

# Modelling the dynamics of motion integration with a new luminance-gated diffusion mechanism

Émilien Tlapale, Guillaume S. Masson, Pierre Kornprobst

► **To cite this version:**

Émilien Tlapale, Guillaume S. Masson, Pierre Kornprobst. Modelling the dynamics of motion integration with a new luminance-gated diffusion mechanism. [Research Report] RR-6944, INRIA. 2010, pp.42. inria-00360277v3

**HAL Id: inria-00360277**

**<https://hal.inria.fr/inria-00360277v3>**

Submitted on 3 Mar 2010

**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

*Modelling the dynamics of motion integration with a  
new luminance-gated diffusion mechanism*

Émilien Tlapale — Guillaume S. Masson — Pierre Kornprobst

**N° 6944 — version 3**

initial version February 2009 — revised version March 2010

. Computational Medicine and Neurosciences .

A large, light gray, stylized 'R' logo is positioned to the left of the text 'Rapport de recherche'.

*Rapport  
de recherche*



## Modelling the dynamics of motion integration with a new luminance-gated diffusion mechanism

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

Theme : Computational Medicine and Neurosciences  
Équipe-Projet NeuroMathComp

Rapport de recherche n° 6944 — version 3 — initial version February 2009 —  
revised version March 2010 — 39 pages

**Abstract:** The dynamics of motion integration show striking similarities when observed at neuronal, psychophysical, and oculomotor levels. Based on the inter-relation and complementary insights given by those dynamics, our goal was to test how basic mechanisms of dynamical cortical processing can be incorporated in a dynamical model to solve several aspects of 2D motion integration and segmentation. Our model is inspired by the hierarchical processing stages of the primate visual cortex: we describe the interactions between several layers processing local motion and form information through feedforward, feedback, and inhibitive lateral connections. Also, following perceptual studies concerning contour integration and physiological studies of receptive fields, we postulate that motion estimation takes advantage of another low level cue, which is luminance smoothness along edges or surfaces, in order to gate recurrent motion diffusion. With such a model, we successfully reproduced the temporal dynamics of motion integration on a wide range of simple motion stimuli: line segments, rotating ellipses, plaids, and barber poles. Furthermore, we showed that the proposed computational rule of luminance-gated diffusion of motion information is sufficient to explain a large set of contextual modulations of motion integration and segmentation in more elaborated stimuli such as chopstick illusions, simulated aperture problems, or rotating diamonds. As a whole, in this paper we proposed a new basal luminance-driven motion integration mechanism as an alternative to less parsimonious models, we carefully investigated the dynamics of motion integration, and we established a distinction between simple and complex stimuli according to the kind of information required to solve their ambiguities.

**Key-words:** recurrent cortical model, 2D motion integration, luminance, motion perception, temporal dynamics

\* Emilien.Tlapale@sophia.inria.fr

† guillaume.masson@incm.cnrs-mrs.fr

‡ Pierre.Kornprobst@sophia.inria.fr

## Modélisation de la dynamique d'intégration du mouvement par un nouveau mécanisme de diffusion modulé par la forme

**Résumé :** Les dynamiques d'intégration du mouvement montrent de frappantes similarités lorsqu'elles sont observées aux niveaux neuronaux, psychophysique, et oculomoteur. Partant des relations entre ces dynamiques, ainsi que des vues complémentaires qu'elles nous offrent, notre but est de tester comment des mécanismes basiques de traitement cortical peuvent être combinés dans un modèle dynamique capable de résoudre différents aspects de l'intégration et de segmentation du mouvement 2D. Notre modèle est inspiré des niveaux de traitement hiérarchiques du cortex visuel du primate : nous décrivons ainsi les interactions entre différentes couches de traitement du mouvement et de la forme à travers des connexions directes, de rétroaction, ainsi que des connexions latérales inhibitives. De plus, à partir de différentes études concernant l'intégration des contours et des études physiologiques sur les champs récepteurs, nous formulons l'hypothèse selon laquelle l'estimation du mouvement utilise une autre indication de bas niveau, la similarité de luminance sur les bords ou dans les surfaces, afin de moduler la diffusion du mouvement. À partir d'un tel modèle, nous avons pu reproduire avec succès les dynamiques temporelles d'intégration du mouvement sur un grande variété de stimuli en mouvement : des lignes en translation, des ellipses en rotations, des grilles, ainsi que des enseignes de barbier. De plus, nous montrons que la règle de calcul proposée, qui module la diffusion du mouvement par une information de luminance, est suffisante pour expliquer un grand ensemble de modulations contextuelles d'intégration et de segmentation du mouvement sur des stimuli plus complexes tels que les illusions de baguettes, les problèmes d'ouvertures simulés, ou les diamants en rotation. En définitive, dans ce papier nous proposons un mécanisme d'intégration du mouvement modulé par la forme en tant qu'alternative à des modèles moins parcimonieux, nous analysons avec attention les dynamiques d'intégration du mouvement, et nous établissons une distinction entre des stimuli simples et complexes, en fonction des informations requises pour résoudre leurs ambiguïtés.

**Mots-clés :** modèle cortical rétroactif, intégration du mouvement 2D, luminance, perception du mouvement, dynamiques temporelles

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Description of the model</b>	<b>6</b>
2.1	Model rationale . . . . .	6
2.2	Model overview . . . . .	7
2.3	Local motion estimation . . . . .	8
2.4	General connectivity . . . . .	9
2.5	Form modulated diffusion . . . . .	10
2.6	Read-out for comparing model performance with biological data	10
2.7	Implementation details . . . . .	12
<b>3</b>	<b>Experimental results</b>	<b>13</b>
3.1	Dynamics of motion integration for line-drawing objects . . . . .	13
3.2	Dynamics of pattern motion using plaids . . . . .	17
3.3	Motion integration on gratings with different apertures . . . . .	19
3.4	Influence of form on selective motion integration . . . . .	21
<b>4</b>	<b>Discussion</b>	<b>25</b>
4.1	Temporal dynamics of motion integration . . . . .	26
4.2	Luminance smoothness: a simple rule for gating motion information	27
4.3	Modelling the neural dynamics of motion integration . . . . .	29
4.4	Limitations of the model . . . . .	31

## 1 Introduction

Natural scenes present many sources of ambiguities, which must be solved in order to extract reliable information that can be used to control behaviour. Correctly integrating different local features is a key point to solve these ambiguities and therefore, understanding its neural dynamics is an important goal of visual neurosciences. Motion processing has offered a powerful framework to investigate it at many levels (Lorenceanu, 2010). Indeed, in order to compute the global motion of an object embedded in a complex surrounding, artificial motion processing systems as well as visual cortex, take local motion estimates as input. As a consequence, they must deal with numerous 1D features corresponding to edges and, generally fewer, 2D features such as corners or line-endings for example. One computational 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 are compatible with the physical translation of the object containing that edge. As pointed out by Wallach (1935), 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. After several decades of intensive research on 2D motion perception and its neural substrates (see Masson & Ilg (2010) for a collection of reviews), it is still highly controversial whether or not, and how, the brain uses these different types of local motion cues to recover the global motion of the surface of interest (see Bradley & Goyal (2008) for a recent review).

Several different computational rules for motion integration have been proposed. Geometrical solutions such as the Intersection of Constraints (IOC) can recover the exact global velocity vector from the different edges motions (Fennema & Thompson, 1979). Several studies have proposed that the primate visual system use a similar computation (Adelson & Movshon, 1982). It remains however unclear how the visual system can implement the IOC rule. Moreover, the fact that perceived direction does not always correspond, at least for short stimulus durations (Yo & Wilson, 1992), to the IOC solution has supported alternative models that emphasise the role of local 2D features (Löffler & Orbach, 1998) or second-order motion cues (Wilson et al., 1992) when computing global motion.

While the computational rules actually used by the brain are still highly disputed, there are numerous physiological evidences that cortical area V1 implements local motion computation and feeds an integrative stage such as area MT (Born & Bradley, 2005). In macaque area MT, neurons solving the aperture problem (i.e. responding 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 (Movshon et al., 1985; Pack & Born, 2001; Pack et al., 2004; Smith et al., 2005). This property contrasts with the findings that V1 neurons mostly respond to the direction orthogonal to the orientation of the edge drifting across their receptive field (Movshon et al., 1985), albeit some neurons seems to act as local features detectors such as end-stopped cells (Hubel & Wiesel, 1962; Pack et al., 2003). 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 by most of the two-stage feedforward models (Movshon et al., 1985; Wilson et al., 1992; Simoncelli & Heeger, 1998; Rust et al., 2006).

The first aspect often ignored by existing models is that motion integration is intrinsically a spatial process. Since most of the natural objects are rigid, propagating non-ambiguous motion information along edges as well as inside surfaces is an essential aspect of motion integration (Hildreth, 1983b; Nakayama & Silverman, 1988; Grzywacz & Yuille, 1991; Weiss & Adelson, 2000). The role of such diffusion process has only been investigated in a small number of biologically-inspired models. Grossberg and colleagues have investigated how local form and motion cues can be integrated through recurrent diffusion (Grossberg & Mingolla, 1985; Berzhanskaya et al., 2007). The various versions of their model succeed to solve the aperture problem in many different instances of motion stimuli investigated psychophysically (Lorenceanu & Shiffrar, 1992; Castet et al., 1993). However, they heavily rely on many different sub-types of local feature detectors. A similar solution was also developed by Bayerl & Neumann (2005, 2007), albeit with a more simple and realistic motion computation algorithm. Still, the strategy used for more complex stimuli relies on finding local 2D features and excluding some of them (for instance T-junctions) from the integration process. Such computational rules have not yet been demonstrated in the cortical processing of 2D moving patterns. Here, we propose a dynamical model providing a simple solution for 2D motion integration by using a minimalist set of biological properties such as recurrent connectivity between layers working at different scales and the combination of low-level cues about visual surfaces properties such as luminance smoothness and local features motion. Instead

of implementing a set of highly selective feature/shape analysers, our approach favours an abstract representation of form information, based on luminance smoothness in the image. Such an abstract description might fuse both contour and surface representations, which have been found in cortical areas V1 and V2 (Rossi et al., 1996; Kinoshita & Komatsu, 2001; Tani et al., 2003). It also offers a simple solution for the edge versus surface (or global) smoothness constraints used by different models of motion integration (see Weiss & Adelson (2000) for a review). We propose that both representations contribute in the gating of motion information diffusion in order to solve the aperture problem both within and across apertures.

The second aspect often ignored (see for instance (Weiss & Fleet, 2001; Rust et al., 2006)) is that biological computation of global motion is highly dynamical. When presented with simple lines, plaids or barber poles, the perceived direction reported by human observers will shift over time. For example it was shown by Castet et al. (1993); Shiffrar & Lorenceau (1996) that initial perception is strongly biased towards the direction orthogonal to the edge orientation (or a vector average solution when several edges are available). This initial bias reflects the strong influence of 1D motions in the earliest glimpse. However, some 200ms after stimulus motion onset, perceived direction matches the true translation of the object. Similar dynamics have been found with other types of 2D moving patterns such as plaids for instance (Yo & Wilson, 1992). Similar temporal dynamics has been found with smooth pursuit eye movements: initial direction followed the orthogonal motion whereas later tracking direction matched the target trajectory (Masson & Stone, 2002; Wallace et al., 2005). Such dynamics can reflect the dynamical neural solution to the aperture problem. Over a time course of several tens of milliseconds, area MT neurons solve the aperture problem, so that late but not early preferred direction corresponds to pattern motion direction (Pack & Born, 2001; Pack et al., 2004; Smith et al., 2005). This is our goal to reproduce such dynamics. Moreover, our model suggests that the dynamics of spatial integration and the time course of motion perception can be intrinsically linked, without the need of postulating fixed delay in local form processing such as end-stopping or feature extraction (Wilson et al., 1992).

This paper is organised as follows. In Section 2 we describe our model that consists in a reduced set of layers that are recurrently connected. Mathematically, our model is a set of nonlinear coupled integro-differential equations. We implement a simple computational rule for motion integration and segmentation: motion diffusion is gated by the local luminance profile. In Section 3 we document the behaviour of this model against a large class of synthetic motion stimuli that have been selected to reproduce key aspects of primate motion processing. The dynamics of the model was compared to that observed at different levels: population of direction selective cells in area MT, ocular tracking behaviour and human psychophysics. By doing so, we aimed at demonstrating the generic aspect of our computational solution and linking these different dynamics observed across multiple scales. Finally, in Section 4 we discuss the model, its plausibility, and its limitations.



## 2 Description of the model

### 2.1 Model rationale

Our goal was to test how several basic mechanisms of cortical processing can be implemented in a dynamical model to solve several aspects of 2D motion integration and segmentation. We avoided implementing any type of specific local feature detectors such as line terminators, corners, T-junctions and so on. For this purpose, we postulated that both local 2D motion computation and global 2D motion integration or segmentation are dynamically solved using only low-level image features such as local motion and luminance orientation. To do so, our model relies on two main characteristics which are biologically-grounded.

The first characteristic is that motion information is extracted and processed at different spatial scales within layers that are recurrently interconnected. As illustrated in Figure 1, our model implements three layers of motion processing. The first layer extracts local motion energy through spatio-temporal filtering, corresponding to simple and complex cells of the primary visual cortex (Simoncelli & Heeger, 1998). They form the input to a second layer which computes local direction and speed of motion. Some complex cells in primary visual cortex have been shown to perform such local velocity computation (Priebe et al., 2006). The third layer implements MT neurons, which integrate motion over larger portions of the image through the convergence of L2 cells (cortical layer 2). Our MT-like L3 neurons have larger receptive fields and are tuned for lower spatial frequencies and higher speed than striate-like L2 cells. This fact is consistent with the view that V1 and MT stages operate at different scales (see Born & Bradley (2005) for a review). Feed-forward models of motion integration are heavily rooted on such evidence (Simoncelli & Heeger, 1998; Rust et al., 2006; Wilson et al., 1992; Löffler & Orbach, 1998) and we will compare our results to one of them. However, V1 and MT areas are recurrently interconnected (see Sillito et al. (2006) for a review). Existing models have shown that such recurrent connectivity can play a role in solving the aperture problem in synthetic and natural sequences (Chey et al., 1997; Bayerl & Neumann, 2004, 2005) as well as implementing contextual effects observed in V1 and MT neurons (see Angelucci & Bullier (2003) for a review). One main innovative aspect is propose a true dynamical model: we do not run our process only when a frame arrives but instead define a continuous model working in continuous time, so that we obtain dynamics measurements allowing comparison with neural and behavioural time courses.

The second characteristic is that our model postulates that the brain takes advantage of another low level cue, luminance smoothness along edges or surfaces, to gate recurrent motion diffusion. Thus, contrary to previous recurrent models of motion integration using isotropic diffusion, our model dynamically constrains the diffusion of motion information along some specific orientation in the image. Indeed, perceptual studies of contour integration and physiological studies of receptive field surround effect in cortical layer 2/3 neurons provide evidence for facilitatory effects that are much stronger in regions of visual space that lie along the axis of preferred orientation than in region off axis (Nelson & Frost, 1985; Field et al., 1993; Polat & Sagi, 1993; Kapadia et al., 1995). There are evidence for involving both lateral connections (Bosking et al., 1997; Stettler et al., 2002) and recurrent input (Lee & Nguyen, 2001; Lee & Mumford,

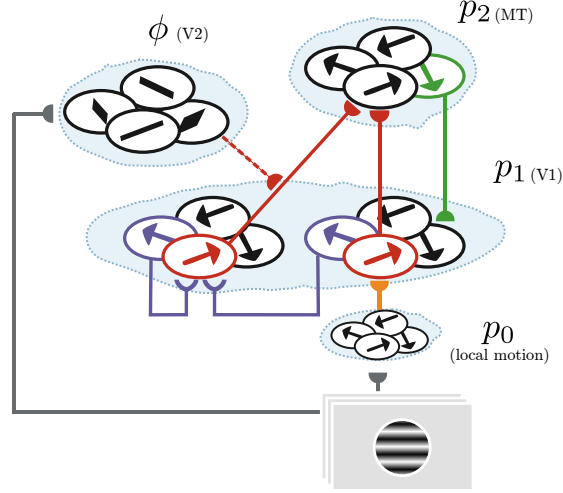


Figure 1: A schematic view of the model showing the interactions of the different cortical layers. The motion integration (layers  $p_0$ ,  $p_1$ ,  $p_2$ ) system is gated by the luminance (layer  $\phi$ ), the gating being represented by the *dashed* line.

2003) from higher computational stages in these non-isotropic interactions. Our goal herein was not to model the detailed connectivity (albeit this might have a profound impact of the exact temporal dynamics) but rather to explore how such luminance-gated motion diffusion can be useful in a large class of object motion integration and segmentation.

## 2.2 Model overview

Our model is represented in Figure 1. It implements the interactions between several layers processing local motion and form information (i.e. static luminance distribution). The model estimates dynamically the velocity information given an input grey level image sequence denoted by:

$$I : (t, x) \in \mathbb{R}^+ \times \Omega \rightarrow I(t, x) \in [0, 1], \quad (1)$$

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

Then 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. Here we will define two types of layers. The first type of layers is related to motion and their activity is denoted by:

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

where  $\mathcal{V}$  represents the space of possible velocities. Each function  $p_i$  can be interpreted as the state of a cortical area retinotopically organised 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 estimation 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 1).

The second type of layer is related to form. Its activity is denoted by:

$$\phi(t, x, \theta) \in \mathbb{R}^+ \times \Omega \times [0, 2\pi) \rightarrow \mathbb{R}^+. \quad (3)$$

It represents the local orientation of the luminance profile from position  $x$  in the direction  $\theta$ . Note that function  $\phi$  is an abstract way to encode form information. Such function can be seen as a description of V2 neuron properties which can represent local orientation of edges from changes in luminance (see Lennie and Movshon, 2005 for a review) but also can encode surface brightness (see Paradiso et al. (2006) for a review). In future development of the model, such function can also be extended to form information extracted from other cues such as colour, texture and so on.

The coupling between layers, as illustrated in Figure 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 the work by Chey et al. (1997); Bayerl & Neumann (2004); Berzhanskaya et al. (2007). Feedforward connections transmit information from layers closer to the eye to layers deeper in the system while feedback connections connect back to the areas closer to the eye. Lateral connections are inhibitory and provide each neuron with an input from its neighbourhood. The following paragraphs give more details on the different layers and their connections.

### 2.3 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 different degrees of biological plausibility (Reichardt, 1957; Van Santen & Sperling, 1985; Watson & Ahumada, 1985; Adelson & Bergen, 1985). Here, we define the input motion detectors,  $p_0$ , using a motion energy model which is an efficient way to extract local motion with spatio-temporal filtering kernels corresponding to neuronal receptive fields (Heeger, 1988; Simoncelli & Heeger, 1998; Rust et al., 2006). The choice of filtering has two main advantages over simpler correlations techniques: First, spatio-temporal filters can handle a larger class of input stimuli due to their wider frequency tuning. Second, fast techniques can be used to estimate local motion due to the properties of steerability and separability properties of certain energy filters (Freeman & Adelson, 1991; Simoncelli & Heeger, 1998; Derpanis & Gryn, 2005). In addition, mechanisms to combine the output of such filters have been largely studied. For instance, the *donut mechanism* is described and studied in (Simoncelli & Heeger, 1998) and Alexiadis & Sergiadis (2008).

More precisely, our local motion input is based on an energy model computed from the filters by Derpanis & Gryn (2005), namely the second derivative of a Gaussian and its Hilbert transform. Thanks to the property of those filters it is easy to steer them to any other orientation using an interpolation mechanism. We combined the output of those filters using the approach presented in Alexiadis & Sergiadis (2008). This choice is motivated by the well-defined theoretical framework that the authors developed for basis filter combination,

as well as the easiness to apply these filters. Briefly, the expression of the filter response is given by:

$$f^r(t, x, v) = \sum_{n=0}^N \left( \sum_{m=1}^M t_m^r(s_n^r(v))(y_m^r * I)(t, x, v) \right)^2, \quad r \in o, e, \quad (4)$$

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

Then, based on the expression (4), we defined the activity (energy) of our first layer  $p_0$  by:

$$p_0(t, x, v) = f^o(t, x, v) + f^e(t, x, v). \quad (5)$$

## 2.4 General connectivity

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

$$\frac{\partial p_1}{\partial t} = -\lambda_1 p_1 + S_1 \left( \lambda_1^f p_0 + \lambda^b p_0 p_2 - \lambda_1^l G_{\sigma_1^l} * \int_{\mathcal{V}} p_1(t, x, w) dw \right), \quad (6)$$

$$\frac{\partial p_2}{\partial t} = -\lambda_2 p_2 + S_2 \left( \lambda_2^f \int_{\Omega} K(t, x, y) p_1(t, x, y) dy - \lambda_2^l G_{\sigma_2^l} * \int_{\mathcal{V}} p_2(t, x, w) dw \right), \quad (7)$$

where  $G_\sigma$  is a Gaussian function of variance  $\sigma$ ,  $\lambda$ 's and  $\sigma$ 's are constants, and  $S_i(u) = (1 - p_i) \max(0, u)$ , and  $K$  is defined by:

$$K(t, x, y) = G_{\sigma_2^f}(|x - y|) \phi(t, x, \widehat{xy}), \quad (8)$$

where  $\widehat{xy}$  denotes the angle between the vector  $\overrightarrow{xy}$  and the horizontal axis, and  $|\cdot|$  is the norm operator.

The three main characteristics of our model (6–8) are summarised as follows:

- **Luminance-gated diffusion**, which is the main novelty of our model. Rather than diffusing motion information isotropically from  $p_1$  to  $p_2$  (7) in order to model wider receptive fields at the integration stage (Bayerl & Neumann, 2004), we defined an anisotropic diffusion depending on local form information. Since this aspect is essential in our model, we will describe it in more details in the next section.
- **Feedback**, from  $p_2$  to  $p_1$ , which is modulated by  $\lambda_b p_o$  (6) in a multiplicative way as in Bayerl & Neumann (2004). Therefore we used a modulating rather than driving feedback, similar to that found in studies of the motion processing system in primates (Sillito et al., 2006).

- **Lateral inhibition**, which is modelled by the terms  $-\lambda G_\sigma * \int p(t, x, w)$  for both layers  $p_1$  and  $p_2$ . All neurons at a given local neighbourhood and for all possible velocities inhibit each other. Such short-range lateral inhibition, usually called recurrent inhibition, leads to a winner-take-all mechanism (Dayan & Abbott, 2001; Yuille & Grzywacz, 1989). Instead of the divisive inhibition as found in some models (Nowlan & Sejnowski, 1994; Bayerl & Neumann, 2004), we implemented a subtractive inhibition. However, note that divisive inhibition can be viewed as the steady-state solution of a dynamical system using subtraction as inhibitory mechanism.

## 2.5 Form modulated diffusion

In order to estimate  $p_2$ ,  $p_1$  is integrated in a spatial neighbourhood using the weight  $K(t, x, y)$ , defined in (8). This weight is composed of two terms. The first term,  $G_{\sigma_r}(|x - y|)$ , weights the connectivity depending on the distance between  $x$  and  $y$ . The second term,  $\phi(t, x, \widehat{xy})$ , is related to the form information. For example, if we want to express an isotropic integration, not depending on the luminance, then we can simply define:

$$\phi(t, x, \theta) = 1, \quad \theta \in [0, 2\pi[. \quad (9)$$

Instead, in this paper, we propose that the integration depends on the form so that the layer  $\phi$  is defined by:

$$\phi(t, x, \theta) = \int_{\Omega} G_{\sigma_x}(x - z) G_{\sigma_\theta}(\theta - \widehat{xz}) G_{\sigma_s}(I(x) - I(z)) dz, \quad \theta \in [0, 2\pi[. \quad (10)$$

This layer  $\phi$  describes the luminance smoothness at position  $x$  and along the direction  $\theta$ . In (10), the term  $G_{\sigma_x}(x - z) G_{\sigma_\theta}(\theta - \widehat{xz})$  defines an oriented spatial neighbourhood around  $x$  (see Figure 2a). The last term, namely  $G_{\sigma_s}(I(x) - I(z))$ , corresponds to a brightness similarity measure describing form information using luminance as a criterion.

A representation of the layer  $\phi$  for all the directions and for a given set of sampled positions is shown in Figure 2b. The main property of  $\phi$  is to facilitate integration inside similar spatial structures of the image, a property shared by neurons as observed in both psychophysics (e.g. Shiffrar et al. (1995); Lorenceau & Alais (2001)) and cell recordings in macaque area MT (e.g. Huang et al. (2007)). Another interesting property is that the extension of the integration also depends on the local contrast: The neighbourhood becomes wider at low contrast than at high contrast similar to the changes in receptive field size with contrast, as observed for instance in macaque area MT (Pack et al., 2005). Such abstract representation of form information presents several key advantages in the context of 2D motion integration. Motion integration inside spatial structures is not only performed along borders (see Figure 2b), but also propagates inside isoluminant regions.

## 2.6 Read-out for comparing model performance with biological data

Our model estimates a distributed activity response: each function  $p_i$  can be interpreted as the state of a cortical area that is retinotopically organised to

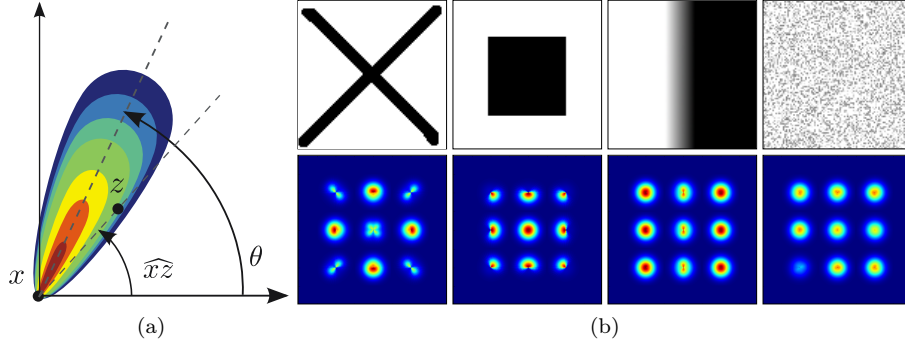


Figure 2: Luminance information. (a) Illustration of the oriented spatial neighbourhood around  $x$  in the direction  $\theta$  used to compute  $\phi$ . Luminance in this oriented neighbourhood is compared with the luminance at the origin  $x$ . (b) Diffusion of information for different spatial structures. Upper row gives a set of input images with different luminance distribution. Lower row shows a representation of  $K$  indicating for a given set of sampled position, the weight by which their neighbourhood is integrated.

provide at each position  $x$  the instantaneous activity of a neuron tuned for the velocity  $v$ , as shown in Figure 3a.

Since such a distributed representation is hard to interpret and analyse, we first define an optical flow like representation. To do so, we average at each position the population response across all velocities, thus obtaining a single vector. Thus, a velocity field  $m_i$  can be extracted from any layer  $p_i$  by:

$$m_i(t, x) = \frac{\sum_{v \in \mathcal{V}} p_i(t, x, v)v}{\sum_{v \in \mathcal{V}} p_i(t, x, v)}, \quad i \in 1, 2. \quad (11)$$

Then, this velocity field can be represented either by arrows (see Figure 3b) or by a colour coded image indicating speed and direction (see Figure 3d). Here, we used the Middlebury colour code (Baker et al., 2007) as illustrated in Figure 3c. This colour code emerged as the *de facto* standard in the optical flow computer vision community and it is motivated by colour perception experiments. The Middlebury colour code associates a single colour to each velocity  $v = (v_x, v_y)$ . The direction of the velocity corresponds to the hue of the velocity, for instance yellow for downward velocities, while the speed of the velocity is encoded in the saturation of the colour, whiter for slower speeds.

Based on this velocity field, another way of interpreting the model output and its dynamic is to define a read-out such as the perceived direction  $w(t) \in \mathbb{R}^2$ . Given  $w(t)$ , one can compare the model performances with the dynamics of biological motion processing gathered at different levels: physiological, psychophysical and behavioural. To do so, we defined a simple read-out from the activity in layer  $p_2$ , by averaging the velocity field over space and at a given time, with a temporal smoothing defined by the following dynamical equation:

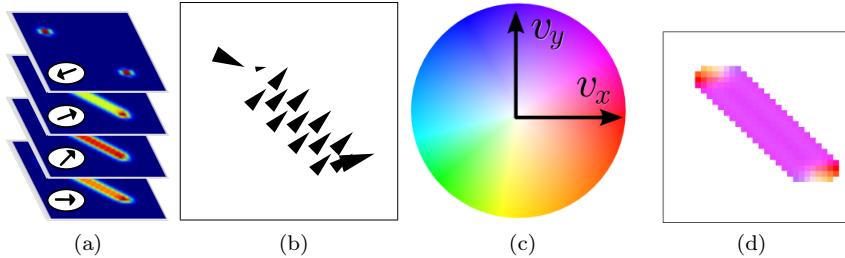


Figure 3: Representation of motion layers. (a) Illustration of the motion representation in the layer  $p_i$  for a translating bar with responses for the different preferred velocities at each location. (b) A sampled velocity field associated to distribution  $p_i$ . (c) The Middlebury colour code (Baker et al., 2007) used for continuous dense output. (d) A colour-based representation of the velocity field.

$$\frac{dw}{dt}(t) = \lambda \left( \sum_{x \in \Omega} m_2(t, x) - w(t) \right), \quad (12)$$

where  $m_2$  is defined by (11). Thanks to the definition of this read-out, we will define in Section 3.1 an estimated direction errors, so that direct comparisons with biological data will be possible.

## 2.7 Implementation details

This section describes the implementation details of our model. As soon as implementation is concerned, time is discrete so that the input grey level sequence is given by a set of images at different times. Here we assume that the images are sampled every 100 msec. The set of possible velocities  $\mathcal{V}$  also needs to be sampled. Herein we chose  $\mathcal{V} = [-3, 3]^2 \in \mathbb{Z}^2$  so that the velocities are sampled in a  $7 \times 7$  pixels grid.

The model defined by equations (6)–(10) was fully specified by a set of 14 parameters. These parameters, whose values are given in Table 1, were found by matching the time scale dynamics of psychophysical experiments. The simple line drawing stimuli were used to fit the parameters that were then kept constant for all other motion stimuli. Note that the isotropic diffusion described in equation (9) was tuned in order to get results similar to the luminance-gated model, at least on the simple line drawing stimuli used during the fitting procedure.

In addition to the time scale matching procedure, we also investigated the role of the parameters. For instance, the  $\lambda_1^i$  and  $\lambda_2^i$  parameters representing the weight of the inhibition are necessary to achieve a *winner-take-all* like mechanism (Yuille & Grzywacz, 1988; Dayan & Abbott, 2001). We evaluated the acceptable range for those inhibition parameters to be between 1.8 and 8.0.

Finally, to speed up the simulations we used the GPGPU technology. Since the anisotropic diffusion process depends on input stimulus, our model requires high computational cost. Thus conventional CPU implementation is too slow for performing extensive model testing. We were able to take advantage of the parallel nature of our model, where the same kind of computation is done at

Equation (6)	$\lambda_1 = 2.0$	$\lambda_1^f = 1.0$	$\lambda_1^b = 24.0$	$\lambda_1^l = 4.0$	$\sigma_1 = 2.0$
Equation (7)	$\lambda_2 = 2.0$	$\lambda_2^f = 16.0$		$\lambda_2^l = 4.0$	$\sigma_2 = 2.0$
Equation (8)	$\sigma_2^f = 8.0$				
Equation (10)	$\sigma_x = 12.0$	$\sigma_\theta = \pi/8$	$\sigma_s = 0.4$		

Table 1: Chosen parameters setting.

every spatial position. In other words, this method and the way it was implemented, allows to process arbitrarily large stimuli, in pixel resolution, which is not the case in recent proposed approaches (see e.g. Berzhanskaya et al. (2007) where the authors consider  $60 \times 60$  binary images).

### 3 Experimental results

Our goal is to demonstrate how a minimal model can qualitatively reproduce a wide set of motion integration and segmentation phenomena as observed at different levels: neuronal, psychophysical and oculomotor behaviour. This multi-level extent is important because the different dynamics are inter-related and give complementary insights about the neuronal solution of the aperture problem and the selective integration process (see Masson & Ilg (2010) for a complete review). Herein we document the performance of our model for a wide range of synthetic motion stimuli already used for investigating brain dynamics of 2D motion integration and segmentation. We qualitatively reproduced the neural dynamics of these phenomena, in particular their time courses. Results were obtained for full-contrast motion stimuli but several simple changes in image geometry were tested, based on previous psychophysical work. In particular, we investigated how a simple luminance-gated diffusion can solve motion integration within and across apertures and how motion integration can be modulated by the contextual organisation of the visual scene without the need for a depth ordering mechanism based on binocular disparity for instance. We systematically compared the model output obtained with or without the form layer  $\phi$ , in order to demonstrate the key role of anisotropic diffusion driven by local luminance information.

The results are organised as follows. First we present results on the dynamics of motion integration obtained with classical simple stimuli made of line drawings. Then we continue our exploration by using different plaid patterns. Next, we describe the effect of the aperture shape on 2D information and its consequence for motion perception. Lastly, we investigate the response of our model on more complex stimuli involving several objects to see how luminance-gating diffusion can solve several aspects of motion integration and segmentation.

#### 3.1 Dynamics of motion integration for line-drawing objects

##### The translating bar stimulus

The dynamics of motion integration and the role of form-based disambiguation mechanisms can be illustrated with the simplest example of the aperture problem in motion perception: a translating bar stimulus as shown in Figure 4a. For



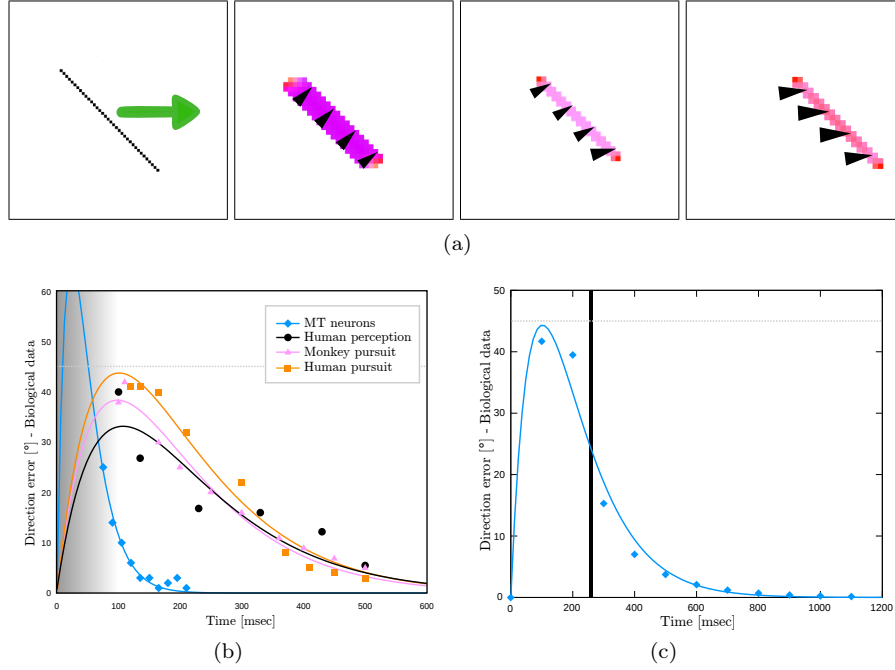


Figure 4: Example of a translating bar: model output and comparison with biological data. (a) Temporal evolution of  $m_1$  for a  $45^\circ$  tilted bar, moving rightward. (b) Temporal dynamics of the observed direction error for MT neuron direction selectivity, human perceived direction, human and macaque tracking direction. Data have been re-plotted from Pack & Born (2001); Lorenceau et al. (1993); Wallace et al. (2005); Born et al. (2006), respectively. Both discrete measurements and best fit are shown. Each data set was fitted with  $f(t) = A \frac{x}{\tau} \exp\left(-\frac{x}{\tau}\right)$ . *Shaded region* corresponds to unreliable data. *Dotted line* corresponds to a  $45^\circ$  error. (c) For comparison, estimated direction error in our model. Dark vertical bar corresponds to the steepest decrease.

short durations, its perceived direction is biased towards the direction orthogonal to its orientation. Such perceptual bias is corrected for longer durations (Lorenceau et al., 1993; Castet et al., 1993; Shiffrar & Lorenceau, 1996). Consistently, it has been 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 msec (Masson & Stone, 2002; Wallace et al., 2005). Similar results were obtained in monkeys by Born and colleagues (Born et al., 2006). Interestingly, when presented with a set of small oriented bars, direction selectivity of MT neurons exhibit the same temporal dynamics: their optimal direction slowly rotating from the component orthogonal to the bars orientation to the 2D motion direction over a 150 msec response period (Pack & Born, 2001).

The *observed direction error* was defined as the difference between the true translation direction of the object and the observed motion direction. Such velocity error has often been used to describe the dynamics of motion integration at these different levels: a population of MT neurons (Born & Bradley, 2005;

Pack & Born, 2001; Pack et al., 2004), the perceived direction (Lorenceanu et al., 1993; Castet et al., 1993; Shiffrar & Lorenceanu, 1996) or the tracking direction of smooth pursuit eye movements (Wallace et al., 2005; Born et al., 2006; Montagnini et al., 2006). A representative set of biological data is illustrated in Figure 4b where time course of direction error is replotted for these different scales: MT population (Pack & Born, 2001), perceived direction in human observers (Castet et al., 1993) and voluntary pursuit in both humans (Montagnini et al., 2006) and monkeys (Born et al., 2006). It should be noted that since the observed direction error is an *angular* error computed from motion, it is highly imprecise during the first dozen of milliseconds. At that period of time, responses are slow, noisy and rapidly varying so that computation of the effective angles becomes unstable. This is particularly true for the dynamics of MT neurons, as direction selectivity vary very rapidly over time. It is for this reason that in Figure 4b a shaded region denotes the initial unreliable period.

Then, in order to compare our results with experimental data, let us define the *estimated direction error*, which is the difference between the angles of the true translation direction and our global read-out  $w(t)$  defined by Equation (12). As illustrated in Figure 4c, applying our model to the translating bar stimulus reproduced several of the phenomena described above. Initial estimation was dominated by local ambiguous (1D) motion measurements, as shown by the velocity field in Figure 4a. We found a smooth 2D motion diffusion inside the bar as shown by the gradual evolution of the velocity fields (Figure 4a, from left to right). Thus our model can solve the aperture problem at both local and global scales. We plotted in Figure 4c the estimated direction error for this translating bar. After a short period of time where the direction error stays constant at about  $40^\circ$ , the estimate of the global motion converged to the true direction (i.e. a null direction error) with an exponential decay. It should be noted that the dynamics we observed at output stage of our model closely mimicked the experimental data measured for both pursuit and perception (Figure 4b).

### Variations of the translating bar

Next, we introduced two variations to test the model behaviour. First, we broke the bar into an increasing number of line segments. The resulting behaviour is shown in Figure 5a. Similarly to what has been found in psychophysical experiments (Lorenceanu et al., 1993), introducing more line-endings both reduced the initial bias in the global motion estimation (from  $44$  to  $32^\circ$ ) and produced a faster exponential decay of the direction error. Wallace et al. (2005) found similar changes when filling a moving diamond with 2D texture elements. On the contrary, smoothing the luminance profile by applying a Gaussian filter along the bar orientation reduced the contrast of line-endings (see Figure 5b) and thus resulted in a larger initial bias, reaching the asymptotic error of  $45^\circ$  and a somewhat longer time constant for error reduction. Similar results were reported with smooth pursuit eye movement in humans (Wall & Danielsson, 1984).

Second, we tested the influence of noise to model output. An additive Gaussian noise pattern, with different variances was introduced to the line motion. The direction error was estimated at a single point in time, corresponding to the steepest decrease in the best fitting function, as illustrated in Figure 4c by a dark vertical bar. Figure 6a plots this instantaneous direction error as a function of the variance of the noise distribution. Similar to the effects of contrast

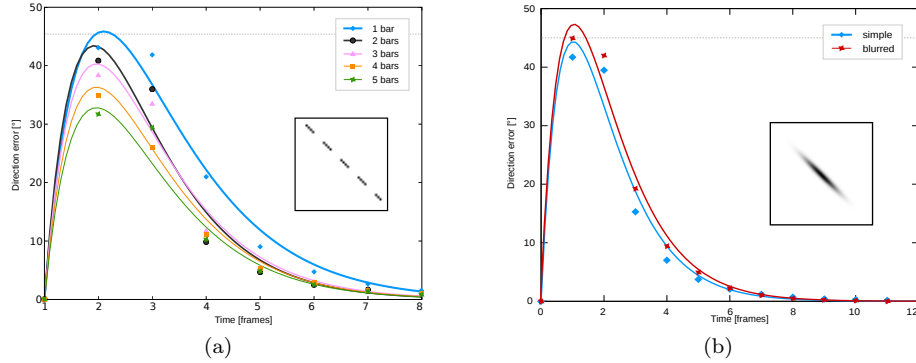


Figure 5: Varying strength of 2D motion cues. (a) A tilted line is broken into small segments, introducing more line-endings. Direction error is plotted against time for different numbers of segments in broken tilted line. (b) A tilted line is filtered with an elongated Gaussian window, which reduces contrast of the line-endings. The smoothed bar elicits larger initial direction error (i.e. larger bias) and a slower time course for computing the exact translation of the bar.

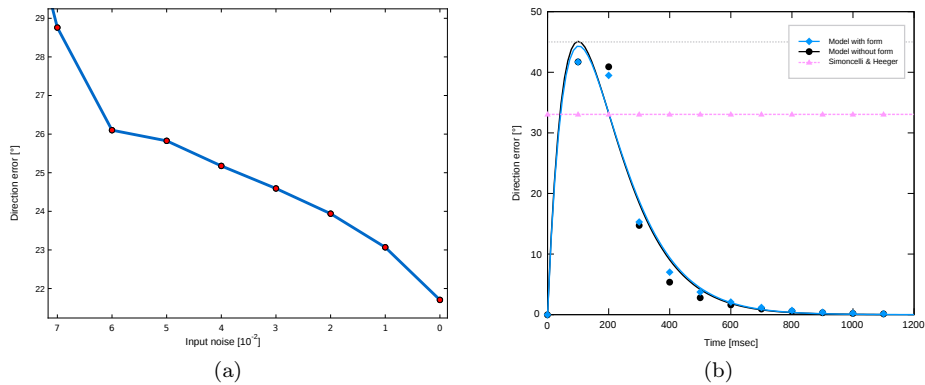


Figure 6: (a) Influence of noise on the dynamics in order to simulate contrast effect. The early response direction error is shown. *Dotted line* corresponds to a  $45^\circ$  error. (b) Comparison of the presented form-modulated model on a simple translating bar with a similar model without form modulation and the model of Simoncelli & Heeger (1998).

which have been observed in both psychophysical studies (Castet et al., 1993) and behavioural studies (Wallace et al., 2005), higher levels of noise resulted in larger initial biases. Moreover, the neural solution of the aperture problem was slower. Similar effects can be observed by changing the input gain of the model.

For those examples, the luminance-gated diffusion has little impact upon the dynamics because only one edge was present and therefore diffusion naturally occurred along it, as illustrated in Figure 6b. We will see that luminance-gated diffusion becomes critical when multiple edges are present and therefore when some solution have to be eliminated. Still our model performs much better than

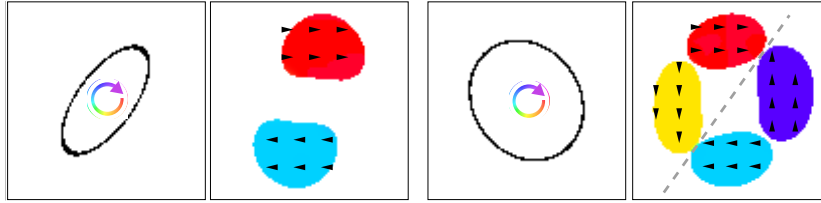


Figure 7: Model response to the gelatinous ellipses. (a) We first process a thin ellipse of ratio 9:20 and the resulting motion (b) is compatible with rotation. (c) We then process a thick ellipse of ration 3:4 and the resulting motion (d) is a deformation incompatible with rotation: *left downward yellow patch* and *violet upward right patch* should be inverted, and not pointing towards the *diagonal line*.

the static model of Simoncelli & Heeger (1998)<sup>1</sup> that we use for comparison. Output of their model is plotted as continuous dotted line in Figure 6b. First, our model predicted a larger initial bias, which is more consistent with psychophysical and behavioural data. Second, thanks to its dynamics, our model can to solve the aperture problem despite the fact that only one 1D edge was present in this simple stimulus, contrarily to the model of Simoncelli & Heeger (1998).

### Rotating ellipses

To conclude this section with line-drawing objects, we want to briefly mention that similar psychophysical observations were made with other types of line-drawing objects (Lorenceau & Shiffrar, 1992; Shiffrar et al., 1995; Masson & Stone, 2002). Our output was always consistent with experimental data, for both initial bias estimate and time course. One interesting example is given by rotating ellipses (Wallach et al., 1956). Weiss & Adelson (2000) investigated motion perception with this type of motion stimuli to probe non-local constraints on models of human motion analysis. The authors showed that narrow and fat ellipses are perceived differently at slow speeds. While “narrow” ellipses are correctly perceived as rigidly rotating, “fat” ones are perceived as deforming non-rigidly with a strong bias towards the directions orthogonal to the long axis of the ellipse. As illustrated in Figure 7, our model reproduces this behaviour as shown by the crude illustration of the velocity flow field. Global motion estimation changed from rotation to expansion with respect to the aspect ratio of the ellipse. With fat ellipse, expansion was found along the long axis of the object. These dynamics were found in absence of the form layer as well.

## 3.2 Dynamics of pattern motion using plaids

Plaid patterns have been largely studied to elucidate 2D motion integration both at psychophysical level (Adelson & Movshon, 1982; Ferrera & Wilson, 1990; Gorea & Lorenceau, 1991; Yo & Wilson, 1992) and physiological level (Movshon et al., 1985; Rodman & Albright, 1989). One interesting aspect of plaid motion is that, depending on the relative direction of the two components, different

<sup>1</sup>Available on-line at <http://www.cns.nyu.edu/~eero/MT-model.php>

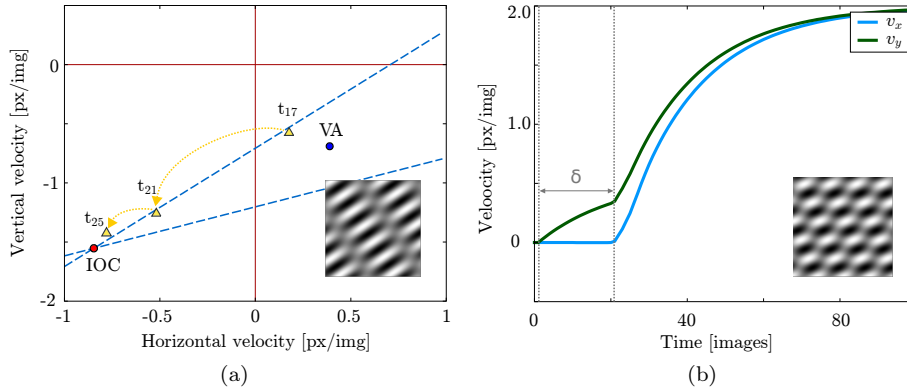


Figure 8: Model responses to plaid pattern motion. (a) Model output obtained with a Type II plaid where the two component directions are separated by  $25^\circ$  (from Bowns (1996)). The temporal dynamics is illustrated by the instantaneous output direction at three different points in time (triangles). In the same plot, the predictions made by the vector average (VA) and intersection-of-constraints (IOC) models are illustrated. (b) Response to an unikinetic plaid as described in Masson & Castet (2002). The initial response following the moving plaid switches with time. Note that we observe a delay  $\delta$  between the vertical and horizontal responses as described in Masson & Castet (2002) for eye movements.

perceived directions can be predicted from the different computational solutions proposed so far: vector averaging (VA), intersection-of-constraints (IOC) or 2D feature tracking (2Dft). Moreover, recent studies showed that direction tuning of pattern-selective cells in area MT shift from components to patterns motion direction over several dozens of milliseconds, further illustrating the fact that solving the aperture problem is a dynamical process (Pack & Born, 2001; Smith et al., 2005). Such neuronal dynamics could explain why perceived direction (Yo & Wilson, 1992) as well as eye tracking direction (Masson & Castet, 2002) shift over time from the vector average prediction to the true pattern motion direction.

Therefore, our model shows a similar dynamics when tested with Type I, Type II (Ferrera & Wilson, 1990) as well as unikinetic plaid patterns (Gorea & Lorenceau, 1991). Figure 8a illustrates the model output in response to a Type II plaid such as used in Bowns (1996). These plaid patterns have been used to separate the predictions made by either the vector average or the IOC models. Initial global estimate of the model output was nearly aligned with the VA prediction. Over time, this estimate gradually shifts toward the IOC prediction, so that at the end of the simulation, the true direction of the plaid pattern is decoded, independently of the component motion direction. Figure 8b illustrates the model performance for another type of plaid. With unikinetic plaids, the IOC solution cannot be applied since only one component is drifting. VA solution collapse to the 1D direction of the drifting component. However, reliable motion information can be extracted by tracking the 2D features (blobs) created at the intersections between the static and drifting gratings. Again, the model output dynamically evolved from the VA solution (i.e. the 1D motion direction)

to the actual pattern motion direction as predicted by the 2Dft model. Interestingly, we found that 2D motion direction was not seen from earliest model output. The influence of motion signals biasing the global estimate towards the global 2D pattern motion was seen only after a fixed delay (indicated by  $\delta$  in Figure 8b), similar to that observed by Masson & Castet (2002) in humans and Barthelemy et al. (2009) in monkeys.

### 3.3 Motion integration on gratings with different apertures

Other aspects of 2D motion signals integration can be investigated with gratings drifting 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 longer axis of the aperture. This phenomenon is known as the barber pole illusion (Wallach, 1935). The bias depends of the aspect ratio, defined by ratio between the long and short axes of the aperture, and increases with it. Moreover, human ocular tracking (Masson et al., 2000b) as well as neuronal responses, gradually evolved from local motion direction (i.e. orthogonal to grating orientation) to global motion direction (i.e. along the aperture long axis) (Pack et al., 2004).

Our model can reproduce these different aspects of motion integration for barber poles (see Figure 9). In all the tested stimuli, a horizontal grating was drifted in the upward direction. Only the shape of the aperture through which the grating was viewed was changed. As illustrated by velocity flow fields obtained at different times, motion flow was first dominated by 1D motion information, but later all local measurements became coherent with the 2D perceived direction. This dynamics can be further illustrated by plotting the time course of the direction error: the estimated global motion was first driven by grating motion direction but then slowly rotated until being aligned with the long axis of the aperture.

This role of local 2D motion cues in driving the final perceived motion was nicely demonstrated by indenting the longer axis of a barber pole (Power & Moulden (1992); Kooi (1993); Masson et al. (2000a) see also Lorenceau & Shiffrar (1992)). Perceived direction changes towards the grating motion direction as the size of the indentation increases. Our model simulated such behaviour. As illustrated in Figure 9b, changing the aperture local geometry introduced new local motion signals, which dominated the global motion direction. As a consequence, global motion remained coherent with the grating motion direction. Note that similar results were also obtained with gratings presented behind a circular aperture (see Figure 9c).

Barber pole motion stimuli with an aspect ratio of 1:1 (i.e. a square aperture) unveil two interesting phenomena. First, short stimulus duration results in a perceived motion direction, as well as a tracking direction that are consistently aligned with grating motion direction across trials (Castet et al., 1999; Masson et al., 2000a). Second, with long motion durations, perceived direction becomes multi-stable, alternating between grating motion direction and motion along one or the other axis of the aperture. Castet et al. (1999) demonstrated stochastic fluctuations in the perceived direction of barber poles with aspect ratio 1:1, yielding to a broad distribution in performance when computed over a large set of trials. Then, perceived direction spanned between the three possible solutions

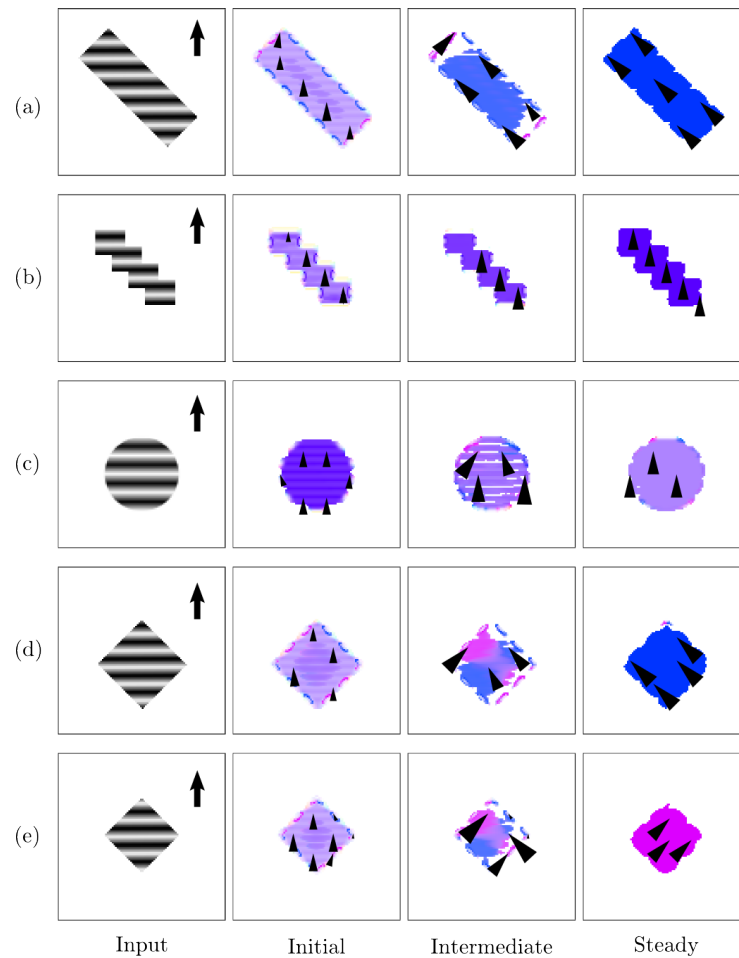


Figure 9: Model response to upward moving gratings presented behind various aperture shapes. For a given row each column shows respectively one image of the stimulus, and three representation of the motion field for the initial response, intermediate response, and steady state. We tested the following apertures: (a) Tilted rectangular aperture with an aspect ratio of 3:1. (b) The aperture edges are indented to locally change the direction of terminator motions. (c) Circular aperture. (d) Square aperture. (e) Slightly smaller square aperture.

Stimulus	Figure	Need layer $\phi$ ?
Line segments	5	No
Ellipses	7	
Gratings	9	
Dotted square	10	Yes
Chopsticks	11	
Diamonds	12	

Table 2: Overview of form-modulation influence on the considered stimuli. All the described stimuli can be classified, with respect to form modulation, as either simple or complex depending on the information they need to lead to the correct percept and dynamics. For complex stimuli, motion information alone is not enough to reproduce the visual cortex mechanisms.

aforementioned. We ran successive simulations with a barber pole of constant aspect ratio 1:1 but introducing small fluctuations in either the input image sequence  $I$  or the input local motion  $p_0$ . For instance, slightly changing the size of the square aperture resulted in a dramatic change in global motion estimation, switching from left- to right-upward direction (see Figure 9d,e). Introducing a small additive Gaussian noise (average: 0.5, variance: 0.02) into  $p_0$  resulted in similar switches. Thus, small changes in stimulus characteristics can lead to totally different estimates of global motion in our dynamical model.

Lastly, we investigated the role of the layer  $\phi$  in motion integration for this kind of stimulus. Dynamics was largely unaffected when ignoring luminosity smoothness information, apart from a small increase in the time constant of the direction error. Thus, dynamics of motion integration between local 1D and 2D motion cues can be largely explained by the winner-take-all mechanism through lateral connectivity and motion diffusion through feedback connectivity. As shown in Table 2, the three kind of stimuli, namely line segments, ellipses, and gratings, do not require the proposed extra luminance gating, as opposed to the more complex stimuli of the following section.

### 3.4 Influence of form on selective motion integration

Previous models of form-motion integration have shown that form information is important for integrating motion across apertures. Here, we investigated how luminance-gated motion diffusion can be used in integrating local motion signals that belong to a given object. Our model can reproduce some key aspects of motion integration *versus* segmentation by testing its response to a large class of motion stimuli used in both psychophysics and neurophysiology.

In this paper, we focus on 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 (Huang et al., 2007; Shiffrar & Lorenceau, 1996). Second, a large bulk of psychophysical data suggests that motion features are discarded when they do not belong to the moving surface (i.e. when they are extrinsic) (Shiffrar et al., 1995; Lorenceau & Shiffrar, 1992; Shimojo et al., 1989). Our model must then be able to selectively integrate motion signals that belong to the moving surface of interest and avoid propagation of local 2D motion signals that are not intrinsic to it.



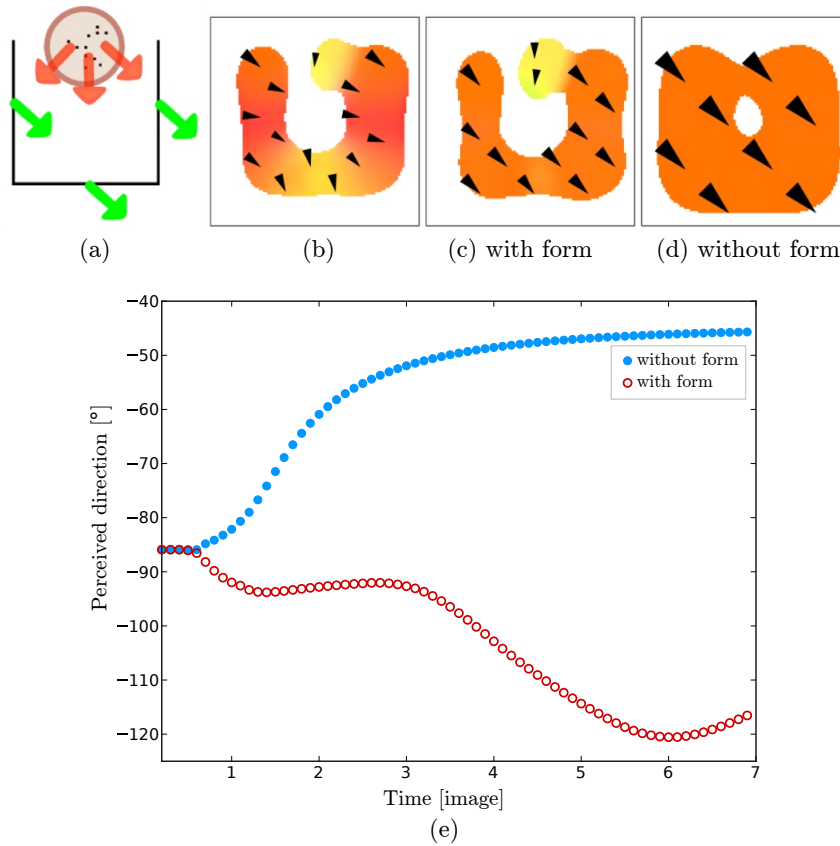


Figure 10: Model response to the motion stimulus proposed by Huang et al. (2007) to test selective motion integration in area MT. (a) A square moving diagonally downward and to the right is presented together with a patch of moving dots instead of the upper segment of the square (see text for more details). (b) Initial model response illustrated by velocity field,  $m_2$ , computed over the first few images showing that edge motion estimates are biased by the aperture problem. (c) Result obtained with luminance-gated diffusion. (d) Result with isotropic diffusion. (e) Perceived direction  $w(t)$  computed inside the dotted region.

### Preventing capture: the dotted square stimulus

In Figure 10, we considered the stimulus proposed by Huang et al. (2007) and we tested how selective is motion integration performed by area MT neurons. The stimulus is described as follows: a square moving in the lower right direction has its upper edge removed and replaced by a set of points moving randomly downward; the velocity of the moving points spans the velocity distribution existing at the centre of an edge due to the aperture problem. Our model gives results similar to those observed with MT neurons recordings: 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 biased the initial motion direction at the centre of the three edges (see Figure 10b). As illustrated in Figure 10c, the aperture problem was correctly solved so that at the end of the simulation, all three edges moved coherently along the 2D translation axis, i.e. diagonally downward and to the right. Notice that motion direction of the patch remained unaffected at all iterations. In brief, two sets of object motion coexists without capture. However, in the isotropic diffusion experiment, random dot patch motion was captured by downward drift of the edges (compare Figure 10c and 10d).

### Influence of context: the chopsticks

In the next example, we considered the chopstick illusion in order to illustrate the influence of form information onto the selective integration of motion information (Anstis, 1990). The first stimulus consists in two horizontally translating bars, as shown in Figure 11a: Thus we introduce two sets of non ambiguous motions arising from the end of lines (i.e. horizontal motion), and from the bars intersection (i.e. vertical motion). In Figure 11a, we illustrate the velocity field  $m_1$  estimated at different times. Our results are coherent with the phenomena reported by psychophysical experiments: under these conditions, two bars are perceived as moving in opposite directions (Anstis, 1990). We also show that velocity flow fields were coherent at the two different spatial scales  $m_1$  and  $m_2$  showing that feedback allows the model to compute coherent motion representation at different stages along the motion pathway. Removing the  $\phi$  layer, resulted in the opposite motion perception: the computed velocity field corresponded to two bars moving coherently upward, forming a single cross being translated vertically.

In the second stimulus, line-endings were made extrinsic by placing two horizontal occluders at the ends of the chopstick (see Figure 11b). In this case the motion percept consists of a single upward translation. Applying the proposed luminance-gated motion diffusion was enough to reproduce this phenomenon. Figure 11b 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  $m_2$  flow fields) but are not propagated inside the line-drawing figures. On the contrary, 2D motion signals arising the intersection between the two lines were propagated along the edges so that after 20 frames, the two bars are perceived as moving coherently in the upward direction.

Applying an isotropic diffusion resulted in a dramatic change in the output velocity fields: 2D motion signals arising at the intersections between edges

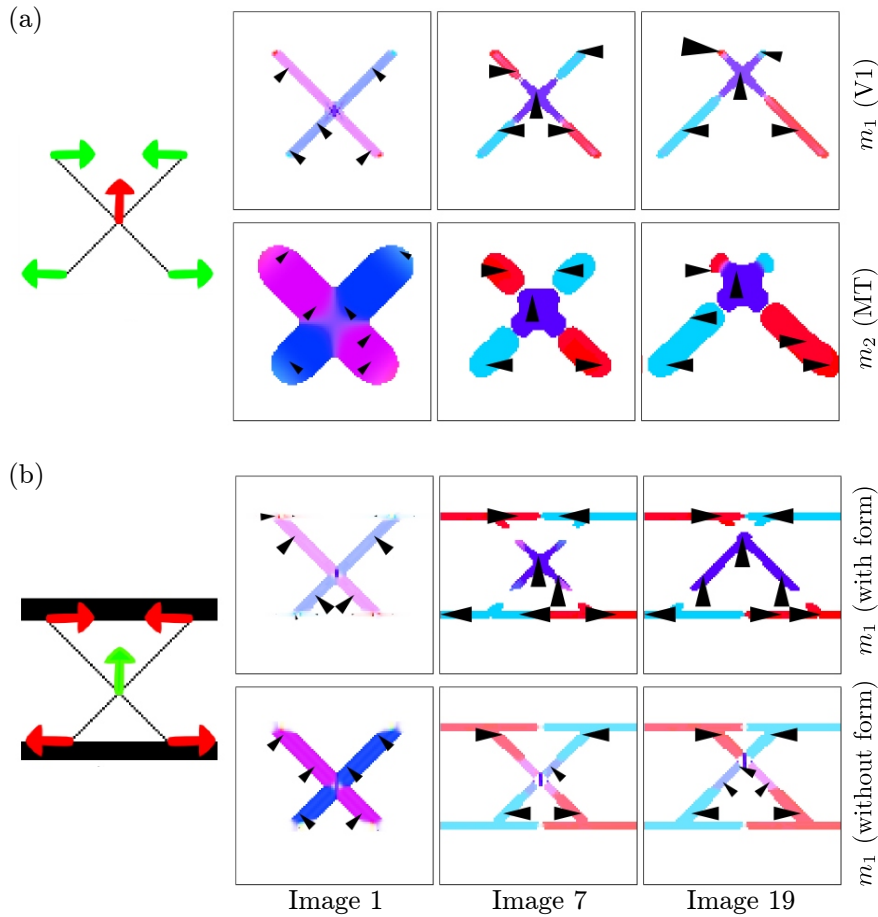


Figure 11: Model response to chopsticks motion. (a) Two tilted and crossing bars are translating in opposite motion direction resulting in two horizontal perceived motions (Anstis, 1990). We display the velocity fields  $m_1$  and  $m_2$  to illustrate the time course of motion computation at two different spatial scales. (b) Model response to occluded chopsticks where two horizontal occluders of different luminance dramatically change the motion percept, leading to a vertical perceived motion. We illustrate model performance as the velocity field  $m_1$  computed at three different times. Upper and lower rows illustrate the results obtained with luminance-gated motion diffusion, or isotropic diffusion respectively.

and occluders were now propagated both along the chopsticks and the edges of the aperture. Such a solution would correspond to the perception of two sticks sliding over each other. Moreover, bars motion captured the occluder edges. This result demonstrates the role of the layer  $\phi$  to implement contextual modulation of motion diffusion, simulating different percepts such as coherent (i.e. one single object) or incoherent (i.e. overlapping objects) motion of the two bars.

### Geometry controlled diffusion: diamonds

Another challenging set of experiments was provided by the study of Lorenceau & Alais (2001) as illustrated in Figure 12. In the original psychophysical study, rotating diamonds like stimuli were displayed to the subjects for long durations. For each stimulus, the subject were asked if the rotation was perceived as clockwise or counterclockwise. The percentage of correct responses have been replotted in Figure 12 (dark grey bars) for the 10 different shapes used in this study. Lorenceau & Alais (2001) found two groups of objects, with performance above and below 80% (horizontal dotted line).

To obtain a result comparable to the rotation coherence described in Lorenceau & Alais (2001), we defined a rotation coherence read-out as follows. First, we decomposed local motion as given by the activity measurements of our model into a radial-rotational space. The bio-plausibility of such a decomposition, as well as its links to human motion percept, have already been described in Barzaza & Grzywacz (2005). This decomposition corresponds to a simple change of coordinates. Then, we computed, via a spatio-temporal average, the global ratio of the rotation over radial motion. Figure 12 plots this ratio for the same 10 shapes (light grey bars). Overall, the different shapes can be grouped similarly into two different sets of stimuli, which are consistent with those obtained from psychophysical experiments. Thus, the model performed better for stimuli that we perceived as being coherent, suggesting a similar solution for motion integration across apertures. However, we found an intriguing mismatch between two stimuli out of ten (as indicated by the two symbols \* in Figure 12), most probably because of the distance between the corresponding line-endings.

## 4 Discussion

In the present study, we proposed a luminance-gated motion diffusion model to solve 2D motion integration and segmentation. We implemented and applied our 4 layers dynamical system to synthetic motion stimuli with the goal to reproduce several key phenomena of 2D motion integration that have been documented by psychophysical, behavioural and neurophysiological studies. In particular, we reproduced the temporal dynamics of motion integration and its dependency upon stimulus characteristics. Furthermore, we showed that the simple computational rule of luminance-gated diffusion of motion information, is sufficient to explain a large set of contextual modulations of motion integration. In the sections below, we first discuss the main contributions of our work concerning the dynamical properties of the solutions and the influence of the luminance in gating the diffusion process. Then we discuss how our model re-

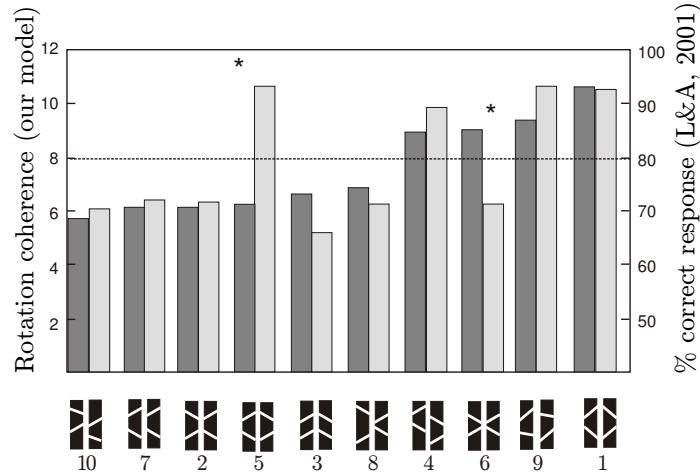


Figure 12: Results on the whole set of stimuli presented in Lorenceau & Alais (2001). All the stimuli are made of four edges varying a diamond shape. The dark grey bars represents the correct rotation response from the psychophysical experiments (see rightward). The light grey bars corresponds to the responses of our model (see leftward axis).

lates to the state of the art, and we conclude by describing two limitations of our model.

#### 4.1 Temporal dynamics of motion integration

Our model successfully reproduced 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, line-drawing objects and barber poles, we found that during the first iterations almost no contribution of 2D motion signals as generated by line-endings or terminators can be seen. This is consistent with the observations made in area MT that early direction tuning of cells is driven by component motions (e.g. (Pack & Born, 2001; Smith et al., 2005; Pack et al., 2004)). At behavioural level, Masson and colleagues found similarly that the earliest phase of ocular following responses to either unikinetic plaids or barber poles is only driven in the direction of grating motion (Masson et al., 2000b; Masson & Castet, 2002; Barthélemy et al., 2008). The origin of such delay between 1D and 2D driven responses has been highly controversial. Some authors attributed it to the delay seen in the emergence of end-stopping properties of V1 neurons (Pack et al., 2003). This temporal dynamics might be related to the timing of the underlying centre-surround interactions (Bair et al., 2003). However, the relative contribution of both lateral and feedback recurrent connectivity to the temporal dynamics of centre-surround interactions is still unclear (see (Angelucci & Bullier, 2003) for a review).

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 have been shown that form-driven responses in area V2 are delayed relative to the fast MT neuronal responses (see Lamme & Roelfsema (2000) for a review). Nevertheless, our sim-

ulations show that any significant contribution of 2D features as emerging from the cortical dynamics must be 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 reflects 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 Lamme & Roelfsema (2000); Bullier (2001).

Next, our model can reproduce the time course of 2D motion integration as evidenced by a large number of studies at both psychophysical (Yo & Wilson, 1992; Castet et al., 1993; Lorenceau et al., 1993), behavioural Masson & Stone (2002); Wallace et al. (2005); Born et al. (2006) and neurophysiological (Pack & Born, 2001; Smith et al., 2005; Pack et al., 2004) 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 the initial bias and time constant of the decay vary with contrast of local non-ambiguous features, line length, barber pole aspect ratio and so on. All these scaling factors affect the dynamics of lateral diffusion. Hence, the recurrent dynamics, which is needed for the diffusion of motion information, can largely explain the observed dynamics of motion integration. Our dynamical model provides a platform to further investigate which biologically realistic neuronal architectures can underlie such computation.

## 4.2 Luminance smoothness: a simple rule for gating motion information

We have implemented a simple mechanism for using form information in the context of motion integration. In particular, we did not implement any complex local features detectors such as end-stopped cells or dipole cells that are found in Berzhanskaya et al. (2007). Here, the layer  $\phi$  indicates directions in the image along which luminance is nearly uniform. Such abstract definition of form information incorporates a form representation as well as a 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 (Hubel & Wiesel, 1962). Abrupt changes in the luminance profile along the contour can be signalled by another type of neurons found in area V1, end-stopped cells (Hubel & Wiesel, 1968; Pack et al., 2004). Albeit neuronal selectivity for more complex shapes can be found at higher hierarchical stages 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 (Berzhanskaya et al., 2007). The layer  $\phi$  used in the present study only signals isoluminant directions and uses this information for guiding motion integration without the need of explicit feature detectors. Moreover, it also implements some kind of surface representation by

signalling luminance smoothness along a wide range of direction. Noteworthy, the layer  $\phi$  implements both contour and surface smoothness constraints using luminance information and therefore offer a simple solution for the need of using both smoothness constraints to efficiently solve motion integration problems (Weiss & Adelson, 2000).

Our model also offers a framework to investigate interactions between luminance and motion processing within the cortical pathways. Recent studies have pointed out that luminance information is encoded at the earliest stage of cortical processing (Rossi et al., 1996; Peng & Van Essen, 2005; Geisler et al., 2007). 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 (Kinoshita & Komatsu, 2001; Tani et al., 2003). Such representation of uniform surfaces based on luminance distribution has been related to brightness perception (Rossi et al., 1996). Our model suggests 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 centre of a disk with edges located far outside the receptive field (Pack & Born, 2002). On the contrary, these cells remained unresponsive to a circle of same diameter. Our model can reproduce this dynamics, thanks to the layer  $\phi$ .

In fact, using luminance information, and in particular the fact that luminance profiles smoothly vary both along single (edges) or multiple (surface) directions might be a very efficient strategy 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 (Geisler, 2008). Considerable attention has been paid to the statistics of contrast distribution and its relationships with the properties of elementary local features detectors (see Simoncelli & Olshausen (2001) for a review). Recent studies have shown the importance of luminance distribution as well (Mante et al., 2005; Frazor & Geisler, 2006) and pinpoint its role in the neural dynamics of local information processing (Mante et al., 2005; Geisler et al., 2007) but also in surface segmentation (e.g. Fine et al. (2003)). 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.

The fact that our model can reproduce many psychophysical observations using a wide range of object shapes (ellipses, chopsticks, line-drawings) stresses the fact that luminance-gating of motion integration is a simple but efficient implementation of interaction between form information and motion information. Motion stimuli used by Lorenceau and Alais (2001) presents the advantages to have identical motion energy. The main difference between the ten stimuli illustrated in Figure 12 was the geometrical relationships between the different segments. Our model produced similar grouping for the same subset of stimuli. This further illustrates the fact that controlling motion diffusion using luminance smoothness can be a simple neural solution for what has been described as form-dependant motion integration. Further work will be conducted to investigate the detailed implementation of this rule.

Lastly our model calls for further experimental and theoretical studies about non-isotropic diffusion of information within or across cortical layers. The fact that point-like processes such as orientation or direction extraction can be in-

terconnected along preferred axis within the cortical sheet has been already suggested by both psychophysical and physiological studies. For instance, the “association field” proposed by Hess and colleagues postulate that contour integration involves facilitatory interactions between orientation-tuned neurons that are colinear and aligned within visual space (see Field et al. (1993); Hess & Field (1999); Hess et al. (2003) for reviews). The colinear facilitatory effects seen for contrast detection (Polat & Sagi, 1993, 1994), static and dynamical contours detection (e.g. Field et al. (1993)) and apparent motion perception (Georges et al., 2002; Seriès et al., 2002) are often thought to be mediated by intra-cortical short-range lateral connections (Bosking et al., 1997; Stettler et al., 2002).

### 4.3 Modelling 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 (Wilson et al., 1992). This model was further expanded to account for barber pole and line motion perception (Löffler & Orbach, 1998). As in any two-motion pathway model, 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 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 barber pole aspect ratios.

On the contrary, the role of diffusion for motion estimation has been investigated thoroughly in the computer vision community. There exists a huge literature concerning the estimation of the so-called optical flow, which is how to estimate accurately the apparent velocity field from videos (see, e.g. (Stiller & Konrad, 1999; Otte & Nagel, 1994; Barron et al., 1994; Aggarwal & Nandhakumar, 1988) for reviews). Almost all of these approaches rely on the brightness consistency assumption leading to the classical optical flow constraint (OFC) 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. For example, one can constrain the smoothness of the solution: the goal is thus to find a compromise between respecting the OFC and having the required degree of smoothness. To do so, one possibility is to define a variational formulation: In this direction, let us mention for example the pioneering work by Horn & Schunk (1981) where smoothness was defined by minimising a quadratic term of the velocity components gradient. The key point here is that choosing a degree of smoothness is equivalent to define the penalty term which will then determine how information is diffused. Interestingly, diffusion is very related to the integration processes discussed in this paper and one can see some analogies. For optical flow, many nonlinear diffusion operators were proposed to prevent models from smoothing the solution across the flow discontinuities (see for example Alvarez et al. (2007); Xiao et al. (2006); Aubert & Kornprobst (1999); Weickert (1997)). But there is yet another set of approaches using also form/luminance modulation for the diffu-



sion process. For example, Hildreth (1983a) presented a model that calculates the velocity field of least variation along a contour in the scene, corresponding a contour smoothness constraint. Similarly, Nagel & Enkelmann (1986) proposed an oriented smoothness constraint in which smoothness is not imposed across steep intensity gradients (edges) in an attempt to handle occlusions. However, as a general observation, models proposed in computer vision ignore the temporal dynamics of motion integration and never try to reproduce visual system properties and behaviour.

Several biologically-inspired models have also been previously designed to investigate the role of motion diffusion in the context of motion integration (Chey et al., 1997; Berzhanskaya et al., 2007; Bayerl & Neumann, 2007). These models are able to capture several aspects of motion integration such as the propagation of feature tracking estimates (Chey et al., 1997; Grossberg et al., 2001). Some of these models implement isotropic motion diffusion by using Gaussian distributions of activity both within layers and between layers through recurrent connectivity (Bayerl & Neumann, 2004, 2007). They can simulate the temporal dynamics of motion integration for simple motion stimuli but cannot render more complex selective motion integration without the need of implementing complex rules such as T-junctions motion cancellation or using distributing motion signal across different depth layers.

Lastly isotropic diffusion 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 (see Berzhanskaya et al. (2007) for the latest version of the model). A similar approach using depth cues was proposed recently by Beck & Neumann (2010). 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 apertures and therefore reproduce the perceived global motion direction observed with motion stimuli such as the occluded diamonds (Shiffrar et al. (1995); Lorenceau & Alais (2001) or the chopsticks Anstis (1990)). Notice that form-motion interaction was used in their model only to disambiguate motion information at the stage of area MT. No feedback was 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 also 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.

Moreover, the FORMOTION model relies heavily on the assignment of each object to a given depth layer. To do so, the authors implemented a complex architecture with six processing stages in the form pathway and seven stages in the motion pathway. Multiple feedforward and feedback interactions are implemented at different levels (Berzhanskaya et al., 2007) and the model postulates the existence of several types of highly specific form and motion detectors. In contrast, in this paper we proposed a minimal 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 four layers, our model can reproduce as many perceptual phenomena as the FORMOTION model.

Our model also implements 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 more complex operations.

#### 4.4 Limitations of the model

We have shown that our model can qualitatively reproduce the dynamics of several key phenomenon of 2D selective motion integration. We successfully applied it on a larger set of motion stimuli than competing recurrent models (Bayerl & Neumann, 2004, 2007). Moreover, the luminance smoothness rule offers a simpler approach than previous models of motion integration (Berzhanskaya et al., 2007). However, the current version of model suffers from two limitations.

First, we cannot model the well-known effects of contrast upon 2D motion integration. Other models had also difficulties in implementing the effects of contrast since almost none neurophysiological experiments have been conducted to investigate the effect of global contrast upon 2D motion integration. Weiss et al. (2001) showed that a Bayesian model of motion integration can mimic the effect of lowering contrast upon the perceived direction. However, their model was not intended to process moving images and therefore lower contrast was directly modelled by a higher variance of the Gaussian distributions forming the velocity likelihoods. Motion energy filters in our model were made insensitive to contrast and, as a consequence, we cannot account for these effects. Moreover, the spatial summation properties of V1 and MT units were not defined as being sensitive to contrast, a factor that could change the dynamical properties of motion diffusion. We attempted to simulate the effect of contrast by adding white noise to the input frames. We found that large additive noise both increased the initial bias towards the vector average prediction and slightly slowed down the time course of direction errors. Both results are consistent with behavioural results Masson & Stone (2002); Wallace et al. (2005); Born et al. (2006). However, a full model should incorporate contrast-dependant local motion filters (such as the one described in Escobar (2009)) as well as contrast-dependent spatial integration mechanisms as found in areas V1 (Sceniak et al., 1999) and MT (Pack et al., 2005).

A second limitation of the model concern the role of other segmentation cues such as depth ordering. The role of binocular disparity could be easily tested in our framework by having both form and motion mechanism being made of disparity-selective neurons. Thus, diffusion could be made anisotropic within different sub-population of neurons tuned for different depths, so that transparent motion for example could be analysed in such a way. Further work will investigate the dynamical properties of motion, luminance and depth cues combination using simple rules such as implemented in the present study.

## Acknowledgements

This research work has received funding from the European Community's Seventh Framework Program under the grant agreement N°215866, project SEARISE, and the Région Provence Alpes Côte d'Azur. GSM was supported by the CNRS, the European Community (FACETS, IST-FET, Vith Framework, N°025213),

and the Agence Nationale de la Recherche (ANR, NATSTATS). We thank Drs. Pascal Mamassian, Éric Castet and Jean Lorenceau for helpful discussions and comments on the manuscript.

## References

- Adelson, E., & Bergen, J. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, *2*, 284–299.
- Adelson, E., & Movshon, J. (1982). Phenomenal coherence of moving visual patterns. *Nature*, *300*, 523–525.
- Aggarwal, J., & Nandhakumar, N. (1988). On the computation of motion from sequences of images — a review. *Proceedings of the IEEE*, *76*, 917–935.
- Alexiadis, D., & Sergiadis, G. (2008). Narrow directional steerable filters in motion estimation. *Computer Vision and Image Understanding*, *110*, 192–211.
- Alvarez, L., Deriche, R., Papadopoulo, T., & Sánchez, J. (2007). Symmetrical dense optical flow estimation with occlusions detection. *International Journal of Computer Vision*, *75*, 371–385.
- Angelucci, A., & Bullier, J. (2003). Reaching beyond the classical receptive field of V1 neurons: horizontal or feedback axons? *J Physiol Paris*, *97*, 141–154.
- Anstis, S. (1990). Imperceptible intersections: The chopstick illusion. In *AI and the Eye* (p. 105).
- Aubert, G., & Kornprobst, P. (1999). A mathematical study of the relaxed optical flow problem in the space BV. *SIAM Journal on Mathematical Analysis*, *30*, 1282–1308.
- Bair, W., Cavanaugh, J., & Movshon, A. (2003). Time course and time–distance relationships for surround suppression in macaque V1 neurons. *The Journal of Neuroscience*, *23*, 7690–7701.
- Baker, S., Scharstein, D., Lewis, J., Roth, S., Black, M., & Szeliski, R. (2007). A database and evaluation methodology for optical flow. In *International Conference on Computer Vision, ICCV'07* (pp. 1–8).
- Barron, J., Fleet, D., & Beauchemin, S. (1994). Performance of optical flow techniques. *The International Journal of Computer Vision*, *12*, 43–77.
- Barthelemy, F., Fleuriet, J., & Masson, G. (2009). Temporal dynamics of 2D motion integration for ocular following in macaque monkeys. *Journal of Neurophysiology*, .
- Barthélemy, F., Perrinet, L., Castet, E., & Masson, G. (2008). Dynamics of distributed 1d and 2d motion representations for short-latency ocular following. *Vision Research*, *48*, 501–522.
- Bayerl, P., & Neumann, H. (2004). Disambiguating visual motion through contextual feedback modulation. *Neural Computation*, *16*, 2041–2066.

- Bayerl, P., & Neumann, H. (2005). Attention and figure-ground segregation in a model of motion perception. *Journal of Vision*, *5*, 659–659.
- Bayerl, P., & Neumann, H. (2007). Disambiguating visual motion by form–motion interaction – a computational model. *International Journal of Computer Vision*, *72*, 27–45.
- Beck, C., & Neumann, H. (2010). Interactions of motion and form in visual cortex – a neural model. *Journal of Physiology - Paris*, (pp. 61–70).
- Berzhanskaya, J., Grossberg, S., & Mingolla, E. (2007). Laminar cortical dynamics of visual form and motion interactions during coherent object motion perception. *Spatial Vision*, *20*, 337–395.
- Born, R., & Bradley, D. (2005). Structure and function of visual area MT. *Annu. Rev. Neurosci*, *28*, 157–189.
- Born, R., Pack, C., Ponce, C., & Yi, S. (2006). Temporal evolution of 2-dimensional direction signals used to guide eye movements. *Journal of Neurophysiology*, *95*, 284–300.
- Bosking, W., Zhang, Y., Schofield, B., & Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *The Journal of Neuroscience*, *17*, 2112–2127.
- Bowns, L. (1996). Evidence for a feature tracking explanation of why type II plaids move in the vector sum direction at short durations. *Vision Research*, *36*, 3685–3694.
- Bradley, D., & Goyal, M. (2008). Velocity computation in the primate visual system. *Nature Reviews Neuroscience*, *9*, 686–695.
- Bullier, J. (2001). Integrated model of visual processing. *Brain Res. Reviews*, *36*, 96–107.
- Castet, E., Charton, V., & Dufour, A. (1999). The extrinsic/intrinsic classification of two-dimensional motion signals with barber-pole stimuli. *Vision Research*, *39*, 915–932.
- Castet, E., Lorenceau, J., Shiffrar, M., & Bonnet, C. (1993). Perceived speed of moving lines depends on orientation, length, speed and luminance. *Vision Research*, *33*, 1921–1921.
- Chey, J., Grossberg, S., & Mingolla, E. (1997). Neural dynamics of motion processing and speed discrimination. *Vision Res.*, *38*, 2769–2786.
- Dayan, P., & Abbott, L. (2001). *Theoretical Neuroscience : Computational and Mathematical Modeling of Neural Systems*. MIT Press.
- Derpanis, K., & Gryn, J. (2005). Three-dimensional nth derivative of gaussian separable steerable filters. In *IEEE International Conference on Image Processing* (pp. 553–556). volume 3.
- Escobar, M.-J. (2009). *Bio-Inspired Models for Motion Estimation and Analysis: Human action recognition and motion integration*. Ph.D. thesis Université de Nice Sophia-Antipolis.

- Fennema, C., & Thompson, W. (1979). Velocity determination in scenes containing several moving objects. *Computer Graphics and Image Processing*, *9*, 301–315.
- Ferrera, V., & Wilson, H. (1990). Perceived direction of moving two-dimensional patterns. *Vision Research*, *30*, 273–287.
- Field, D., Hayes, A., & Hess, R. (1993). Contour integration by the human visual system: evidence for a local "association field". *Vision Research*, *33*, 173–193.
- Fine, I., MacLeod, D., & Boynton, G. (2003). Surface segmentation based on the luminance and color statistics of natural scenes. *Journal of the Optical Society of America A*, *20*, 1283–1291.
- Frazor, R., & Geisler, W. (2006). Local luminance and contrast in natural images. *Vision Research*, *46*, 1585–1598.
- Freeman, W., & Adelson, E. (1991). The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*, 891–906.
- Geisler, W. (2008). Visual perception and the statistical properties of natural scenes. *Annu. Rev. Psychol.*, *59*, 167–192.
- Geisler, W., Albrecht, D., & Crane, A. (2007). Responses of neurons in primary visual cortex to transient changes in local contrast and luminance. *Journal of Neuroscience*, *27*, 5063.
- Georges, S., Seriès, P., Frégnac, Y., & Lorenceau, J. (2002). Orientation dependent modulation of apparent speed: psychophysical evidence. *Vision Research*, *42*, 2757–2772.
- Gorea, A., & Lorenceau, J. (1991). Directional performances with moving plaids: component-related and plaid-related processing modes coexist. *Spatial vision*, *5*, 231–252.
- Grossberg, S., & Mingolla, E. (1985). Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychological review*, *92*, 173–211.
- Grossberg, S., Mingolla, E., & Viswanathan, L. (2001). Neural dynamics of motion integration and segmentation within and across apertures. *Vision Research*, *41*, 2521–2553.
- Grzywacz, N., & Yuille, A. (1991). Theories for the visual perception of local velocity and coherent motion. Bradford Books chapter 16. (pp. 231–252). The MIT Press.
- Heeger, D. (1988). Optical flow using spatiotemporal filters. *The International Journal of Computer Vision*, *1*, 279–302.
- Hess, R., & Field, D. (1999). Integration of contours: new insights. *Trends in Cognitive Sciences*, *3*, 480–486.

- Hess, R., Hayes, A., & Field, D. (2003). Contour integration and cortical processing. *Journal of Physiology-Paris*, *97*, 105–119.
- Hildreth, E. (1983a). The detection of intensity changes by computer and biological vision systems. *Computer Vision, Graphics, and Image Processing*, *22*, 1–27.
- Hildreth, E. (1983b). *The measurement of visual motion*. Ph.D. thesis MIT.
- Horn, B., & Schunk, B. (1981). Determining optical flow. *Artificial Intelligence*, *17*, 185–203.
- Huang, X., Albright, T., & Stoner, G. (2007). Adaptive surround modulation in cortical area MT. *Neuron*, *53*, 761–770.
- Hubel, D., & Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat visual cortex. *J Physiol*, *160*, 106–154.
- Hubel, D., & Wiesel, T. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*, 215.
- Kapadia, M., Ito, M., Gilbert, C., & Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron*, *50*, 35–41.
- Kinoshita, M., & Komatsu, H. (2001). Neural representation of the luminance and brightness of a uniform surface in the macaque primary visual cortex. *Journal of Neurophysiology*, *86*, 2559–2570.
- Kooi, T. (1993). Local direction of edge motion causes and abolishes the barberpole illusion. *Vision Research*, *33*, 2347–2351.
- Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*, 571–579.
- Lee, T., & Mumford, D. (2003). Hierarchical bayesian inference in the visual cortex. *J. Opt. Soc. Am. A*, *20*.
- Lee, T., & Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. *Proceedings of the National Academy of Sciences*, *98*, 1907.
- Löffler, G., & Orbach, H. (1998). Computing feature motion without feature detectors: A model for terminator motion without end-stopped cells. *Vision Research*, *39*, 859–871.
- Lorenceau, J. (2010). From moving contours to object motion: Functional networks for visual form/motion processing. In Masson & Ilg (2010).
- Lorenceau, J., & Alais, D. (2001). Form constraints in motion binding. *Nature Neuroscience*, *4*, 745–751.
- Lorenceau, J., & Shiffrar, M. (1992). The influence of terminators on motion integration across space. *Vision Research*, *32*, 263–273.

- Lorenceanu, J., Shiffrar, M., Wells, N., & Castet, E. (1993). Different motion sensitive units are involved in recovering the direction of moving lines. *Vision Research*, *33*, 1207–1207.
- Mante, V., Frazor, R., Bonin, V., Geisler, W., & Carandini, M. (2005). Independence of luminance and contrast in natural scenes and in the early visual system. *Nature Neuroscience*, .
- Masson, G., & Castet, E. (2002). Parallel motion processing for the initiation of short-latency ocular following in humans. *The journal of neuroscience*, *22*, 5147–5163.
- Masson, G., & Ilg, U. (Eds.) (2010). *Dynamics of Visual Motion Processing. Neuronal, Behavioral, and Computational Approaches* (1st ed.). Springer Verlag.
- Masson, G., Rybarczyk, Y., Castet, E., & Mestre, D. (2000a). Temporal dynamics of motion integration for the initiation of tracking eye movements at ultra-short latencies. *Visual Neuroscience*, *17*, 753–767.
- Masson, G., Rybarczyk, Y., Castet, E., & Mestre, D. (2000b). Temporal dynamics of motion integration for the initiation of tracking responses at ultra-short latencies. *Visual Neuroscience*, *17*, 754–767.
- Masson, G., & Stone, L. (2002). From following edges to pursuing objects. *Journal of neurophysiology*, *88*, 2869–2873.
- Montagnini, A., Spering, M., & Masson, G. (2006). Predicting 2D target velocity cannot help 2D motion integration for smooth pursuit initiation. *J Neurophysiol*, *96*, 3545–3550.
- Movshon, J., Adelson, E., Gizzi, M., & Newsome, W. (1985). The analysis of visual moving patterns. *Pattern recognition mechanisms*, (pp. 117–151).
- Nagel, H., & Enkelmann, W. (1986). 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.
- Nakayama, K., & Silverman, G. (1988). The aperture problem. II. spatial integration of velocity information along contours. *Vision Research*, *28*, 747–753.
- Nelson, J., & Frost, B. (1985). Intracortical facilitation among co-oriented, co-axially aligned simple cells in cat striate cortex. *Experimental Brain Research*, *61*, 54–61.
- Nowlan, S., & Sejnowski, T. (1994). Filter selection model for motion segmentation and velocity integration. *J. Opt. Soc. Am. A*, *11*, 3177–3199.
- Otte, M., & Nagel, H. (1994). Optical flow estimation: Advances and comparisons. In J.-O. Eklundh (Ed.), *Proceedings of the 3rd European Conference on Computer Vision* (pp. 51–70). Springer-Verlag volume 800 of *Lecture Notes in Computer Science*.

- Pack, C., & Born, R. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature*, *409*, 1040–1042.
- Pack, C., & Born, R. (2002). Integration of motion signals over regions of uniform luminance by mt neurons in the alert macaque. *Journal of Vision*, *2*, 412a. Abstract.
- Pack, C., Gartland, A., & Born, R. (2004). Integration of contour and terminator signals in visual area MT of alert macaque. *The Journal of Neuroscience*, *24*, 3268–3280.
- Pack, C., Hunter, J., & Born, R. (2005). Contrast dependence of suppressive influences in cortical area MT of alert macaque. *Journal of Neurophysiology*, *93*, 1809–1815.
- Pack, C., Livingstone, M., Duffy, K., & Born, R. (2003). End-stopping and the aperture problem: Two-dimensional motion signals in macaque v1. *Neuron*, *39*, 671–680.
- Paradiso, M., Blau, S., Huang, X., MacEvoy, S., Rossi, A., & Shalev, G. (2006). Lightness, filling-in, and the fundamental role of context in visual perception. *Progress in Brain Research*, (pp. 109–123).
- Peng, X., & Van Essen, D. (2005). Peaked encoding of relative luminance in macaque areas V1 and V2. *Journal of Neurophysiology*, *93*, 1620–1632.
- Polat, U., & Sagi, D. (1993). Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision Research*, *33*, 993–993.
- Polat, U., & Sagi, D. (1994). The architecture of perceptual spatial interactions. *Vision Research*, *34*, 73–78.
- Power, R., & Moulden, B. (1992). Spatial gating effects on judged motion of gratings in apertures. *Perception*, *21*, 449–449.
- Priebe, N., Lisberger, S., & Movshon, A. (2006). Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex. *The Journal of Neuroscience*, *26*, 2941–2950.
- Reichardt, W. (1957). Autokorrelationsauswertung als funktionsprinzip des zentralnervensystems. *Zeitschrift für Naturforschung*, *12*, 447–457.
- Rodman, H., & Albright, T. (1989). Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). *Experimental Brain Research*, *75*, 53–64.
- Rossi, A., Rittenhouse, C., & Paradiso, M. (1996). The representation of brightness in primary visual cortex. *Science*, *273*, 1104.
- Rust, N., Mante, V., Simoncelli, E., & Movshon, J. (2006). How MT cells analyze the motion of visual patterns. *Nature Neuroscience*, *9*, 1421–1431.



- Sceniak, M. P., Ringach, D. L., Hawken, M. J., & Shapley, R. (1999). Contrast's effect on spatial summation by macaque V1 neurons. *Nature Neuroscience*, *2*, 733–739.
- Seriès, P., Georges, S., Lorenceau, J., & Frégnac, Y. (2002). Orientation dependent modulation of apparent speed: a model based on the dynamics of feedforward and horizontal connectivity in V1 cortex. *Vision Research*, *42*, 2781–2798.
- Shiffrar, M., Li, X., & Lorenceau, J. (1995). Motion integration across differing image features. *Vision Research*, *35*, 2137–2146.
- Shiffrar, M., & Lorenceau, J. (1996). Increased motion linking across edges with decreased luminance contrast, edge width and duration. *Vision Research*, *36*, 2061–2067.
- Shimojo, S., Silverman, G., & Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Research*, *29*, 619–26.
- Sillito, A., Cudeiro, J., & Jones, H. (2006). Always returning: feedback and sensory processing in visual cortex and thalamus. *TRENDS in Neurosciences*, *29*, 307–316.
- Simoncelli, E., & Heeger, D. (1998). A model of neuronal responses in visual area MT. *Vision Research*, *38*, 743–761.
- Simoncelli, E., & Olshausen, B. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, *24*, 1193–1216.
- Smith, M., Majaj, N., & Movshon, A. (2005). Dynamics of motion signaling by neurons in macaque area MT. *Nature Neuroscience*, *8*, 220–228.
- Stettler, D., Das, A., Bennett, J., & Gilbert, C. (2002). Lateral connectivity and contextual interactions in macaque primary visual cortex. *Neuron*, *36*, 739–750.
- Stiller, C., & Konrad, J. (1999). Estimating motion in image sequences. *Signal Processing Magazine, IEEE*, *16*, 70–91.
- Tani, T., Yokoi, I., Ito, M., Tanaka, S., & Komatsu, H. (2003). Functional organization of the cat visual cortex in relation to the representation of a uniform surface. *Journal of Neurophysiology*, *89*, 1112–1125.
- Van Santen, J., & Sperling, G. (1985). Elaborated reichardt detectors. *Journal of the Optical Society of America A*, *2*, 300–320.
- Wall, K., & Danielsson, P. (1984). A fast sequential method for polygonal approximation of digitized curves. *cvgip*, *28*, 220–227.
- Wallace, J., Stone, L., & Masson, G. (2005). Object motion computation for the initiation of smooth pursuit eye movements in humans. *Journal of Neurophysiology*, *93*, 2279–2293.
- Wallach, H. (1935). Über visuell wahrgenommene Bewegungsrichtung. *Psychological Research*, *20*, 325–380.

- Wallach, H., Weisz, A., & Adams, P. (1956). Circles and derived figures in rotation. *The American Journal of Psychology*, *69*, 48–59.
- Watson, A., & Ahumada, A. (1985). Model of human visual-motion sensing. *J Opt Soc Am A*, *2*, 322–342.
- Weickert, J. (1997). Coherence-enhancing diffusion of colour images. *7th National Symposium on Pattern Recognition and Image Analysis*, .
- Weiss, Y., & Adelson, E. (2000). Adventures with gelatinous ellipses – constraints on models of human motion analysis. *Perception*, *29*, 543–566.
- Weiss, Y., & Fleet, D. (2001). Velocity likelihoods in biological and machine vision. In *Probabilistic Models of the Brain: Perception and Neural Function* (pp. 81–100). MIT Press.
- Wilson, H., Ferrera, V., & Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience*, *9*, 79–97.
- Xiao, J., Cheng, H., Sawhney, H., Rao, C., & Isnardi, M. (2006). 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.
- Yo, C., & Wilson, H. (1992). Perceived direction of moving two-dimensional patterns depends on duration, contrast and eccentricity. *Vision Research*, *32*, 135–47.
- Yuille, A., & Grzywacz, N. (1988). A computational theory for the perception of coherent visual motion. *nat*, *333*, 71–74.
- Yuille, A., & Grzywacz, N. (1989). A winner-take-all mechanism based on presynaptic inhibition feedback. *Neural Computation*, *1*, 334–347.



---

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

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex  
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier  
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq  
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique  
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex  
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex  
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex  
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex

---

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