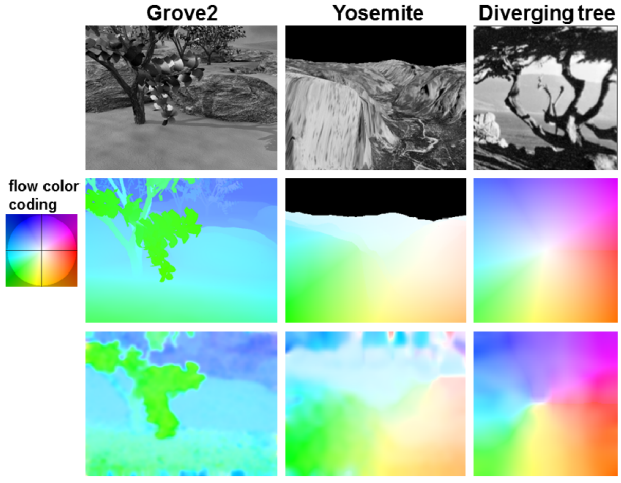

Figure 6: The top row shows the first frame of the input sequences used to our study. The middle row shows the color coded ground truth flow of each image. The bottom row shows the estimated optic flow of our algorithm. In the left-most column we visualize the color coded ground truth flow that indicate speed and direction.

the algorithm without them. Then, to evaluate the influence of feedback from MT, we execute the algorithm without FB.

Fig. 3 and Fig. 4 show estimated optic flow for three real image sequences (Mequon, Schefflera and Wooden), two synthetic image sequences (Grove and Urban) and one modified stereo data (Teddy). In the left-most column we see the flow color coding, i.e., the color coded ground truth flow that indicate speed and direction (Middlebury color code). It associates a single color with each possible velocity and the direction of the velocity corresponds to the hue of the velocity (whiter for slower speeds). We included in Fig. 3 and Fig.4 images of the results presented in Table 1.

Table 2: Error statistics for motion perception estimation for a variety of synthetic and real-world image sequences (Dataset 2).

	Our Method		**Our Method without the exponents α and β**		**Our Method without the FB from MT**	
	AE	*EE*	*AE*	*EE*	*AE*	*EE*
RubberWhale	6.59	0.22	7.87	0.26	6.91	0.23
Hydrangea	5.36	0.55	8.75	1.18	5.67	0.71
Dimetrodon	4.31	0.27	11.46	0.71	4.31	0.27
Grove2	5.98	0.50	9.50	0.88	9.14	0.75
Grove3	11.84	1.37	15.76	1.98	14.68	1.70
Urban2	22.61	3.11	35.05	6.51	29.38	5.18
Urban3	18.91	3.09	29.38	5.19	28.91	4.72
Venus	12.96	1.25	21.47	2.21	12.96	1.25
Yosemite	5.43	0.28	8.71	0.46	7.34	0.41
Translating tree	1.33	0.09	3.75	0.23	1.81	0.11
Diverging tree	5.58	0.16	8.00	0.24	5.31	0.15

Table 3: Comparisons of the average angular errors of our model and some state-of-the-art bio-inspired works in the literature.

	Our Method	**Raudies 2012[a]**	**Solari 2014[b]**	**Dellen 2011[c]**	**Solari 2015[d]**
RubberWhale	6.59	-	17.1	9.8	10.20
Hydrangea	5.36	-	-	9.3	5.96
Dimetrodon	4.31	-	-	-	-
Army	8.97	14.5	-	-	12.0
Mequon	8.89	11.4	-	-	10.7
Schefflera	10.9	16.8	-	-	15.6
Wooden	8.30	11.7	-	-	16.6
Grove	8.16	16.6	-	-	6.51
Grove2	5.98	-	-	-	4.28
Grove3	11.84	-	-	-	9.72
Urban	22.9	27.3	-	-	16.2
Urban2	22.61	-	-	-	14.51
Urban3	18.91	-	-	-	15.11
Yosemite	5.43	9.77	11.4	3.75	3.41
Translating tree	1.33	-	-	0.52	-
Diverging tree	5.58	-	6.7	3.82	-
Venus	12.96	-	-	-	-
Teddy	19.9	37.9	-	-	12.3

[a] [47]
[b] [63]
[c] [64]
[d] [65]

Fig. 5 and Fig. 6 show estimated optic flow for three real image sequences (RubberWhale, Hydrangea and Dimetrodon) and for three synthetic image sequences (Grove2, Yosemite and Diverging tree), respectively. In the left-most column we see the flow color coding (Middlebury color code). We included in Fig. 5 and Fig. 6 images of the results presented in Table 2.

The performance of the algorithm is comparable to the [47, 63–65] obtained by some state-of-the-art bio-inspired approaches. Various approaches exist in the literature that test only the widely used simple synthetic Yosemite sequence with ground true flow, which is very limited. Our study focuses on comparing the results with works that present several tests. According to Table 3 the proposed algorithm has better results than the estimated optic flow results in [47, 63].

Except Teddy and Venus, that are stereo images pairs, our proposed algorithm has better average angular error results for real world images (the first seven images of the Table 3) than for synthetic images, if we compare our results with the results presented in [64, 65].

Summarizing the results, the simulations suggest that the additional nonlinear components (exponents) and the feedback projections of MT cells to V1 produce more accurate motion detection results.

5 Discussion

Our model is inspired by the idea of our previous work [1], where we represented RFs of V1 cells through logarithmic Gabor functions. Although our previous approach [1] uses the energy model [16] to describe the structure of the complex cells, we choose the ERD [14, 23] to extract the motion information from input image sequences in this work. We opt for ERD to compute movement because the energy model [16] is a more complicated reformulation of the ERD. This biologically plausible model was easier to configure and it permitted to develop an algorithm in MATLAB with a reasonable processing time. We spent 165 seconds to obtain the estimated optic flow of Yosemite sequence (Fig. 6).

The model for V1 neuron is a relatively simple extension of the ERD whose definitions use an exponent in their formulation. Table 2 demonstrates the results of our algorithm without the exponents (α and β). If we compare with the results of the proposed algorithm (Table 2), we can observe that the nonlinear interaction incorporated to the ERD architecture contributes to improve the output data.

To examine the importance of the FB connections from MT to V1 we test our algorithm without the FB (Table 2). Our results show that FB influences more the relative performance of synthetic sequences than our real world images.

We proposed a different implementation of the inhibitory center-surround interactions in the V1 area, based on some evidence that end-stopped units can module surround suppression [32, 33, 35]. This interpretation is appealing because nonspecific suppression can come from a combination of cells [33] and we modeled these properties that extend outside their CRF as a divisive normalization.

The algorithm performs well on real world images (Army, Mequon, Schefflera, Wooden, RubberWhale, Hydrangea and Dimetrodon), which contain motion discontinuities, rigid and non-rigid motion, hidden texture, thin structures, areas with little texture and shadows [59]. The best-fitting parameter values were chosen for computing the flow in real world images. On the other hand, our method has limitations with synthetic image sequences containing motions with large displacements. An aspect to be considered about the bio-inspired approaches presented in the Experimental Results section [64, 65] is that they have better results for synthetic images, however, their performance is poor for real images. That is a consequence of the over smoothy sequences observed on the results available. A challenge in biologically inspired models is to find good methods that could help to correctly find motion information across the various datatypes in the benchmark [59].

In summary, this work presents a bio-inspired model where a modified ERD with exponentiation, feedback from MT, and shunting inhibition (end-stopped units) affecting V1 neurons is proposed. The output of V1 feed the MT neurons, the spatial resolution is reduced, followed by a normalization and a subsequent pooling to find the most active neurons for motion detection. Our model is consistent with known properties of neurons in V1 and MT brain areas and we suggested an orderly arrangement and a combination of methods, which is different from existing state-of-the-art methods.

As future work, we are considering to extend the proposed model by adding a stage to represent the neurons of V2 in order to investigate whether the overall results can be improved. This investigation can answer whether the properties of V2 neurons are similar to MT neurons and whether they can combine the signal information following a very nonlinear integrative mechanism [3, 66].

References

[1] F. d. C. e. C. Faria, J. Batista, H. Araújo, Stereoscopic Depth Perception Using a Model Based on the Primary Visual Cortex, PLoS ONE, 2013, 8(12), e80745

[2] M. Antonelli, A. Gibaldi, F. Beuth, A. J. Duran, A. Canessa, M. Chessa, F. Solari, A. P. del Pobil, F. Hamker, E. Chinellato, S. P. Sabatini, A hierarchical system for a distributed representation of the peripersonal space of a humanoid robot, IEEE Transactions on Autonomous Mental Development, 2014, 6, 259–273

[3] J. H. Maunsell, D. C. van Essen, The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey, The Journal of Neuroscience, 1983, 3, 2563–2586

[4] J. H. R. Maunsell, D. C. van Essen, Functional Properties of Neurons in Middle Temporal Visual Area of the Macaque Monkey. I. Selectivity for Stimulus Direction, Speed, and Orientation, Journal of Neurophysiology, 1983, 49, 1127–1147

[5] J. A. Movshon, W. T. Newsome, Visual Response Properties of Striate Cortical Neurons Projecting to Area MT in Macaque Monkeys, The Journal of Neuroscience, 1996, 16, 7733–7741

[6] J. H. R. Maunsell, D. C. van Essen, Functional Properties of Neurons in Middle Temporal Visual Area of the Macaque Monkey. II. Binocular Interactions and Sensitivity to Binocular Disparity, Journal of Neurophysiology, 1983, 49, 1148–1167

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

[8] R. T. Born, Center-Surround Interactions in the Middle Temporal Visual Area of the Owl Monkey, Journal of Neurophysiology, 2000, 84, 2658–2669

[9] L. L. Lui, J. A. Bourne, M. G. P. Rosa, Spatial Summation, End Inhibition and Side Inhibition in the Middle Temporal Visual Area (MT), Journal of Neurophysiology, 2007, 97, 1135–1148

[10] G. C. DeAngelis, R. D. Freeman, I. Ohzawa, Length and Width Tuning of Neurons in the Cat's Primary Visual Cortex, Journal of Neurophysiology, 1994, 71, 347–374

[11] R. T. Born, D. C. Bradley, Structure and Function of Visual Area MT, Annual Review Neuroscience, 2005, 28, 157–189

[12] D. G. Albrecht, W. S. Geisler, Motion selectivity and the contrast-response function of simple cells in the visual cortex, Visual Neuroscience, 1991, 7, 531–546

[13] T. Poggio, W. Reichardt, Considerations on Models of Movement Detection, Kybernetik, 1973, 13, 223–227

[14] J. P. H. van Santen, G. Sperling, Elaborated Reichardt detectors, Journal of the Optical Society of America A, 1985, 2(2), 300–321

[15] W. Reichardt, Evaluation of optical motion information by movement detectors, Journal of Comparative Physiology A, 1987, 161, 533–547

[16] E. H. Adelson, J. R. Bergen, Spatiotemporal energy models for the perception of motion, Journal of the Optical Society of America A, 1985, 2(2), 284–299

[17] A. B. Watson, A. J. A. Jr., Model of human visual-motion sensing, Journal of the Optical Society of America A, 1985, 2(2), 322–342

[18] N. C. Rust, V. Mante, E. P. Simoncelli, J. A. Movshon, How MT cells analyze the motion of visual patterns, Nature Neuroscience, 2006, 9(11), 1421–1431

[19] M. Silies, D. M. Gohl, T. R. Clandinin, Motion-Detecting Circuits in Flies: Coming into View, Annual Review of Neuroscience, 2014, 37, 307–327

[20] A. Borst, M. Egelhaaf, Detection Visual Motion: Theory and Models, in F. A. Miles, J. Wallman (Ed.), Visual Motion and Its Role in the Stabilization of Gaze, Elsevier Science Publishers B.V., 1993

[21] B. Krekelberg, Motion Detection Mechanisms, in A. Basbaum et al (Ed.), The Senses: A Comprehensive Reference, Elsevier Inc, Oxford, 2008

[22] A. Borst, T. Euler, Seeing Things in Motion: Models, Circuits, and Mechanisms, Neuron, 2011, 71, 974–994

[23] B. Hassenstein, W. Reichardt, Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers Chlorophanus., Z Naturforsch, 1956, 11b, 513–524

[24] N. Petkov, E. Subramanian, Motion detection, noise reduction, texture suppression and contour enhancement by spatiotemporal Gabor filters with surround inhibition, Biological Cybernetics, 2007, 97, 423–439

[25] F. Raudies, H. Neumann, A Bio-Inspired, Motion-Based Analysis of Crowd Behavior Attributes Relevance to Motion Transparency, Velocity Gradients, and Motion Patterns, PLos ONE, 2012, 7(12), e53456

[26] S. Jain, Performance Characterization of Watson Ahumada Motion Detector Using Random Dot Rotary Motion Stimuli, PLos ONE, 2009, 4(2), e4536

[27] A. Pavan, A. Contillo, G. Mather, Modelling adaptation to directional motion using the Adelson-Bergen energy sensor, PLos ONE, 2013, 8(3), e59298

[28] M.-J. Escobar, P. Kornprobst, Action recognition via bio-inspired features: The richness of center-surround interaction, Computer Vision and Image Understanding, 2012, 116, 593–605

[29] R. M. Haefner, B. G. Cumming, Adaptation to natural binocular disparities in primate V1 explained by a generalized energy model, Neuron, 2008, 57, 147–158

[30] A. Borst, J. Haag, D. F. Reiff, Fly Motion Vision, Annual Review of Neuroscience, 2010, 33, 49–70

[31] N. C. Rust, O. Schwartz, J. A. Movshon, E. P. Simoncelli, Spatiotemporal Elements of Macaque V1 Receptive Fields, Neuron, 2005, 46, 945–956

[32] J. R. Cavanaugh, W. Bair, A. Movshon, Nature and Interaction of Signals From the Receptive Field Center and Surround in Macaque V1 Neurons, Journal of Neurophysiology, 2002, 88, 2530–2546

[33] D. J. Heeger, Normalization of cell responses in cat striate cortex, Visual Neuroscience, 1992, 9, 181–197

[34] E. P. Simoncelli, D. J. Heeger, A model of neuronal responses in visual area MT, Vision Research, 1998, 38, 743–761

[35] J. M. G. Tsui, J. N. Hunter, R. T. Born, C. C. Pack, The Role of V1 Surround Suppression in MT Motion Integration, Journal of Neurophysiology, 2010, 103, 3123–3138

[36] C. C. Pack, M. S. Livingstone, K. R. Duffy, R. T. Born, End-Stopping and the Aperture Problem: Two-Dimensional Motion Signals in Macaque V1, Neuron, 2003, 39, 671–680

[37] D. J. Field, Relations between the statistics of natural images and the response properties of cortical cells, Journal of the Optical Society of America A, 1987, 4, 2379–2394

[38] P. Kovesi, Image features from phase congruency, Journal of Computer Vision Research, 1999, 1, 1–26

[39] F. Raudies, E. Mingolla, H. Neumann, A Model of Motion Transparency Processing with Local Center-Surround Interactions and Feedback, Neural Computation, 2011, 23, 2868–2914
[40] F. Mechler, D. S. Reich, J. D. Victor, Detection and Discrimination of Relative Spatial Phase by V1 Neurons, The Journal of Neuroscience, 2002, 22(14), 6129–6157
[41] M. Carandini, D. J. Heeger, Summation and division by neurons in primate visual cortex, Science, 1994, 264(5163), 1333–1336
[42] C. Koch, Biophysics of Computation: Information Processing in Single Neurons, 2004
[43] A. L. Hodgkin, A. F. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve, The Journal of Physiology, 1952, 117(4), 500–544
[44] P. Bayerl, H. Neumann, Disambiguating Visual Motion Through Contextual Feedback Modulation, Neural Computation, 2004, 16, 2041–2066
[45] T. Brosch, H. Neumann, Computing with a Canonical Neural Circuits Model with Pool Normalization and Modulating Feedback, Neural Computation, 2014, 26, 2735–2789
[46] C. Beck, H. Neumann, Combining Feature Selection and Integration – A Neural Model for MT Motion Selectivity, PLoS ONE, 2011, 6(7), e21254
[47] F. Raudies, H. Neumann, Developing and Applying Biologically-Inspired Vision Systems: Interdisciplinary Concepts., 2012, 121–153
[48] J. D. Bouecke, E. Tlapale, P. Kornprobst, H. Neumann, Neural Mechanisms of Motion Detection, Integration, and Segregation: From Biology to Artificial Image Processing Systems, EURASIP Journal on Advances in Signal Processing 2011, 2011
[49] A. Borst, M. Egelhaaf, J. Haag, Mechanisms of Dendritic Integration Underlying Gain Control in Fly Motion-Sensitive Interneurons, Journal of Computational Neuroscience, 1995, 2, 5–18
[50] M. Maruyama, D. D. Palomo, A. A. Ioannides, Stimulus-Contrast-Induced Biases in Activation Order Reveal Interaction Between V1/V2 and Human MT+, Human Brain Mapping, 2009, 30, 147–162
[51] A. M. Sillito, J. Cudeiro, H. E. Jones, Always Returning: Feedback and Sensory Processing in Visual Cortex and Thalamus, Trends in Neurosciences, 2006, 29(6), 307–316
[52] J.-M. Hupé, A. C. James, P. Girard, S. G. Lomber, B. R. Payne, J. Bullier, Feedback Connections Act on the Early Part of the Responses in Monkey Visual Cortex, Journal of Neurophysiology, 2001, 85, 134–145
[53] A. Angelucci, J. B. Levitt, E. J. S. Walton, J.-M. Hupé, J. Bullier, J. S. Lund, Circuits for Local and Global Signal Integration in Primary Visual Cortex, The Journal of Neuroscience, 2002, 22(19), 8633–8646
[54] J. M. G. Tsui, C. C. Pack, Contrast sensitivity of MT receptive field centers and surrounds, Journal of Neurophysiology, 2011, 106, 1888–1900
[55] S. Deneve, P. E. Latham, A. Pouget, Reading population codes: a neural implementation of ideal observers, Nature Neuroscience, 1999, 2(8), 740–745
[56] S. Treue, K. Hol, H.-J. Rauber, Seeing multiple directions of motion – physiology and psychophysics, Nature Neuroscience, 2000, 3(3), 270–276
[57] A. Pouget, P. Dayan, R. Zemel, Information processing with population codes, Nature Reviews Neuroscience, 2000, 1, 125–132
[58] S. Baker, D. Scharstein, J. P. Lewis, IEEE International Conference on Computer Vision, 2007
[59] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, R. Szeliski, A database and evaluation methodology for optical flow, International Journal of Computer Vision, 2011, 92, 1–31
[60] J. L. Barron, D. J. Fleet, S. S. Beauchemin, Performance of optical flow techniques, International Journal of Computer Vision, 1994, 12, 43–77
[61] M. Otte, H.-H. Nagel, European Conference on Computer Vision, 1994, 49–60
[62] D. J. Fleet, A. D. Jepson, Computation of component image velocity from local phase information, International Journal of Computer Vision, 1990, 5, 77–104
[63] F. Solari, M. Chessa, S. P. Sabatini, An integrated neuromimetic architecture for direct motion interpretation in the log-polar domain, Computer Vision and Image Understanding, 2014, 125, 37–54
[64] B. Dellen, F. Wörgötter, A local algorithm for the computation of image velocity via constructive interference of global Fourier components, International Journal of Computer Vision, 2011, 92, 53–70
[65] F. Solari, M. Chessa, N. V. K. Medathati, P. Kornprobst, What can we expect from a V1-MT feedforward architecture for optical flow estimation?, Signal Processing: Image Communication, 2015
[66] P. Lennie, Single units and visual cortical organization, Perception, 1998, 27, 889–935