



HAL
open science

A neural model of luminance-gated recurrent motion diffusion for 2D motion integration and segmentation

Émilien Tlapale, Guillaume S. Masson, Pierre Kornprobst

► **To cite this version:**

Émilien Tlapale, Guillaume S. Masson, Pierre Kornprobst. A neural model of luminance-gated recurrent motion diffusion for 2D motion integration and segmentation. [Research Report] RR-6944, 2009. inria-00360277v2

HAL Id: inria-00360277

<https://inria.hal.science/inria-00360277v2>

Submitted on 10 Jun 2009 (v2), last revised 3 Mar 2010 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

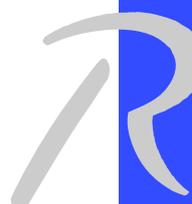
*A neural model of luminance-gated recurrent motion
diffusion for 2D motion integration and
segmentation*

Émilien Tlapale — Guillaume S. Masson — Pierre Kornprobst

N° 6944

February 2009

Thème BIO



*R*apport
de recherche



A neural model of luminance-gated recurrent motion diffusion for 2D motion integration and segmentation

Émilien Tlapale* , Guillaume S. Masson[†] , Pierre Kornprobst[‡]

Thème BIO — Systèmes biologiques
Projet NeuroMathComp

Rapport de recherche n° 6944 — February 2009 — 46 pages

Abstract: We propose a model of motion integration modulated by luminance information, which is able to explain the percept on a large class of motion stimuli facing the aperture problem. This model is related to other multi-layer architectures incorporating both feedforward, feedback and inhibitive lateral connections and is inspired by the motion processing cortical areas in the primate (V1, V2, MT). Our main contribution is to propose a new anisotropic integration model where motion diffusion through recurrent connectivity between layers working at different spatial scales is gated by the luminance distribution in the image. This simple model offers a competitive alternative to models based on a large set of cortical layers implementing specific form or motion features detectors. We demonstrate that the proposed approach produces results compatible with several psychophysical experiments concerning not only the resulting global motion percept but also the motion integration dynamics. It can also explain several properties of MT neurons regarding the dynamics of selective motion integration. As a whole, the paper affords an improved motion integration model which is numerically tractable and reproduces key aspect of cortical motion integration.

Key-words: dynamics, form, GPGPU, motion, perception

* Emilien.Tlapale@sophia.inria.fr

† Guillaume.Masson@incm.cnrs-mrs.fr

‡ Pierre.Kornprobst@sophia.inria.fr

Un modèle neural récurrent de diffusion du mouvement selon la luminance pour l'intégration la segmentation du mouvement 2D

Résumé : Nous proposons un modèle d'intégration du mouvement modulée par la forme capable d'expliquer la perception d'une large classe de stimuli en mouvement dans lesquels se pose le problème de l'ouverture. Notre modèle peut être relié à d'autres architectures multi-couches incorporant à la fois des connexions montantes, descendantes et localement inhibitives, et inspirées par les zones corticales dédiées au mouvement chez le macaque (V1, V2, MT). Notre principale contribution est de proposer un nouveau modèle d'intégration anisotropique du mouvement dans lequel la diffusion entre les différentes couches travaillant à des échelles spatiales différentes est conduite par la distribution de luminance dans l'image. Ce modèle simple offre une alternative efficace aux modèles basés sur un grand ensemble de couches et incorporant des détecteurs spécifiques pour la forme ou le mouvement. Nous démontrons que l'approche proposée produit des résultats compatibles avec différentes expériences menées en psychophysique non seulement pour la perception mouvement global résultante mais aussi pour la dynamique d'intégration du mouvement. Elle est également capable d'expliquer différentes propriétés de l'intégration sélective du mouvement des neurones MT. En définitive, ce papier propose un modèle d'intégration du mouvement amélioré et numérique simulable capable de reproduire les aspects clés de l'intégration du mouvement.

Mots-clés : dynamique, forme, GPGPU, mouvement, perception

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 5 |
| 2 | Description of the model | 9 |
| 2.1 | Model overview | 9 |
| 2.2 | Local motion estimation | 10 |
| 2.3 | General connectivity | 11 |
| 2.4 | Form-modulated diffusion | 12 |
| 3 | Model implementation, read-out and testing | 15 |
| 3.1 | Model parameters and implementation | 15 |
| 3.2 | Model read-out and comparison with biological data | 16 |
| 4 | Results | 19 |
| 4.1 | Dynamics of motion integration | 19 |
| 4.2 | Gratings, apertures and the barberpole illusion | 20 |
| 4.3 | Influence of form on selective motion integration | 26 |
| 4.4 | Results on real sequences | 30 |
| 5 | Discussion | 33 |
| 5.1 | Modeling the neural dynamics of motion integration | 34 |
| 5.2 | Luminance smoothness: a simple rule for gating motion integration | 36 |

Chapter 1

Introduction

In the field of bio-inspired motion processing models, one can distinguish two main classes of approaches. In the first class, models are primarily high level and do not focus on the precise anatomical or functional properties of the visual system. Their goal is to propose some general mechanisms which are the fundamental principles in order to reproduce some psychophysical results. Bayesian models such as [73, 74, 45] provide a good illustration of these high level models. Instead, in the second class, models tend to reproduce some of the key features of the visual system in term of structure and connectivity. Some examples include [63, 13, 16] and our model is also in that branch.

Natural scenes present many sources of ambiguities that must be solved in order to extract a reliable information that can be used to control behavior. Correctly integrating different local features is a key point to solve these ambiguities and therefore, understanding its neural dynamics is crucial. Motion processing has offered a powerful framework to investigate it at many levels. Indeed, In order to compute the global motion of an object within a complex scene, artificial motion processing systems as well as the visual cortex, take local motion estimates as input. Therefore, they have to deal with numerous 1D features corresponding to edges and, generally fewer, 2D features such as corners or line-endings for example. The problem is that 1D features lead to the well known *aperture problem*: edge motion seen through a restricted aperture is highly ambiguous so that an infinite number of visual velocity vectors that are compatible with the physical translation of the object containing that edge. As pointed out by Wallach [71] a spatial integration of 1D features with different orientations can be used to reconstruct this true translation. But 2D features can also be extracted as their motion seen through the same aperture size is not ambiguous. Several different computational rules for motion integration have been proposed over the last three decades. Geometrical solutions such as the Intersection of Constraints (IOC) can recover the exact global velocity vector from the different edges motions [24]. Several studies have proposed that the primate visual system use a similar computation [2]. It remains however unclear how the the visual system can implement the IOC rule. Moreover, the fact that perceived direction does not always correspond, at least initially, to the IOC solution

have supported alternative models that emphasize the role of local 2D features [39] when computing global motion.

While the computational rules actually used by the brain is still highly disputed, there are many physiological evidence that cortical area V1 implements local motion computation and feeds an integrative stage such as area MT. In macaque, neurons solving the aperture problem (i.e. respond to the true motion of a complex pattern and not the normal direction of one of its component) have been found by many different studies, using different 2D motion stimuli [47, 54, 55, 65]. This contrast with the findings that V1 neurons mostly respond to the direction orthogonal to the orientation of the edge drifting across their receptive field [47] albeit some neurons seems to act as local features detectors such as end-stopped cells [34, 55]. Thus, there seems to be a good intuition that 2D motion computation is a two-stage mechanism with local extraction feeding global integration.

There are however two aspects that have been largely ignored within these two-stage feed-forward models. First, motion integration is intrinsically a spatial process. Since objects are rigid, propagating non ambiguous motion information along edges as well as inside surfaces is an essential aspect of motion integration. The role of such diffusion process has been investigated in a small number of biologically-inspired models. Grossberg and colleagues have investigated how local form and motion cues that be integrated through recurrent diffusion [29, 16]. The various versions of this model succeed to solve the aperture problem in many different instances of motion stimuli investigated psychophysically. However, they rely on many different sub-types of local form feature detectors. A similar solution was used by [15, 14] albeit with a more simple and realistic motion computation algorithm.

The aforementioned models cannot deal simply with the complementary aspect of motion integration: motion segmentation. To do so, they need to implement complicated architectures where depth or scale specific layers interplay to segment the object of interest for which global motion were to be computed. One reason why they need such specific segmentation mechanism is that all of them are based on a symmetrical (i.e. Gaussian) diffusion process: they extract specific 2D features, compute their motion and through recurrent connectivity between detection and integration stages, diffuse it in every direction. Computer vision, as well as biological experiment have suggested that an non-isotropic diffusion can be much more powerful in solving the aperture problem and computing global motion. Herein, we want to propose such a simple mechanism based on only two elementary local dimension: orientation and direction. Thus, the model to be presented here instantiates another example of form-motion interaction model but where non-isotropic diffusion play a key role. Moreover, instead of implementing a set of highly selective form detectors, we rather preferred to define an abstract representation of form information, based on luminance smoothness in the image. Such abstract description might collapse both contour and surface representations which have been found in cortical areas V1 and V2 [59, 36, 67]. We propose that both representations contribute in the gating of motion information diffusion to solve the aperture problem within and across apertures.

A second aspect which is often ignored by many 2D motion integration models (e.g. [74, 60] but see [21, 15]) is that biological computation of global motion is highly dynamical.

When presented with simple lines, plaids or barberpoles, perceived direction reported by human observers will shift over time. That initial perception is strongly biased towards the direction orthogonal to the edge orientation (or a vector average solution when several edges are available) reflects the strong influence of 1D motions in the earliest glimpse. Over a time course of several tens of *ms*, area MT neurons solve the aperture problem. As a consequence, perceived as well as tracking directions match the true 2D motion direction only 200 to 300 *ms* after target motion onset [41, 44, 70].

In the present article, we suggest how the spatial integration and the dynamics of motion perception are intrinsically linked and we propose a dynamical model based on a diffusion process gated by low-level, luminance-based features. Our goals were to propose a simplified and tractable model of form-motion integration and segmentation which is based on elementary local features and consist of a small set of layers to be described by dynamical equations. We tested this model against a large class of synthetic motion stimuli to reproduce key aspects of biological motion processing and its dynamics. Lastly, we demonstrate that such a simple model can deal with aperture problem in real videos.

Chapter 2

Description of the model

2.1 Model overview

Our model, represented in Figure 2.1, describes the interactions between several layers processing local motion and form information. The state of each layer is described by a scalar-valued function corresponding to a level of activity at each spatial position and for each velocity.

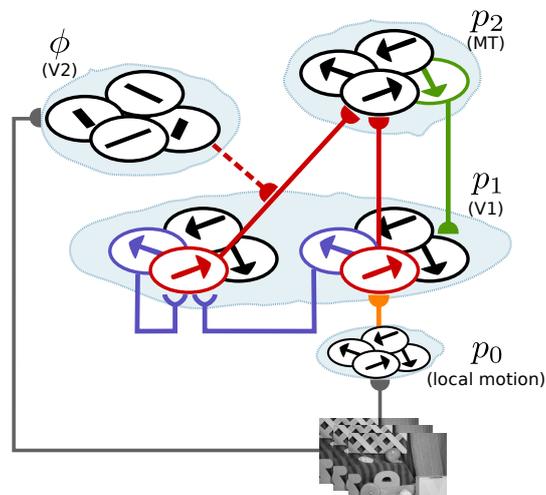


Figure 2.1: Schematic view of the proposed model showing the interactions of the different cortical layers. The motion integration (p_0 , p_1 , and p_2) system is modulated (*dashed red connection*) by a form information (ϕ_I).

The model estimates dynamically the velocity information given an input gray level video sequence:

$$I : (t, x) \in \mathbb{R}^+ \times \Omega \mapsto I(t, x) \in \mathbb{R}^+,$$

where t is the time, $x = (x_1, x_2)$ denotes the spatial position within the 2D-spatial domain $\Omega \subset \mathbb{R}^2$.

The first set of layers is related to motion and is denoted by:

$$p_i : (t, x, v) \in \mathbb{R}^+ \times \Omega \times \mathcal{V} \rightarrow p_i(t, x, v) \in [0, 1], i \in \{0, 1, 2\},$$

where \mathcal{V} represents a space of possible velocities and each function p_i can be interpreted as the state of a cortical area retinotopically organized which describes at each position x the instantaneous activity of a neuron tuned for the velocity v . In brief, layer p_0 implements a local motion estimate through spatio-temporal filtering. These local measurements are integrated to compute local velocity at two different spatial scales in layers p_1 and p_2 . The two layers can be seen as an implementation of detection and integration stages that correspond to cortical areas V1 and MT (see Figure 2.1).

The second type of layer is related to form and denoted by:

$$\phi_I(t, x, \theta) \in \mathbb{R}^+ \times \Omega \times [0, 2\pi[\rightarrow \mathbb{R}^+,$$

which represents the local orientation of the luminance profile from position x in the direction θ . Note that function ϕ_I is an abstract way to encode form information. Such function can be seen as describing V2 neuron properties which can represent orientation of edges from different cues such as luminance, relative motion or disparity.

The coupling between layers, as illustrated in Figure 2.1 defines the connectivity rules between the different layers using a set of coupled differential equations. With that respect, our model follows some previous contributions such as [13, 16]. Feedforward connections transmit information from layers closer to the input to layers deeper in the system while feedback connections connect back to the areas closer to the input. Lateral connections are inhibitive and give to each neuron an input from its neighborhood. The following paragraphs give more detail on the different layers and their connections.

2.2 Local motion estimation

The initial stage of every motion processing system is to compute local motions cues as input to the system. Various models of motion detection have been proposed in the literature, with various degrees of biological plausibility [58, 68, 72, 1]. Here, we define p_0 , the input motion detectors, using an energy model. It provides an efficient way to extract local motion with spatio-temporal filtering kernels corresponding to neuronal receptive fields [31, 63, 60]. We preferred spatio-temporal filtering over simple motion correlators since they can handle a larger class of input stimuli. Moreover fast techniques can be used to estimate local motion due to the properties of steerability and separability properties of certain energy

filters [27, 63, 23] In addition mechanisms to combine the output of such filters into a local motion estimate, the *donut mechanism* [63], have been largely studied (e.g. [4]).

Our local motion input can thus be defined as:

$$f^r(t, x, v) = \sum_{n=0}^{N^r} \left(\sum_{m=1}^{M^r} t_m^r(s_n^r(v))(y_m^r * I)(t, x) \right)^2, r \in \{o, e\}, \quad (2.1)$$

$$p_0(t, x, v) = f^o(t, x, v) + f^e(t, x, v), \quad (2.2)$$

where f^o and f^e are the odd and even responses of the filters, N is the order of the chosen filters, $M = \frac{(N+1)(N+2)}{2}$, y_m are a set of precalculated filters independent of the chosen velocity, and s_n are vectors on frequency plane corresponding to the velocity v combined with the weights given by t_m .

2.3 General connectivity

Given an activity p_0 , the core of our model is defined by the interaction between the two cortical layers p_1 and p_2 , which is modeled by two coupled differential equations:

$$\frac{\partial p_1}{\partial t} = -\lambda_1^l p_1 + S_1 \left(\lambda_1^f p_0 + \lambda^b p_0 p_2 - \lambda_1^i G_{\sigma_1^i}^x * \int_{\mathcal{V}} p_1(t, x, w) dw \right), \quad (2.3)$$

$$\frac{\partial p_2}{\partial t} = -\lambda_2^l p_2 + S_2 \left(\lambda_2^f \int_{\Omega} K_I(t, x, y) p_1(t, y, v) dy - \lambda_2^i G_{\sigma_2^i}^x * \int_{\mathcal{V}} p_2(t, x, w) dw \right), \quad (2.4)$$

$$\text{with } K_I(t, x, y) = G_{\sigma_2^f}(x - y) \phi_I(t, x, \widehat{xy}), \quad (2.5)$$

where G_σ is a Gaussian function of variance σ , \widehat{xy} denotes the angle between the vector $y\vec{x}$ and the horizontal axis and $\lambda, \sigma,$ denote some constants and $S_i(u) = (1 - p_i) \max(0, u)$.

The three main characteristics of this model are as follows:

- **Feedback** from p_2 to p_1 is modulated by $\lambda_b p_o$ in (2.3), in a multiplicative way as in [13]. Thus we use a modulating rather than driving feedback as found in studies of the motion processing system in primates [62].
- **Lateral inhibition** is modeled by the terms $G_{\sigma^i}^x * \int_{\mathcal{V}} p_i(t, x, w) dw$ for both layers p_1 and p_2 . All neurons at a given local neighborhood for all possible velocities inhibits each other. Moreover intra-layers inhibition remains purely local by adding a spatial term: neurons at close retinotopic locations inhibit each other. Such short-range lateral inhibition, usually called recurrent inhibition, leads to a winner-take-all mechanism [22, 78]. Instead of the divisive inhibition found in some models [50, 13], we implemented a subtractive inhibition because ... Note finally, that divisive inhibition can be viewed as the steady-state solution of a dynamical system using subtraction as inhibition mechanism.

- **Form modulated motion diffusion** is the main proposal of our model. Rather than using an isotropic motion diffusion from p_1 to p_2 (2.4) to model wider receptive fields (e.g. [13]), we forced diffusion to be non-isotropic by making it dependent upon local form information. We describe below how we implemented the extraction of form information and how it affects recurrent motion integration.

2.4 Form-modulated diffusion

In order to estimate p_2 , p_1 is integrated in a spatial neighborhood using the weights $K_I(t, x, y)$, as defined in (2.5), which are composed by two terms. The first term weights the connectivity depending on the distance between x and y . The second term is related to the form information defined by:

$$\phi_I(t, x, \theta) = \int_{\Omega} G_{\sigma_x}(x - z) G_{\sigma_{\theta}}(\theta - \widehat{xz}) G_{\sigma_s}(I(t, x) - I(t, z)) dz. \quad (2.6)$$

Function ϕ_I describes the power of diffusion at position x and in the direction θ . In (2.6) the term $G_{\sigma_x}(x - z) G_{\sigma_{\theta}}(\theta - \widehat{xz})$ defines an oriented spatial neighborhood around x (see Figure 2.2). The last term, namely $G_{\sigma_s}(I(t, x) - I(t, z))$ corresponds to a luminosity similarity measure describing form information.

A representation of ϕ_I for all the directions and for a given set of sampled positions is shown in Figure 2.3. The main property of ϕ_I is that it facilitate integration inside similar spatial structures, a property shared by neurons as observed in both psychophysics [40] and cell recordings in macaque area MT [33]. The extension of the integration also depends on the local contrasts: the neighborhood is wider at low contrast than at high contrast (see [52]). Such abstract representation of form information presents several key advantages in the context of 2D motion integration. First, motion integration inside spatial structures is not only performed along borders (see Figure 2.3a), but also propagates inside isoluminant regions (Figure 2.3b). Second, in contrast to some previous work [16] ϕ_I handles smoothness constraints as illustrated in Figure 2.3(c).

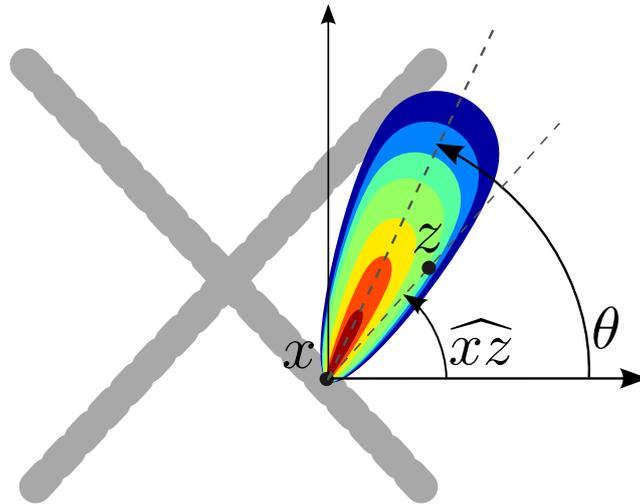


Figure 2.2: Illustration of the oriented spatial neighborhood around x in the direction θ used to compute ϕ_I . Luminosity in this oriented neighborhood is compared with the luminosity at the center.

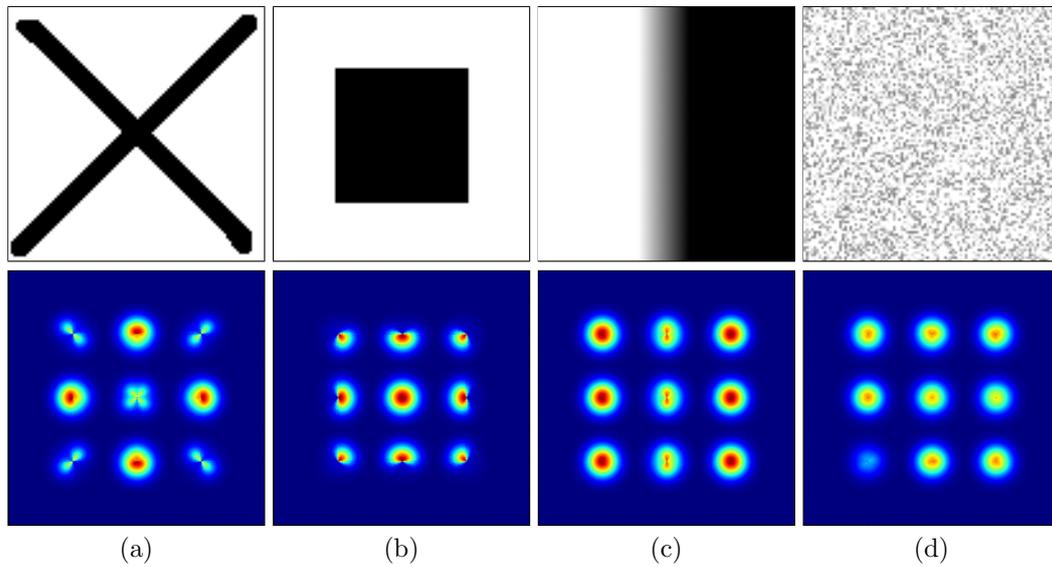


Figure 2.3: *First row* gives a set of input images. *Second row* shows a representation of K_I indicating for a given set of sampled position, the weight by which their neighborhood is integrated.

Chapter 3

Model implementation, read-out and testing

3.1 Model parameters and implementation

The model (2.3)–(2.6) is fully described by a set of 14 parameters that have been manually defined and whose values are given in Table 3.1.

| | | | | | |
|----------------|-------------------|-------------------------|--------------------|-------------------|-----------------|
| Equation (2.3) | $\lambda_1^l = 2$ | $\lambda_1^f = 1$ | $\lambda_1^b = 24$ | $\lambda_1^i = 4$ | $\sigma_1 = 2$ |
| Equation (2.4) | $\lambda_2^l = 2$ | $\lambda_2^f = 16$ | | $\lambda_2^i = 4$ | $\sigma_2 = 12$ |
| Equation (2.5) | $\sigma_x = 8$ | | | | |
| Equation (2.6) | $\sigma_x = 12$ | $\sigma_\theta = \pi/8$ | $\sigma_s = 0.4$ | | |

Table 3.1: Parameters setting used in our model (2.3)–(2.6).

On the following results we choose our discretized velocity space \mathcal{V} to be all the integer pairs $v = (v_x, v_y)$ in a 7×7 regularly spaced grid, although any kind of velocity space can be used in our model. A discretization procedure has to be applied since we work on dynamical equations: we choose the Runge-Kutta algorithm. Moreover since the input is not continuous but is made of successive frames, and because we want more precision than the coarse input, we need intermediary frames. For simplicity, we did not interpolate but choose input similar to the previous frame for all intermediate frames before the next one. We discretized the system with 10 intermediary time steps between two input frames, not including the intermediary frames of the Runge-Kutta algorithm.

Finally, to speed up the simulations we use the GPGPU technology. Due to the anisotropic diffusion process being dependent on input stimulus, our model requires highly computational cost. Thus conventional CPU implementation would be too slow for performing extensive model testing. We implemented our model on graphic cards using NVIDIA's CUDA

technology. We can then take advantage of the parallel nature of our model where the same kind of computation is done for every spatial position. Except for GPU kernels, all the code is written in Python using the SciPy library [35].

3.2 Model read-out and comparison with biological data

The result given by our model is a *distributed* activity response: each function p_i can be interpreted as the state of a cortical area retinotopically organized which describes at each position x the instantaneous activity of a neuron tuned for the velocity v as in Figure 3.1 (a). In order to represent a velocity field, we average at each position the population response across all velocities, thus obtaining a single vector. A motion map v_i was extracted from any layer p_i using:

$$v_i(t, x) = \frac{\sum_{v \in \mathcal{V}} p_i(t, x, v) v}{\sum_{v \in \mathcal{V}} p_i(t, x, v)}, \quad (3.1)$$

and can be represented either by a vector field (b) or a color image (d) with a given color code to represent velocities. Here we use the Middlebury color code [10] (see Figure 3.1 (c)).

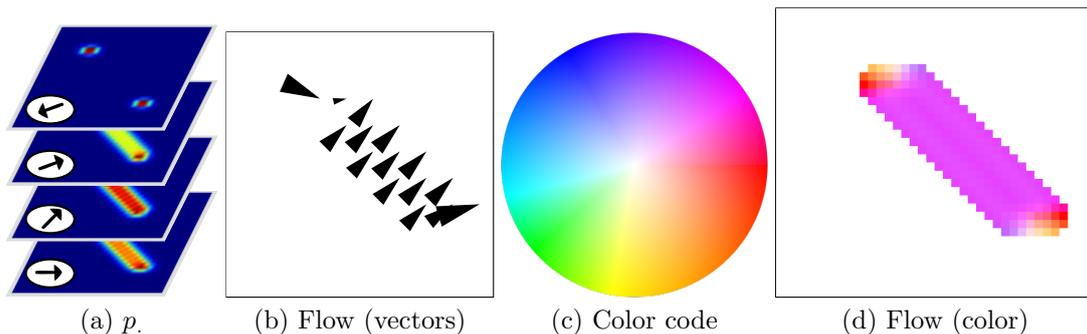


Figure 3.1: Illustration of the motion representation on the layer p_0 for a translating bar. (a) Responses for the different preferred velocities. (b) Subsampled velocity field associated to distribution p , shown in (a) and estimated by (3.1). (c) Middlebury color code [10] used for continuous dense output. (d) Color-based representation of the velocity field in (b).

Another way of visualizing model output and its dynamic is to define as a *read-out* a single output vector $w(t) \in \mathbb{R}^2$. This would also allow to compare the model performances and dynamics with biological results collected at different (i.e. physiological, psychophysical and behavioral) levels. To do so, we propose a simple read-out of activity in layer p_2 , based on averaging the velocity field over the whole stimulus for each frame with a temporal smoothing defined by the dynamical equation:

$$\frac{\partial \mathbf{w}}{\partial t}(t) = \lambda \left(\sum_{x \in \Omega} \mathbf{v}_2(t, x) - \mathbf{w}(t) \right), \quad (3.2)$$

where \mathbf{v}_2 is defined by (3.1).

Given \mathbf{w} , we compute a global direction error as the difference between the true translation direction of the object and the global estimation of the model, $\dot{\mathbf{w}}(t)$. Such velocity error has often been used to describe the dynamics of motion integration at different levels such as single MT neurons [54, 55, 65], perceived direction [41] or tracking eye movements [70, 17, 46]. Thus, by computing a time-dependent global direction error, it becomes possible to qualitatively compare model performance with psychophysical or behavioral experiments.

Chapter 4

Results

Since our goal was to test our model against a wide range of synthetic motion stimuli used for investigating brain dynamics of 2D motion integration and segmentation, we will report model output for several key phenomena of such process. We qualitatively reproduced the neural dynamics of these phenomena, in particular their time courses. Model performances were obtained for full-contrast motion stimuli but different simple changes in image geometry were tested, based on previous psychophysical works. We also systematically compared model output with, or without layer ϕ_I to demonstrate the key role of non-isotropic diffusion constrained by local form luminance information.

We first report results obtained for testing the dynamics of motion integration. Classical motion stimuli such as line-drawing objects, bars, barber-poles, or plaids have been used. Second, we describe the experiments ran with occluded and superimposed motion stimuli to demonstrate that our model can also solved motion segmentation problem without the need of complicated segmentation mechanisms. Lastly, we tested our model against realistic movies as a proof that it can be used in such circumstances with reasonable computing performance.

4.1 Dynamics of motion integration

The dynamics of motion integration and the role of form-based disambiguation mechanisms can be illustrated with the simple example of of the aperture problem in motion perception: a translating bar stimulus as shown in Figure 4.1. When set into motion for short durations, its perceived direction is biased towards the direction orthogonal to its orientation. Such perceptual bias is corrected for longer durations [41, 20]. Similarly, one of us demonstrated that initial tracking direction exhibits the same bias and that eye-tracking direction converges towards the true 2D object motion direction over a period of about 300 *ms* [44, 70]. Similar results were obtained in monkeys by Born and colleagues [17]. Interestingly, when presented with a set of small oriented bars, direction selectivity of MT neurons exhibit the

same temporal dynamics, the optimal direction rotating from the component orthogonal to the bars orientation to the 2D motion direction [54].

As illustrated in Figure 4.1, applying our model to the translating bar stimulus shown in (a), reproduced several of the phenomena described above. Initial first estimation was dominated by local ambiguous (1D) motion measurements, as shown by the velocity field at Frame 0. However, we obtained a gradual 2D motion diffusion inside the bar as shown by the progressive changes in the velocity fields obtained between Frames 0 and 59 (b–e). Thus our model can solve the aperture problem at both local and global scales. The dynamics can be further described by plotting global read-out w defined in (3.2) and shown in (e). To be consistent with the following experiments we prefer to use the direction error between the global read-out w and the 2D global translation direction of the bar, against time 4.1 (g). After a short period of time where the direction error stays constant at about 40° , the estimate of the global motion converged to the true direction (i.e. direction error = 0°) with an exponential decay.

Broken lines of increasing number of segments were also used as shown in Figure 4.2 (a). Similarly to what has been found in psychophysical experiments [41], introducing more line-endings both reduced the initial bias in the global motion estimation and produced a faster exponential decay in the direction error (b). On the contrary, smoothing the luminance profile by applying a Gaussian filter along the bar orientation reduces the contrast of line-endings (c) and thus resulted in a larger initial bias, reaching the asymptotic error of 45° and a longer time constant for error reduction (d). Similar results have been shown at behavioral level in humans [70].

4.2 Gratings, apertures and the barberpole illusion

Different aspects of the 2D motion signals integration can be investigated with gratings moving through different kinds of apertures. For instance, when a moving grating is seen through a rectangular aperture, human observers report a perceived global motion direction that is tilted towards the long axis of the aperture. This phenomenon is known as the *barberpole illusion*. The rotation is stronger as the aspect ratio (i.e., the ratio between the long and short axes of the aperture) is increased. Moreover, human ocular tracking [43] as well as neuronal responses gradually evolved from local (i.e. orthogonal to grating orientation) to global (i.e. along the aperture long axis) direction [55].

Our model can reproduce these different aspects of motion integration for barberpoles, as shown in Figure 4.4. As illustrated by velocity flow fields obtained at different time steps, motion flow is first dominated by 1D motion information while at a much later time step, all local measurements were coherent with the 2D, perceived direction. This dynamics is further illustrated by plotting the time course of the direction error: the estimated global motion is first driven by grating motion direction but then slowly rotates to be aligned with the long axis of the aperture. Right-hand plot of (a) illustrate the early response at $t = 15$ for several barberpole aspect ratios, ranging from 1:4 to 1:7 (Figure 4.5). Higher aspect ratios elicit smaller initial biases and faster convergence.

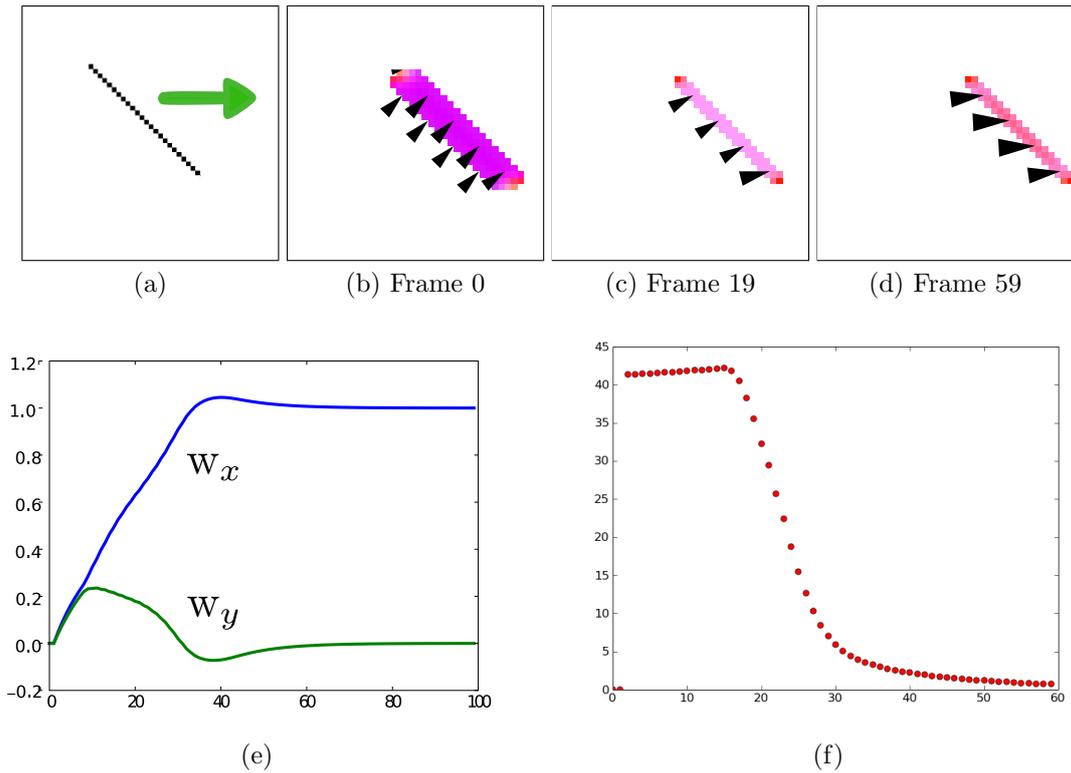


Figure 4.1: Response of the model on a horizontally translating bar presented in (a). (b)-(d) Evolution of the velocity field $v_1(t, x)$. (e) Temporal read-out providing a global motion similar to the one get in eye movements computed by averaging the velocity field. Green and blue correspond respectively to w_x (horizontal) and w_y (vertical) components. (f) Mean tracking error computed as the angular difference between w and the 2D translation direction of the bar.

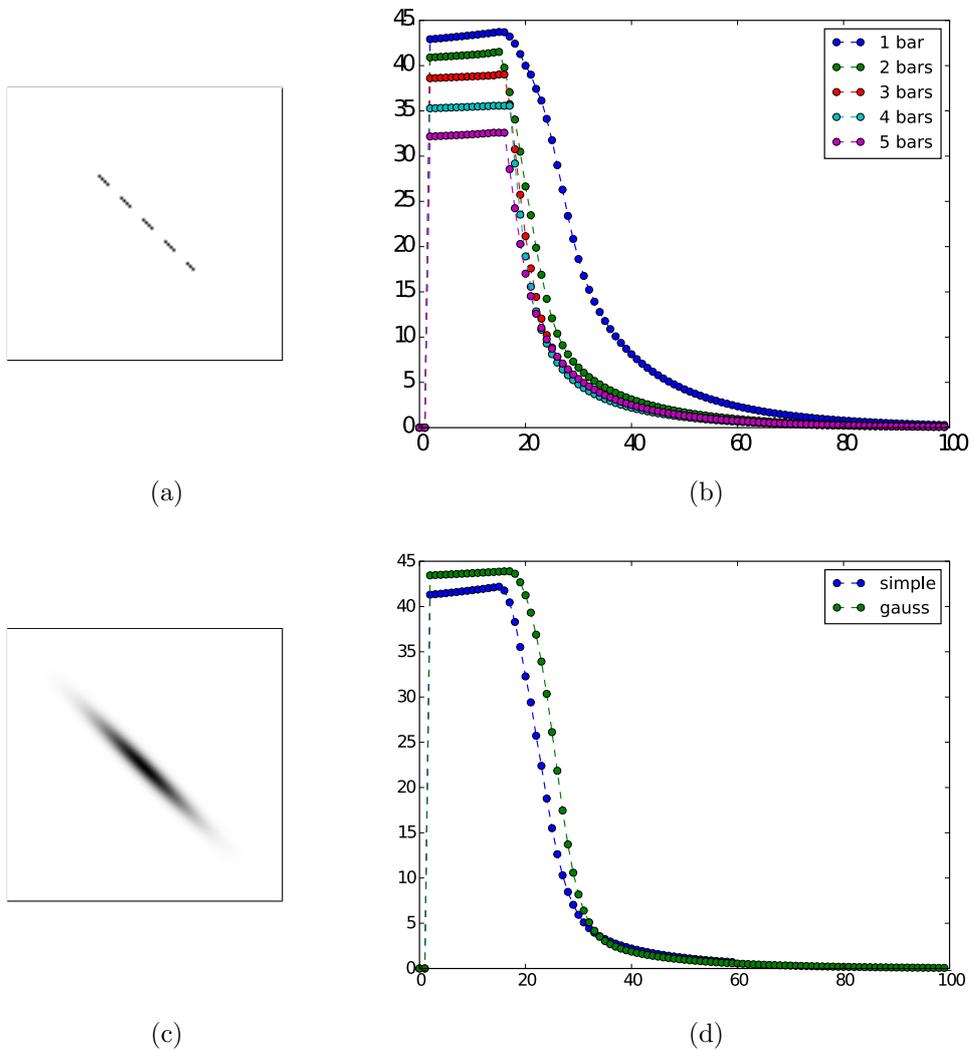


Figure 4.2: (a) Translating broken line stimulus. (b) Direction error plotted against time for different number of cut in the original line. (c) Translating smoothed bar stimulus. (d) The smoothed bar dynamic is slower than the high contrast translating bar dynamic.

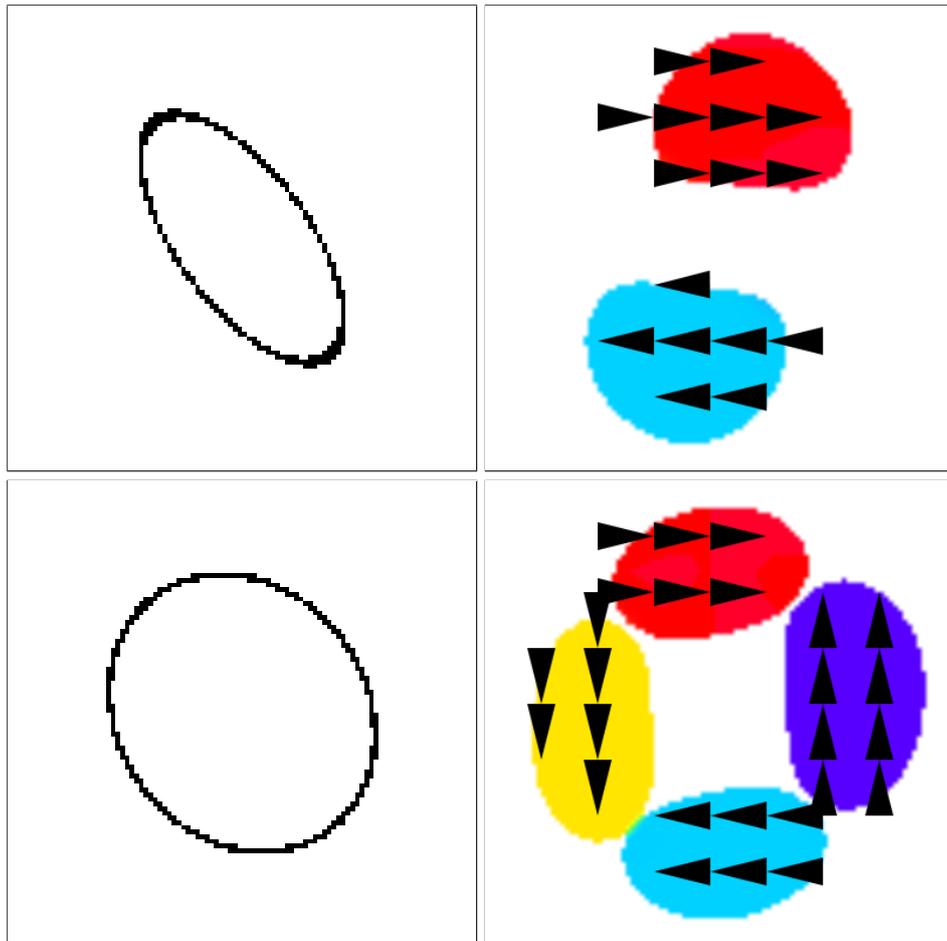


Figure 4.3: Output v_2 of our model for rotating ellipses. On the *first row* a thin ellipse of ratio 9:20 is used and the resulting motion is compatible with rotation. On the *second row* a rotating thick ellipse of ratio 3:4 results in a deforming motion instead of rotation.

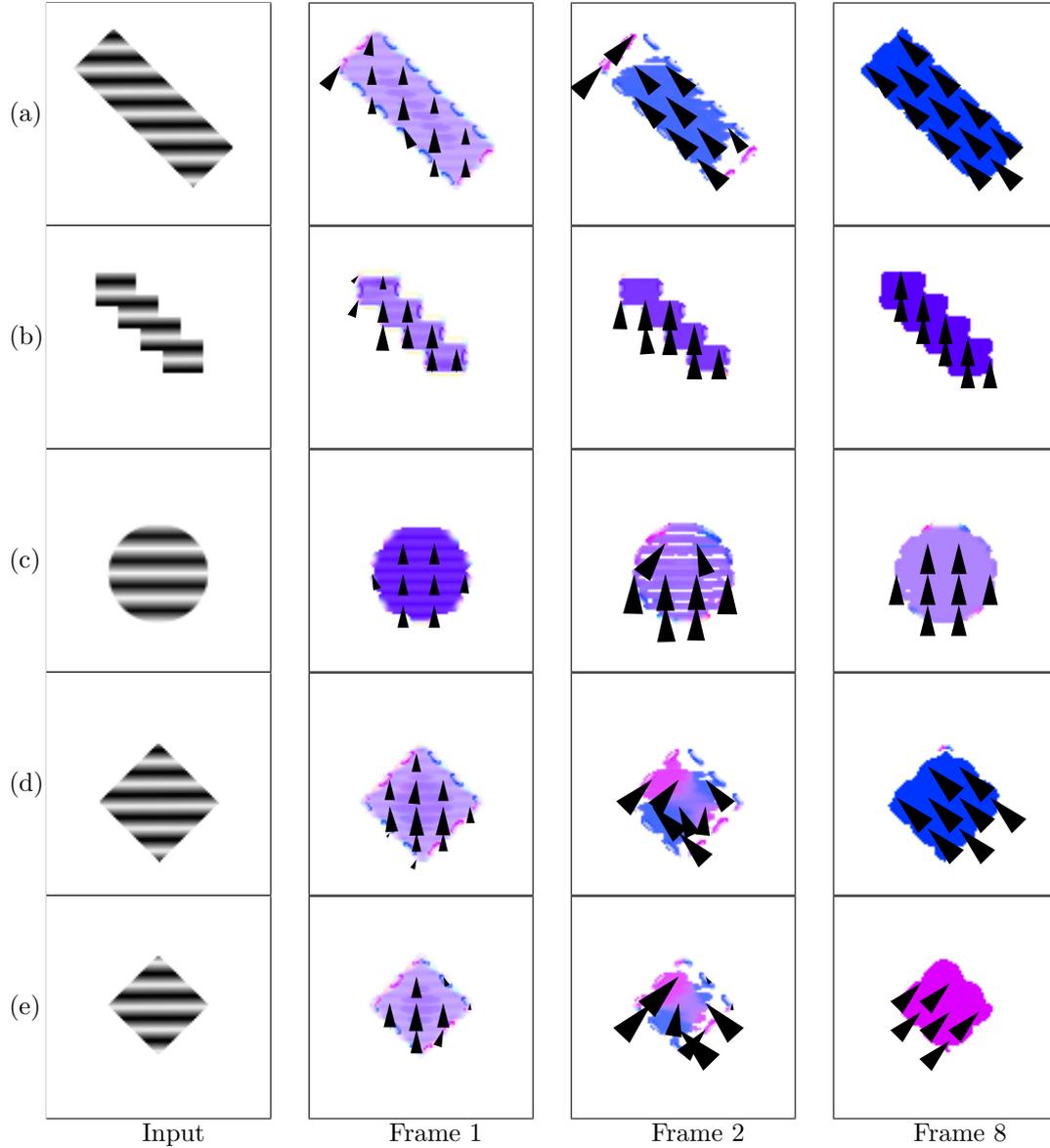


Figure 4.4: (a) uses a grating moving through a rectangular aperture: the *barberpole illusion*. We show the output v_1 , an average colored version of p_1 . (c) shows how changing the aperture and the 2D motion cues it generates influences the perceived motion. (d) considers a grating moving through a square aperture. In (e) a similar grating with different characteristics, size and position, exhibits a totally different output. This stimulus is multistable. (f) presents a grating through a circular aperture.

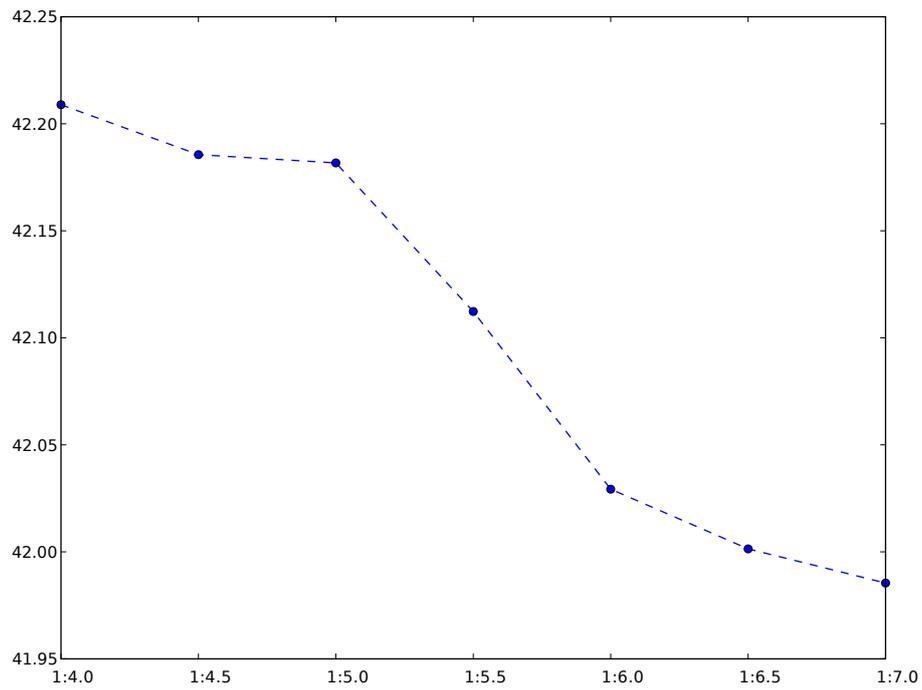


Figure 4.5: Varying the aspect ratio of a barberpole stimulus changes its dynamic. We used ratios ranging from 1:4 to 1:7 and plotted the direction error during the early response ($t=15$). One can see that higher aspect ratios elicited smaller initial biases and faster convergence.

That local 2D motion cues generated by the shape of the aperture dictate the final perceived motion was nicely demonstrated by indenting the long axis of a barberpole [57, 37, 43]. Perceived direction changed towards the grating motion direction as the size of the indentation increased. Our model reproduces a similar behavior. As illustrated in 4.4 (b), changing the aperture geometry introduced new local motion signals that soon dominate the global motion direction and then stay consistent with the grating motion direction. This last global motion was also seen with gratings presented behind a circular aperture (c).

Barberpole motion stimuli with an aspect ratio of 1 (i.e., a square aperture), yield to two interesting phenomena. First, short stimulus presentation gives a perceived motion direction as well as a tracking direction consistent with the grating motion direction [19, 43]. Second, with long motion durations, perceived direction becomes multi-stable, alternating between the grating motion direction to one or the other motion direction found along the edges. Thus, Castet and colleagues [19] showed stochastic fluctuations in the perceived direction of barberpoles with aspect ratio 1:1, yielding to a broad distribution in performance computed over a large set of trials. Perceived directions spanned between the three possible solution: motion along one of the aperture axis or along the grating motion direction. We ran successive model simulation with barberpoles of unity aspect ratio and introducing small fluctuations in either the input image I or the input local motion p_0 . For instance, slightly changing the size of the square aperture resulted in a dramatic changed in global motion estimate, switching from left- to right-upward direction (d-e)). In the same vein, introducing noise into p_0 yield to similar switches. Thus, small changes in stimulus characteristics can lead to a totally different global motion.

4.3 Influence of form on selective motion integration

This later result indicates that biasing motion diffusion with local spatial luminance information plays a role in 2D motion integration. However, motion signals must be integrated within, and only within the object of interest. We explored whether our model can reproduce some key aspects of this motion segmentation by testing its response to a large class of motion stimuli used in both psychophysics and neurophysiology. In particular, we were interested into two aspects of motion integration and segmentation. First, motion signals are integrated only along rigid structure and are not captured by motion from the surrounding [33, 61]. Second, a large bulk of psychophysical data have suggested that features motion are discarded when they do not belong to the moving surface (i.e. when they are extrinsic) [61].

In Figure 4.6 we tested our model with the stimulus used in [33] to test selective motion integration by area MT neurons. A square moving in the lower right direction has its top edge removed and replaced by a set of points moving randomly downward. The points motion directions spanned the velocity distribution existing at the center of an edge due to the aperture problem at the center of an edge. Our model gives results similar to those observed with MT neurons recording: the ambiguity is not solved at the location of the missing edge and the velocity field is thus averaged as a downward motion. Furthermore,

the aperture problem along the three remaining edges was correctly solved so that all local motions are coherent with the right-downward translation of the object. In brief, the two sets of global motion did not captured each other. Removing the form information (ϕ_I) resulted in isotropic motion diffusion and therefore edges motion processing was captured by the downward motion of the random dot patterns.

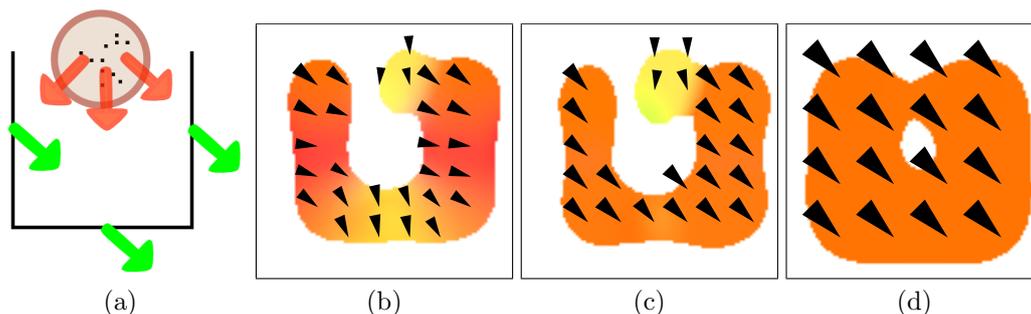


Figure 4.6: (a) Square moving left and down with points moving randomly down at top removed edge (see [33]). (b) Velocity field in p_2 at the beginning with the aperture problem. (c) Iterating does not solve the aperture problem on top. (d) Same output with the form information disabled (isotropic diffusion).

We next used the chopstick illusion to illustrate the influence of form information onto motion processing. The first stimulus was made of two horizontally translating bars, shown in Figure 4.7. We thus have two sets of non ambiguous motion information arising from the end of lines (i.e., horizontal motion), and from the bars intersection (i.e, vertical motion). In Figure 4.7, we illustrate the velocity field v_1 computed at different different points in time. Our results are coherent with the phenomena reported by psychophysical experiments: under these conditions, two bars are perceived as moving in opposite direction. We also show that velocity flow fields were coherent at the two different spatial scales v_1 and v_2 showing that feedback allows to have coherent motion representation at different stages of the motion pathway. Removing the ϕ_I layer, resulted in the opposite motion perception. The two bars moved coherently upward, corresponding to a single cross being translated vertically.

Such coherent motion percept also occurs when line-endings are made extrinsic by placing two horizontal occluders (see Figure 4.8). Applying a simple rule for non-isotropic motion diffusion was enough to reproduce this phenomena. Figure 4.8 illustrates the temporal dynamics of motion integration for the occluded chopstick motion stimulus. Horizontal motion features arising at the intersections between lines and occluders are normally extracted (see v_2 flow fields) but are not propagated inside the line-drawing figures. On the contrary, 2D motion signal arising the intersection between the two lines was propagating along the edges so that after 20 frames, the two bars are perceived as moving coherently in the upward direction. Thus contextual modulations of motion integration can be implemented, simulating

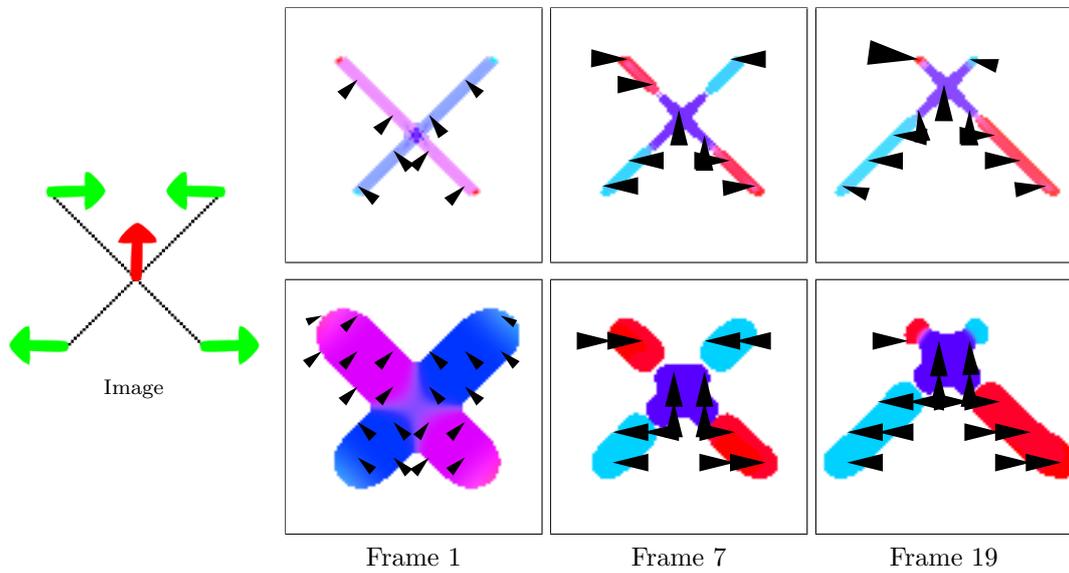


Figure 4.7: Non-occluded chopstick illusion made of two horizontally translating bars. For perception, the end of line 2D information is propagated (*green*), the intersection 2D information is inhibited (*red*). We display the velocity fields v_1 and v_2 obtained from our model respectively in the *first row* and *second row* respectively.

different percepts such as coherent (i.e., one single object) or incoherent (i.e., overlapping objects) motion of the two edges.

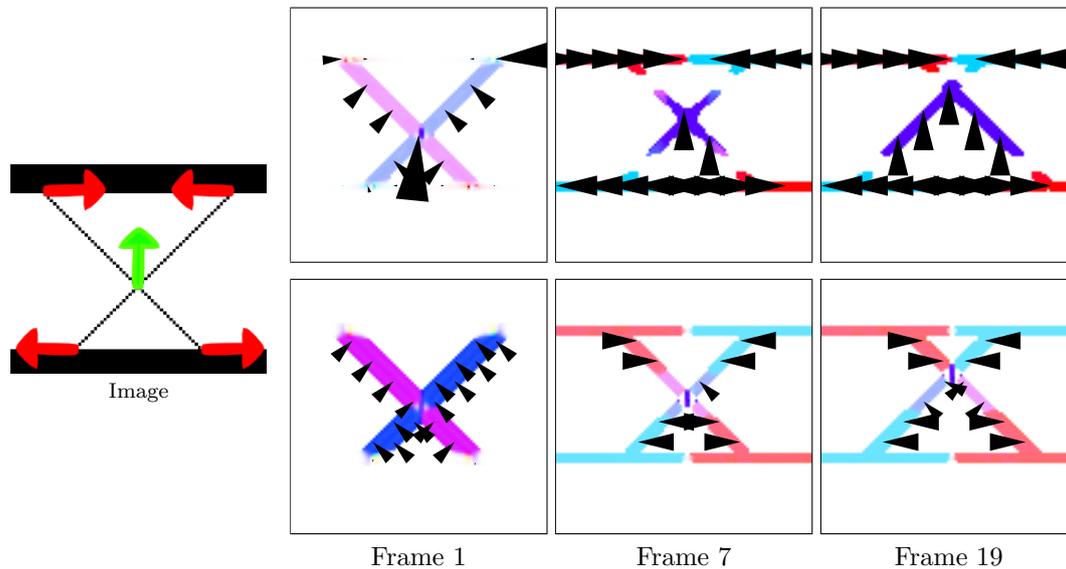


Figure 4.8: Occluded chopstick illusion which adds two rectangular occluders at the end of lines level. This change the perception to a single vertical motion. A result which is also reproduced by our model. We display the velocity fields v_1 obtained from our model in the *first row* using our form modulated diffusion, and with an isotropic diffusion in the *second row*.

Lorenceanu and colleagues [61] introduced the moving diamond stimulus where a rigid diamond is set into rotation behind visible or invisible apertures (see Figure 4.9). They showed that with visible occluders (a), subjects perceived coherent motion of the four edges, forming a rotating diamond. When occluders were made invisible (b), motion coherency largely broke down and subjects reported the percept of four independent edges moving along the vertical axis. We successfully obtained results in close agreement with these psychophysical results: with visible apertures (a), all local motions were coherent with the rotation of a single object. On the contrary, with invisible apertures (b), local motion estimates corresponded to two set of parallel bars moving either upward or downward.

In a final test using synthetic motion sequences, we tested our model performance against rotating ellipses of different aspect ratios.

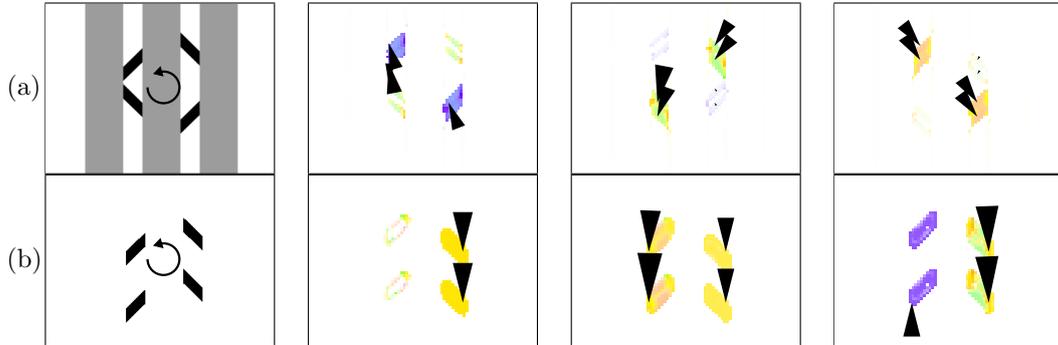


Figure 4.9: Output of our model (v_1) applied to Shiffrar and Lorenceau's diamond illusion [61] for three different frames. With visible occluders, local motion is consistent with the physical rotation of the moving diamond, as shown by the output of our model (v_1). On the contrary, with invisible occluders, motion field is consistent with two translational motions instead of the physical rotation global motion

4.4 Results on real sequences

Since we use motion energy filters instead of simple correlators [13], our model is able to process stimuli containing a broad range of spatio-temporal frequencies. The combination of fast implementation techniques like steerable separable filters [4, 23], and parallel hardware using graphic cards allowed us to apply our dynamical system to more complex and larger inputs such as natural sequences. We used the Hamburg taxi sequence [11] In Figure 4.10 we show the output for the three different layers, p_0 , p_1 , and p_2 , and for three different frames. Clearly, our model is able to unambiguously extract the translation of each moving object from the surrounding. Patchiness of the output velocity flow fields was due to the winner-take-all mechanisms as well as the absence of inter-velocities interactions.

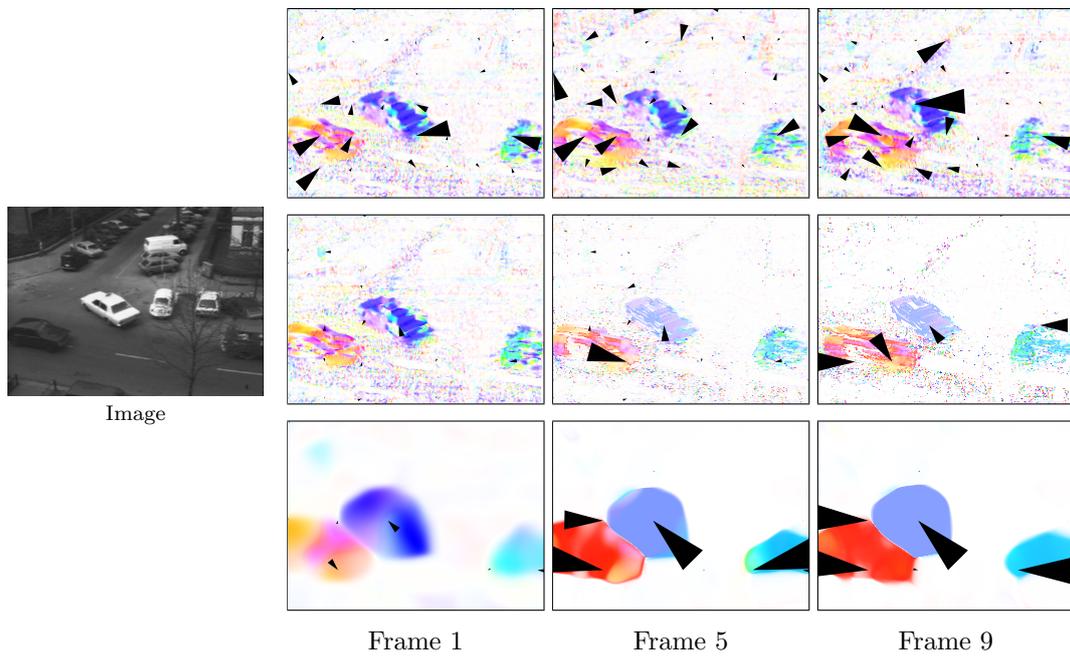


Figure 4.10: Hamburg taxi sequence [11]. We display in the *first row* a velocity field extracted from p_0 . *Second* and *last rows* respectively corresponds to velocity fields from p_1 and p_2 .

Chapter 5

Discussion

In the present study, we designed a form-constrained motion diffusion model to solve 2D motion integration and segmentation. We applied this dynamical system to synthetic motion stimuli to reproduce several key phenomena of 2D motion integration as evidenced by psychophysical, behavioral and neurophysiological studies. In particular, we have been able to reproduce the temporal dynamics of motion integration and its dependency upon stimulus characteristics. Furthermore, we demonstrate that such a simple computational rule, non-isotropic, form-biased diffusion of motion information, is sufficient to explain a large set of contextual modulations of motion integration, in particular those related to selective motion integration, a special case of motion segmentation. Lastly, our model can be successfully applied to natural sequences with reasonable computational performance. Below we first compare our model performance with other existing models which are based on form-motion interactions. Next, we discuss how non-isotropic, form-biased diffusion of motion information implemented by recurrent networks are biologically plausible and give a simple expression of the interactions found between local features in different cortical areas along the primate motion pathway.

Our model successfully reproduces the temporal dynamics of 2D motion integration for a large set of motion stimuli used in investigating visual motion perception and its neuronal basis. First, for lines, plaids and barberpoles, we found that during the first iterations we have almost not contribution of 2D motion signals such as generated by line-endings, blobs or terminators, respectively. This is consistent with the observations made in area MT that the very first *ms* of direction tuning are driven by component motions (e.g. [54, 65, 55]). At behavioral level, Masson and colleagues found similarly that the earliest phase of ocular following responses to either unikinetic plaids or barberpoles is only driven in the direction of grating motion [43, 12]. The origin of such delay between 1D and 2D driven responses has been highly controversial. Some authors have attributed it to the delay seen in the emergence of end-stopping properties of V1 neurons [55]. This temporal dynamics might be related to the timing of the underlying center-surround interactions [9]. However, the relative contribution of both lateral and feedback recurrent connectivity to such temporal dynamics

is still unclear [6]. In our model, we have not implemented specific features detectors, neither their particular temporal dynamics. We have also not implemented specific delay between motion and form pathways although it has been shown that form-driven responses in area V2 are somewhat delayed relative to the fast MT neuronal responses (see [38] for a review). Nevertheless, our simulations show that contribution of 2D features is somewhat delayed, as compared to 1D-driven responses. This could be explained by the poor signal strength of local 2D motion signals as well as by the need to recurrent computation to extract them. The earliest dynamics indeed reflect the time needed for local directions corresponding to 2D features to be amplified and to inhibit the other, nearby ambiguous motion signals. Further work will be done to investigate how the precise timing of 2D motion integration can be simulated by implementing the timing architecture of the early visual pathways [38, 18].

Next, our model can reproduce the temporal dynamics of 2D motion integration as evidenced by a large number of studies at both psychophysical [77, 20, 41], behavioral [44, 70, 17] and neurophysiological [54, 65, 55] levels. In brief, the estimate of global motion, as computed by our simple read-out mechanism, gradually shifts over time. Following an exponential decay, direction error decreases from the initial bias towards 1D motion (or its vector average for multiple edges/components pattern) to the actual 2D translation of the object. Both initial bias and time constant of the decay varies with contrast of local non-ambiguous features, line length, barberpole aspect ratio and so on. All these scaling factors affect the dynamics of lateral diffusion. Thus recurrent dynamics needed for the diffusion of motion information can largely explain the observed dynamics of motion integration. It provides a platform to further investigate biologically realistic neuronal architectures can underlie such computation, within a single and simple paradigm of visual computation.

5.1 Modeling the neural dynamics of motion integration

Several other models have been designed to simulate the temporal dynamics of 2D motion integration. A first attempt was made by Wilson and coworkers to explain the transition of perceived direction between vector average and IOC solutions for type II plaids in human observers [75]. This model was further expanded to account for barberpole motion perception [39]. As all two-motion pathways models, they postulate that 1D and 2D motion features are extracted through parallel pathways, the later being delayed. Such delay, and the winner-take-all competition performed at the integration stage as thought to be sufficient to explain the temporal dynamics of 2D motion integration. These models do not implement any diffusion process and therefore global motion does not correspond to an homogeneous velocity flow fields. They clearly miss the spatial properties of motion integration and therefore cannot account to geometrical changes such as line lengths or barberpole aspect ratios.

On the contrary, the role of diffusion has been previously investigated by different groups, in both computer vision and computational neurosciences. Estimating motion from image sequences is a central problem for many applications such as compression, image registration, structure from motion, robotics, video denoising or colorization, ego-motion or estimation

of time to collision. For this reason, solving the aperture problem has been also widely considered in the computer vision community and many kinds of approaches have been considered to estimate the so-called optical flow [66, 51, 11, 3]. Almost all these approaches use the classical brightness constancy assumption that relates the gradient of brightness to the components of the local flow to estimate the optical flow. Because this problem is ill-posed, additional constraints are usually required. One possible solution to account for the smoothness constraint in general is to define a variational formulation or a model based on partial differential equations, as shown in [8]. For example, one of the earliest models was proposed by Horn and Schunk [32] where smoothness was defined by minimizing a quadratic term of the velocity components gradient.

Interestingly, several general properties of variational approaches can be related to computations within cortical maps [69]. The major one is that in both cases, one manipulates local processes with two main ingredients: a local dynamic and a diffusion in a local neighborhood, exactly what happens in a neural network. Diffusion is certainly the key aspect of variational approaches, and it is very related to the integration processes that we have defined in this paper. In terms of optic flow models, diffusion is often used to prevent the methods from smoothing the solution across the flow discontinuities (see for example [5, 76, 7]). To do so, diffusion is most often isotropic. Another set of approaches exists which is more related to our introduction of form/luminance modulation of diffusion. Nagel and Enkelmann [49, 48] propose an oriented smoothness constraint in which smoothness is not imposed across steep intensity gradients (edges) in an attempt to handle occlusions.

Several models have been previously designed to investigate the role of motion diffusion in the context of motion integration [21, 16, 14]. These models are able to capture several aspects of motion integration such as the propagation of feature tracking estimates [21, 30]. Some of these models implement isotropic motion diffusion by Gaussian distributions of activity both within layers and between layers using recurrent connectivity [14]. They can simulate the temporal dynamics of motion integration for simple motion stimuli but cannot render more complex selective motion integration without implementation more complex rules such as T-junctions cancellation or using different depth layers. Their model also fails to account for motion grouping across occluders. To solve this latter aspect, Grossberg and colleagues introduced the idea of non-isotropic motion integration that can be biased either by local form information as well as by depth cues ([16]). By doing so, the various version of the model designed by Grossberg and colleagues, also called FORMOTION model, can solve some aspects of motion grouping within and across occluders and therefore reproduce the perceived global motion direction observed with motion stimuli such as the occluded diamonds [61, 40] or the chopsticks. Notice that form-motion interactions is used only to disambiguate motion information at the stage of area MT. No feedback is implemented between areas MT and V1 within the motion pathway, so that local motion information remains constant at the earliest stage of motion processing. Recurrent interactions between motion processing layers are implemented between areas MT and MST to perform motion grouping at the highest spatial scale. Notice that feedback connectivity does exist but only between area MT and the V1 form module to solved local ambiguities in the static

distribution of luminance and thus uses motion information for improving 3D figure-ground separation.

The FORMOTION model relies heavily on the assignment of each object to a given depth layer. To do so, they implement a complex architecture with 6 processing stages in the form pathway and 7 stages in the motion pathway. Multiple, feedforward and feedback interactions are implemented at different levels [16] and postulate several types of highly specific form and motion detectors. In contrast, we have attempted to design a very simplistic model to understand how diffusion of motion information can be constrained using some low-level form information such as smoothness in luminance distribution. With only 4 layers, our model can reproduce as many perceptual phenomena as the FORMOTION model. It also implements in a dynamical recurrent system based on (i) a generic mechanism for extracting local motion and (ii) a simple rule for constraining motion diffusion. We believe that such a powerful model can then be extended to understand how cortical architectures implement these elementary operations.

5.2 Luminance smoothness: a simple rule for gating motion integration

We have implemented a simple mechanism for using form information in the context of motion integration. In particular, we did not implement complex local features detectors such as end-stopped cells or dipole cells such as in [16]. In brief, our form layer ϕ_I indicates directions in the image along which luminance is nearly uniform. Such abstract definition of form information incorporates form representation as well as surface representation. Neurons in the early stage of the visual cortex are known to respond to a specific orientation in the luminance distribution. As a consequence, they can signal a local contour within their receptive field [34]. Abrupt changes in the luminance profile along the contour can be signaled by another type of neurons found in area V1, end-stopped cells [34, 55]. Albeit neuronal selectivity for more complex shapes can be found higher along the ventral cortical pathway, it is still unclear how many different elementary features detectors can be found in the earliest stage of visual form processing. Most of the existing models face this problem since they rely heavily on the implementation of local features detectors to extract contours, shapes and so on and then feed the motion pathway using some non linear interactions [16]. The ϕ_I function used herein that signal isoluminance directions and use this information for guiding motion integration without the need of explicit feature detectors. Moreover, it enables some surface representation by signaling luminance smoothness along a wide range of direction. Recent studies have pointed out that luminance information is encoded in the earliest stage of cortical processing [59, 56, 28]. At population level, patches of neurons are strongly activated by large stimuli of uniform luminance and are located in close relationship with the singular points in the orientation-preference maps [36, 67]. Such representation of uniform surfaces based on luminance distribution has been related to brightness perception [59]. Our model suggest that such population of neurons can also be involved in the spatial

integration of motion information. Interestingly, some MT neurons can signals motion over regions of uniform luminance, corresponding to the center of a disk with edges located far outside the receptive field [53]. On the contrary, these cells remained unresponsive to a circle of same diameter. Our model can reproduce this dynamics, thanks to the ϕ_I function.

Using luminance information, and in particular the fact that luminance profiles smoothly vary both along single (edges) or multiple (surface) directions might be very efficient for computing a global solution for object motion. There is plenty of evidence suggesting a tight linkage between the statistics of natural scenes and the design of the visual system [28]. Considerable attention have been paid to the statistics of contrast distribution and its relationships with the properties of elementary local features detectors (see [64] for a review). Recent studies have shown the importance of luminance distribution as well [42, 26] and pinpoint its role in the neural dynamics of local information processing [42, 28] but also surface segmentation [25]. Our model suggests that further work shall be conducted to better understand how these two aspects of visual objects (i.e. edges and surfaces) can be used to gate motion integration performed within the V1-MT recurrent network.

Acknowledgments

This research work has received funding from the European Community's Seventh Framework Programme under grant agreement n°215866, project SEARISE and the Région Provence-Alpes-Côte d'Azur. G.S.M. was supported by the CNRS, the European Community (FACETS, IST-FET, Sixth Framework, n°025213) and the Agence Nationale de la Recherche (ANR, NATSTATS).

Bibliography

- [1] E.H. Adelson and J.R. Bergen. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2:284–299, 1985.
- [2] EH Adelson and JA Movshon. Phenomenal coherence of moving visual patterns. *Nature*, 300(5892):523–525, 1982.
- [3] J.K. Aggarwal and N. Nandhakumar. On the computation of motion from sequences of images — a review. *Proc. IEEE*, 76(8):917–935, August 1988.
- [4] D.S. Alexiadis and G.D. Sergiadis. Narrow directional steerable filters in motion estimation. *Computer Vision and Image Understanding*, 110(2):192–211, 2008.
- [5] Luis Alvarez, Rachid Deriche, Théo Papadopoulo, and Javier Sánchez. Symmetrical dense optical flow estimation with occlusions detection. *International Journal of Computer Vision*, 75(3):371–385, dec 2007.
- [6] A. Angelucci and J. Bullier. Reaching beyond the classical receptive field of v1 neurons: horizontal or feedback axons? *J Physiol Paris*, 97(2–3):141–154, Mar–May 2003.
- [7] G. Aubert, R. Deriche, and P. Kornprobst. Computing optical flow via variational techniques. *SIAM Journal of Applied Mathematics*, 60(1):156–182, 1999.
- [8] G. Aubert and P. Kornprobst. *Mathematical problems in image processing: partial differential equations and the calculus of variations (Second edition)*, volume 147 of *Applied Mathematical Sciences*. Springer-Verlag, 2006.
- [9] Wyeth Bair, James R. Cavanaugh, and Anthony Movshon. Time course and time–distance relationships for surround suppression in macaque V1 neurons. *The Journal of Neuroscience*, 23(20):7690–7701, aug 2003.
- [10] S. Baker, D. Scharstein, JP Lewis, S. Roth, M.J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. In *International Conference on Computer Vision, ICCV’07*, pages 1–8, 2007.

-
- [11] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. *The International Journal of Computer Vision*, 12(1):43–77, 1994.
- [12] F.V. Barthélemy, L.U. Perrinet, E. Castet, and G.S. Masson. Dynamics of distributed 1d and 2d motion representations for short-latency ocular following. *Vision Research*, 48(4):501–522, 2008.
- [13] P. Bayerl and H. Neumann. Disambiguating visual motion through contextual feedback modulation. *Neural Computation*, 16(10):2041–2066, 2004.
- [14] P. Bayerl and H. Neumann. Disambiguating visual motion by form–motion interaction – a computational model. *International Journal of Computer Vision*, 72(1):27–45, 2007.
- [15] Pierre Bayerl and Heiko Neumann. Attention and figure-ground segregation in a model of motion perception. *Journal of Vision*, 5(8):659–659, sep 2005.
- [16] J. Berzhanskaya, S. Grossberg, and E Mingolla. Laminar cortical dynamics of visual form and motion interactions during coherent object motion perception. *Spatial Vision*, 20(4):337–395, 2007.
- [17] Richard T. Born, Christopher C. Pack, Carlos Ponce, and Si Yi. Temporal evolution of 2–dimensional direction signals used to guide eye movements. *Journal of Neurophysiology*, 95:284–300, 2006.
- [18] J. Bullier. Integrated model of visual processing. *Brain Res. Reviews*, 36:96–107, 2001.
- [19] E. Castet, V. Charton, and A. Dufour. The extrinsic/intrinsic classification of two-dimensional motion signals with barber-pole stimuli. *Vision Research*, 39(5):915–932, 1999.
- [20] E. Castet, J. Lorenceau, M. Shiffrar, and C. Bonnet. Perceived speed of moving lines depends on orientation, length, speed and luminance. *Vision Research*, 33:1921–1921, 1993.
- [21] J. Chey, S. Grossberg, and E. Mingolla. Neural dynamics of motion processing and speed discrimination. *Vision Res.*, 38:2769–2786, 1997.
- [22] P. Dayan and L. F. Abbott. *Theoretical Neuroscience : Computational and Mathematical Modeling of Neural Systems*. MIT Press, 2001.
- [23] K.G. Derpanis and J.M. Gryn. Three-dimensional nth derivative of gaussian separable steerable filters. In *IEEE International Conference on Image Processing*, volume 3, pages 553–556, sep 2005.
- [24] C.L. Fennema and W.B. Thompson. Velocity determination in scenes containing several moving objects. *Computer Graphics and Image Processing*, 9(4):301–315, 1979.

-
- [25] I. Fine, D.I.A. MacLeod, and G.M. Boynton. Surface segmentation based on the luminance and color statistics of natural scenes. *Journal of the Optical Society of America A*, 20(7):1283–1291, 2003.
- [26] R.A. Frazor and W.S. Geisler. Local luminance and contrast in natural images. *Vision Research*, 46(10):1585–1598, 2006.
- [27] W.T. Freeman and E.H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906, 1991.
- [28] W.S. Geisler, D.G. Albrecht, and A.M. Crane. Responses of neurons in primary visual cortex to transient changes in local contrast and luminance. *Journal of Neuroscience*, 27(19):5063, 2007.
- [29] S. Grossberg and E. Mingolla. Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychological review*, 92(2):173–211, 1985.
- [30] S. Grossberg, E. Mingolla, and L. Viswanathan. Neural dynamics of motion integration and segmentation within and across apertures. *Vision Research*, 41(19):2521–2553, 2001.
- [31] David J. Heeger. Optical flow using spatiotemporal filters. *The International Journal of Computer Vision*, 1(4):279–302, January 1988.
- [32] B.K. Horn and B.G. Schunk. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.
- [33] X. Huang, T. D. Albright, and G. R. Stoner. Adaptive surround modulation in cortical area MT. *Neuron*, 53:761–770, March 2007.
- [34] D.H. Hubel and T.N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat visual cortex. *J Physiol*, 160:106–154, 1962.
- [35] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001.
- [36] M. Kinoshita and H. Komatsu. Neural representation of the luminance and brightness of a uniform surface in the macaque primary visual cortex. *Journal of Neurophysiology*, 86(5):2559–2570, 2001.
- [37] T.L. Kooi. Local direction of edge motion causes and abolishes the barberpole illusion. *Vision research(Oxford)*, 33(16):2347–2351, 1993.
- [38] Victor A.F. Lamme and Pieter R. Roelfsema. The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci*, 23:571–579, 2000.

-
- [39] G. Löffler and H.S. Orbach. Computing feature motion without feature detectors: A model for terminator motion without end-stopped cells. *Vision Research*, 39(4):859–871, 1998.
- [40] J. Lorenceau and D. Alais. Form constraints in motion binding. *Nature Neuroscience*, 4:745–751, 2001.
- [41] J. Lorenceau, M. Shiffrar, N. Wells, and E. Castet. Different motion sensitive units are involved in recovering the direction of moving lines. *Vision Research*, 33:1207–1207, 1993.
- [42] V. Mante, R. A. Frazor, V. Bonin, W. S. Geisler, and M. Carandini. Independence of luminance and contrast in natural scenes and in the early visual system. *Nature Neuroscience*, 2005.
- [43] G.S. Masson, Y. Rybarczyk, E. Castet, and D.R. Mestre. Temporal dynamics of motion integration for the initiation of tracking responses at ultra-short latencies. *Visual Neuroscience*, 17(5):754–767, 2000.
- [44] G.S. Masson and L.S. Stone. From following edges to pursuing objects. *Journal of neurophysiology*, 88(5):2869–2873, 2002.
- [45] A. Montagnini, P. Mamassian, L. Perrinet, E. Castet, and G.S. Masson. Bayesian modeling of dynamic motion integration. *Journal of Physiology-Paris*, 101(1-3):64–77, 2007.
- [46] A. Montagnini, M. Spering, and G. S. Masson. Predicting 2d target velocity cannot help 2d motion integration for smooth pursuit initiation. *J Neurophysiol*, 96:3545–3550, 2006.
- [47] JA Movshon, EH Adelson, MS Gizzi, and WT Newsome. The analysis of moving visual patterns. *Experimental Brain Research*, 11:117–151, 1986.
- [48] H-H. Nagel. On the estimation of optical flow: relations between different approaches and some new results. *Artificial Intelligence Journal*, 33:299–324, 1987.
- [49] H.H. Nagel and W. Enkelmann. An investigation of smoothness constraint for the estimation of displacement vector fields from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:565–593, 1986.
- [50] Steven J. Nowlan and Terrence J. Sejnowski. Filter selection model for motion segmentation and velocity integration. *J. Opt. Soc. Am. A*, 11(12):3177–3199, 1994.
- [51] M. Otte and H.H. Nagel. Optical flow estimation: Advances and comparisons. In Jan-Olof Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision*, volume 800 of *Lecture Notes in Computer Science*, pages 51–70. Springer-Verlag, 1994.

-
- [52] C. C. Pack, J. N. Hunter, and R. T. Born. Contrast dependence of suppressive influences in cortical area MT of alert macaque. *Journal of Neurophysiology*, 93(3):1809–1815, 2005.
- [53] C.C. Pack and R.T. Born. Integration of motion signals over regions of uniform luminance by mt neurons in the alert macaque. *Journal of Vision*, 2(7):412a, nov 2002. Abstract.
- [54] Christopher Pack and Richard Born. Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature*, 409:1040–1042, feb 2001.
- [55] Christopher Pack, Andrew Gartland, and Richard Born. Integration of contour and terminator signals in visual area MT of alert macaque. *The Journal of Neuroscience*, 24(13):3268–3280, 2004.
- [56] X. Peng and D.C. Van Essen. Peaked encoding of relative luminance in macaque areas v1 and v2. *Journal of Neurophysiology*, 93(3):1620–1632, 2005.
- [57] R.P. Power and B. Moulden. Spatial gating effects on judged motion of gratings in apertures. *Perception*, 21:449–449, 1992.
- [58] W. Reichardt. Autokorrelationsauswertung als funktionsprinzip des zentralnervensystems. *Zeitschrift fur Naturforschung*, 12:447–457, 1957.
- [59] A.F. Rossi, C.D. Rittenhouse, and M.A. Paradiso. The representation of brightness in primary visual cortex. *Science*, 273(5278):1104, 1996.
- [60] N.C. Rust, V. Mante, E.P. Simoncelli, and J.A. Movshon. How mt cells analyze the motion of visual patterns. *Nature Neuroscience*, 9:1421–1431, 2006.
- [61] M. Shiffrar, X. Li, and J. Lorenceau. Motion integration across differing image features. *Vision Research*, 35(15):2137–2146, 1995.
- [62] Adam M. Sillito, Javier Cudeiro, and Helen E. Jones. Always returning: feedback and sensory processing in visual cortex and thalamus. *TRENDS in Neurosciences*, 29(6):307–316, jun 2006.
- [63] E. P. Simoncelli and D.J. Heeger. A model of neuronal responses in visual area MT. *Vision Research*, 38:743–761, 1998.
- [64] E.P. Simoncelli and B.A. Olshausen. Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24(1):1193–1216, 2001.
- [65] Matthew Smith, Najib Majaj, and Anthony Movshon. Dynamics of motion signaling by neurons in macaque area mt. *Nature Neuroscience*, 8(2):220–228, feb 2005.
- [66] C. Stiller and J. Konrad. Estimating motion in image sequences. *Signal Processing Magazine, IEEE*, 16(4):70–91, 1999.

-
- [67] T. Tani, I. Yokoi, M. Ito, S. Tanaka, and H. Komatsu. Functional organization of the cat visual cortex in relation to the representation of a uniform surface. *Journal of Neurophysiology*, 89(2):1112–1125, 2003.
- [68] J.P.H. van Santen and G. Sperling. Elaborated reichardt detectors. *Journal of the Optical Society of America A*, 2(2):300–320, 1985.
- [69] T. Viéville, S. Chemla, and P. Kornprobst. How do high-level specifications of the brain relate to variational approaches? *Journal of Physiology - Paris*, 101(1-3):118–135, 2007.
- [70] J.M. Wallace, L.S. Stone, and G.S. Masson. Object motion computation for the initiation of smooth pursuit eye movements in humans. *Journal of Neurophysiology*, 93(4):2279–2293, 2005.
- [71] H. Wallach. Über visuell wahrgenommene Bewegungsrichtung. *Psychological Research*, 20(1):325–380, 1935.
- [72] A.B. Watson and A.J. Ahumada. Model of human visual-motion sensing. *J Opt Soc Am A*, 2(2):322–342, February 1985.
- [73] Y. Weiss and E. H. Adelson. Slow and smooth: A Bayesian theory for the combination of local motion signals in human vision. *Center for Biological and Computational Learning Paper*, 1998.
- [74] Yair Weiss and David J. Fleet. Velocity likelihoods in biological and machine vision. In *Probabilistic Models of the Brain: Perception and Neural Function*, pages 81–100. MIT Press, 2001.
- [75] HR Wilson, VP Ferrera, and C Yo. A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience*, 9(1):79–97, jul 1992.
- [76] Jiangjian Xiao, Hui Cheng, Harpreet Sawhney, Cen Rao, and Michael Isnardi. Bilateral filtering-based optical flow estimation with occlusion detection. In *Proceedings of the 9th European Conference on Computer Vision*, Lecture Notes in Computer Science. Springer-Verlag, 2006.
- [77] C. Yo and H.R. Wilson. Perceived direction of moving two-dimensional patterns depends on duration, contrast and eccentricity. *Vision Res*, 32(1):135–47, 1992.
- [78] A.L. Yuille and N.M. Grzywacz. A winner-take-all mechanism based on presynaptic inhibition feedback. *Neural Computation*, 1(3):334–347, 1989.



Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)

Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
<http://www.inria.fr>
ISSN 0249-6399