

Bio-inspired motion estimation – From modelling to evaluation, can biology be a source of inspiration?

Émilien Tlapale, Pierre Kornprobst, Guillaume Masson, Olivier Faugeras, Jan Bouecke, Heiko Neumann

► **To cite this version:**

Émilien Tlapale, Pierre Kornprobst, Guillaume Masson, Olivier Faugeras, Jan Bouecke, et al.. Bio-inspired motion estimation – From modelling to evaluation, can biology be a source of inspiration?. [Research Report] RR-7447, INRIA. 2010. inria-00532894v2

HAL Id: inria-00532894

<https://hal.inria.fr/inria-00532894v2>

Submitted on 8 Nov 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

Bio-inspired motion estimation – From modelling to evaluation, can biology be a source of inspiration?

Émilien Tlapale — Pierre Kornprobst — Guillaume S. Masson — Olivier Faugeras — Jan
D. Bouecke — Heiko Neumann

N° 7447

November 2010

. Computational Medicine and Neurosciences .

 ***rapport
de recherche***

Bio-inspired motion estimation – From modelling to evaluation, can biology be a source of inspiration?

Émilien Tlapale^{*}, Pierre Kornprobst[†], Guillaume S. Masson[‡],
Olivier Faugeras[§], Jan D. Bouecke[¶], Heiko Neumann^{||}

Theme : Computational Medicine and Neurosciences
Équipe-Projet NeuroMathComp

Rapport de recherche n° 7447 — November 2010 — 32 pages

Abstract: We propose a bio-inspired approach to motion estimation based on recent neuroscience findings concerning the motion pathway. Our goal is to identify the key biological features in order to reach a good compromise between bio-inspiration and computational efficiency. Here we choose the neural field formalism which provides a sound mathematical framework to describe the model at a macroscopic scale. Within this framework we define the cortical activity as coupled integro-differential equations and we prove the well-posedness of the model. We show how our model performs on some classical computer vision videos, and we compare its behaviour against the visual system on a simple classical video used in psychophysics. Following this idea, we propose a new benchmark to evaluate models against visual system performance. Baseline results are provided for both bio-inspired and computer vision models. Results confirm the good performance of recent computer vision approaches even on such synthetic stimuli, and also show that taking biology into account in models can improve performance. As a whole, this article affords a considerable insight into how biology can bring new ideas in computer vision at different levels: modelling principles, mathematical formalism and evaluation methodology. Perspectives around this work are promising and cover the addition of delays to constrain propagation as well as the extension of our benchmark to better characterise the visual system performance.

Key-words: motion estimation, motion integration, optical flow, psychophysics, benchmark, neural field

* emilien@tlapale.com

† Pierre.Kornprobst@inria.fr

‡ guillaume.masson@incm.cnrs-mrs.fr

§ Olivier.Faugeras@inria.fr

¶ Jan.Bouecke@uni-ulm.de

|| heiko.neumann@uni-ulm.de

Estimation du mouvement bio-inspirée – De la modélisation à l'évaluation, la biologie peut-elle être une source d'inspiration ?

Résumé : Nous proposons une approche bio-inspirée de l'estimation du mouvement, basée sur de récentes découvertes en neurosciences concernant le flux du mouvement. Notre but est d'identifier les propriétés biologiques clés permettant d'avoir un bon compromis entre la bio-inspiration et l'efficacité computationnelle. Ici, nous choisissons le formalisme des champs neuronaux, qui s'avère être un cadre mathématique adapté pour décrire le modèle à une échelle mesoscopique. Dans ce cadre, nous définissons l'activité corticale par un système d'équations intégro-différentielles couplées, et nous prouvons que le problème est bien posé. Nous montrons les performances de notre modèle sur des vidéos classiques issues de la vision par ordinateur, et nous comparons son comportement à celui du système visuel du primate sur des vidéos utilisées en psychophysique. De là, nous proposons un banc d'essai pour évaluer les performances vis à vis des performances du système visuel. Une série de résultats pour des approches bio-inspirées ainsi que des algorithmes de vision par ordinateur est fournie. Ces résultats confirment les bonnes performances des approches récentes en vision par ordinateur, même sur des stimuli psychophysiques, et montrent aussi que la bio-inspiration peut conduire à des performances accrues. Dans l'ensemble, cet article propose une nouvelle approche pour intégrer des nouvelles idées issues de la biologie en vision par ordinateur à différents niveaux : sur les principes des modèles, dans le formalisme mathématique et dans la méthodologie d'évaluation. Les perspectives autour de ce travail sont prometteuses et comprennent l'ajout de délais pour contraindre la propagation, ainsi que l'extension de notre banc d'essai afin de mieux caractériser les performances du système visuel.

Mots-clés : estimation du mouvement, intégration du mouvement, flot optique, psychophysique, banc d'essai, champs neuronaux

Contents

1	Introduction	3
2	The primate visual system	4
2.1	General considerations	4
2.2	The motion pathway	5
2.3	From biology to bio-inspired models?	5
3	A neural field model for motion estimation	6
3.1	Biological inspiration of the model	6
3.2	Description of maps interactions	8
3.2.1	The neural field framework	8
3.2.2	Local motion estimation	9
3.2.3	Core equations	10
3.3	Existence and uniqueness of the solution	11
3.4	Relations to the state of the art in computer vision	14
4	Results	15
4.1	Implementation details	15
4.2	Natural scenes	15
4.3	Psychophysical stimulus	20
5	Towards a bio-inspired benchmark	21
5.1	Motivation	21
5.2	Database Design	22
5.2.1	The two fundamental questions	22
5.2.2	Line-drawing objects	23
5.2.3	Gratings	23
5.3	Baseline results	24
6	Conclusion	25

1 Introduction

Biological vision shows intriguing characteristics in terms of performance and robustness. For example, the primate visual system is able to handle very large inputs, with around sixty-five millions cones and twice as many rods, whereas typical software input is two or three orders of magnitude smaller in terms of pixels. The luminance range and contrast sensitivity of the primate visual system are hardly comparable to the few bits used to represent grey scale as input to algorithmic software. A final noteworthy feature of the visual system is its capacity to respond correctly to a wide range of spatial and temporal scales, implying sensitivity to a large extent of velocities. This extraordinary biological machinery is moreover capable of achieving fast performance despite its intrinsic slow communication bandwidth.

Motion estimation is a task performed very well by the primate visual cortex. But motion estimation is also a key feature for many vision and robotic applications, and it is an active field of research in the computer vision community. Indeed, there is still a strong need for more accurate and efficient methods, and

the enthusiasm around the estimation of optical flow is clear from the success of the recent benchmark proposed by [Baker et al \(2007\)](#).

In this article, our goal is to contribute towards bridging the gap between algorithmic and biological vision by proposing a bio-inspired motion estimation model based on a suitable mathematical framework. It is our conviction that a breakthrough can be reached by understanding which mechanisms make the visual system so robust when it deals with a large variety of data under changing environmental conditions.

We based our approach on three main ideas. First, the design of our model mimics the functional properties of the main visual cortex layers dedicated to motion. Second, the mathematical framework chosen is well posed and suitable to model cortical layers activity at a macroscopic scale. Third, we show how the evaluation of motion estimation models can be further extended, in the light of the available experimental data in neuroscience.

The article is organised as follows. In [Section 2](#) we revisit the main properties of the visual system focusing on the motion pathway. In [Section 3](#) we present our neural field model defined by a set of two coupled integral equations. We prove that our model is mathematically well-posed and we relate it with some classical approaches in computer vision (based on partial differential equations). In [Section 4](#) we present our results, obtained not only on videos from computer vision but also on a classical stimulus from psychophysical experiments. A step further, in [Section 5](#) we set the basis of a novel evaluation methodology in which motion estimation is compared to human performance. Finally, in [Section 6](#) we give the main perspectives related to this work.

2 The primate visual system

2.1 General considerations

Two main functional streams of visual information processing are classically distinguished in the literature: the *form pathway*, which processes static features; the *motion pathway*, which concerns motion integration and segmentation. Both pathways receive their input from the retina through the lateral geniculate nucleus (LGN) and the primary visual cortex (V1). This segregated organisation is rooted in a similar dichotomy found at the neuronal level where parvocellular and magnocellular neurons exhibit different spatio-temporal bandwidths, colour preferences and luminance contrast sensitivities as well as different conduction times ([Born, 2001](#); [Nassi and Callaway, 2009](#)). The focus of the present model is the cortical motion processing that depends primarily (but not exclusively, see [Nassi and Callaway \(2006\)](#)) upon the inputs for the motion pathway.

The primate visual cortex can be seen as set of densely connected aggregates of neurons known as *cortical maps*. Cortical maps form a highly connected hierarchy with forward and backward streams of information. Each of these cortical maps has been identified as processing a specific information such as texture, colour, orientation or motion. Most of these cortical maps are retinotopically organised: adjacent neurons have receptive fields that cover slightly different, but overlapping portions of the visual field. This spatial organisation of the neuronal responses to visual stimuli generally preserves the topology of the visual input, and cortical maps can be seen as a function of the visual input. The

primary visual cortex (V_1) feeds higher order maps distributed within several extra-striate areas spanning both parietal and temporal lobes of the neo-cortex. The information is transmitted to subsequent cortical maps by convergent forward connections. Note that higher order maps correspond to a coarser analysis of the visual input since they integrate more information by cumulative forward connections. Although much less is known about the properties and role of feedback projections, recent evidence has been accumulated to suggest that feedback indeed plays a crucial role in processing the visual information through context-dependent, modulatory effects or long distance diffusion (Sillito et al, 2006).

2.2 The motion pathway

In the present paper we focus on two specific cortical maps known as V_1 and MT. They both play a crucial role in estimating local motion and computing pattern motion properties. Moreover, their main characteristics are among the most examined in neuroscience, offering an enormous bulk of experimental data at different scales, from single neurons to behaviour. In this section, our goal is to briefly review some of the main features of these maps, emphasising their functional properties and connectivities (Escobar, 2009; Bradley and Goyal, 2008; Carandini et al, 2005; Born and Bradley, 2005).

The majority of inputs to MT comes from the V_1 cortical map, particularly from its layer 4B (Born and Bradley, 2005). While the exact computational rules are still disputed, there is a general view that V_1 and MT implement the forward stream of a two stage motion integration (Born and Bradley, 2005). Recently, it was found that some V_1 complex cells exhibit some speed tuning, although these local estimates of target speeds are of large bandwidth. Lastly, V_1 neurons are highly orientation selective. Both strong orientation selectivity and small receptive field size make V_1 neurons particularly sensitive to the aperture problem.

Another important feature is that MT has many feedback connections to V_1 (Rockland and Knutson, 2000). There are experimental data supporting the view that such feedback play a crucial role in context-dependent processing by shaping centre-surround interactions within V_1 (Hupé et al, 1998; Sillito et al, 2006; Angelucci and Bullier, 2003). Feedback connections are also much faster than the horizontal ones (2–6 m/s versus 0.1–0.2 m/s) according to Grinvald et al (1994). Thus, feedbacks from MT to V_1 are a significant feature to take into account in models.

2.3 From biology to bio-inspired models?

Most models aiming to reproduce motion integration mechanisms are two-stage feed-forward models where V_1 acts as a local-motion detector, and MT implements motion integration by pooling local motion cues (Movshon et al, 1985; Wilson et al, 1992; Simoncelli and Heeger, 1998; Rust et al, 2006). However, these models ignore two essential properties of motion integration.

The first property is that motion integration is fundamentally a spatial process. Various non-ambiguous motion cues need to be integrated and segregated to propagate motion information inside surfaces (Hildreth, 1983; Nakayama and Silverman, 1988; Grzywacz and Yuille, 1991; Weiss and Adelson, 2000). This

spatial property of motion integration has only been investigated in a few bio-inspired models (Grossberg and Mingolla, 1985; Weiss and Adelson, 1998; Bayerl and Neumann, 2004; Escobar et al, 2009).

The second property is that motion integration is fundamentally a dynamical process. When presented with line drawings, plaids or barber poles, the perceived motion direction shifts over time (Yo and Wilson, 1992; Castet et al, 1993; Shiffrar and Lorenceau, 1996). Similar dynamics can be found in smooth pursuit eye movements (Masson and Stone, 2002; Wallace et al, 2005), and reflect neural time courses (Pack and Born, 2001; Pack et al, 2004; Smith et al, 2005). These dynamics however are only beginning to be investigated by modellers such as Montagnini et al (2007); Tlapale et al (2010b).

In this article, our model has two recurrently connected stages. Since we want to consider the dynamics of the processing from a mathematical as well as behavioural perspective, the model is written as a dynamical system of neural field equations. The dynamics is the distinguishing feature without which comparisons to biology would be highly difficult. As a whole, in the context of motion estimation and up to our knowledge, this is the first contribution to propose a bio-inspired model based on the neural field formalism to handle real dynamical scenes.

3 A neural field model for motion estimation

3.1 Biological inspiration of the model

Figure 1 describes the general structure of our model. Given an input k_1 , motion is estimated and integrated at two different spatial scales within two maps (p_1, p_2) that are recurrently interconnected. The connectivity rules between these two maps are written as a set of coupled integral equations (through forward and backward connections). This functional structure is inspired by the biology as discussed below.

The input to our system (denoted by k_1) corresponds to a local motion estimation. Here it is based on a measure of correlations between frames (modified Reichardt detectors, see Section 3.2.2).

The first layer of our model (denoted by p_1) computes local direction and speed of motion. This corresponds to complex cells in primary visual cortex that have been shown to perform local velocity computation (Priebe et al, 2006).

The second layer of our model (denoted by p_2) integrates motion over larger portions of the image and the information is propagated back to the first layer. This corresponds to MT cell properties. Our MT-like cells have larger receptive fields and are tuned to lower spatial frequencies and higher speeds than V1 cells. This fact is consistent with the view that V1 and MT stages operate at different scales (Born and Bradley, 2005). Feed-forward models of motion integration are heavily rooted on such evidence (Simoncelli and Heeger, 1998; Rust et al, 2006; Wilson et al, 1992; Löffler and Orbach, 1998). However, V1 and MT are recurrently interconnected (Sillito et al, 2006) and existing models have shown that such a recurrent connectivity can play a role in solving the aperture problem in synthetic and natural sequences (Chey et al, 1997; Bayerl and Neumann, 2004), as well as implementing contextual effects observed in V1 and MT neurons (Angelucci and Bullier, 2003).

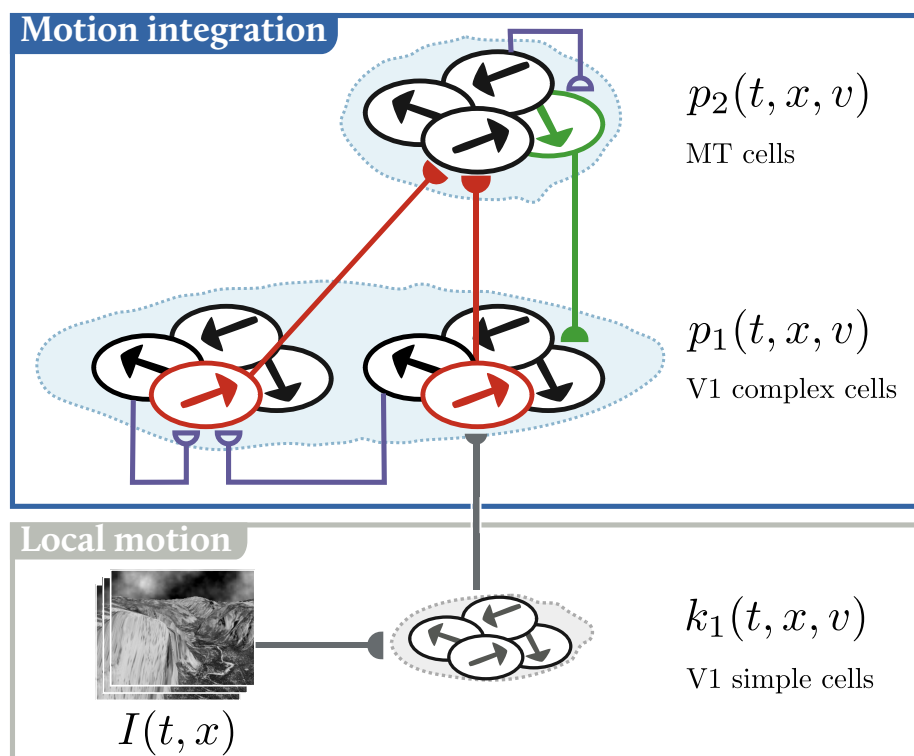


Figure 1: A schematic view of the model showing the interactions of the different cortical layers. From a grey-level input, our model estimates local motion information in k_1 , which is then used in a feed-forward (*red*) / feedback (*green*) loop between p_1 and p_2 . There are also lateral connections in each layer (*blue*).

Beyond the similarities concerning the functional structure, another major innovative aspect of our contribution is to propose a truly dynamical model. More precisely, we are interested in the evolution of motion estimation between frames so that comparison with neural and behavioural time courses becomes possible.

3.2 Description of maps interactions

3.2.1 The neural field framework

One major difficulty in observing or modelling the brain is that its study involves a multi-scale and thus is a multi-disciplinary analysis. As noted by Churchland and Sejnowski (1992), there is a large variety of scales but there is no integrative model yet covering all of them at once. Thus, one has to choose a given scale and define a suitable mathematical framework for that scale.

In this article we consider the macroscopic scale, defining the cortical activity at the population level, for the number of neurons and synapses even in a small piece of cortical area is immense. In order to describe cortical activity at the population level, neural field models are proposed as a continuum approximation of the neural activity.

Since the seminal work by Wilson and Cowan (1972, 1973) and Amari (1977), intensive research has been carried out to extend models and study them mathematically. The general mathematical study of such neural field equations can be very complex and it is still a challenging field of research (Ermentrout, 1998; Coombes, 2005; Faye and Faugeras, 2009; Veltz and Faugeras, 2010).

Our goal is to use this formalism for the problem of motion estimation. Following the general structure described in Section 3.1, the two maps describing the firing rate activity of a continuum of neuron populations are denoted by

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

for $i \in \{1, 2\}$, where Ω is the spatial domain (a bounded open subset of \mathbb{R}^2) and $\mathcal{V} \subseteq \mathbb{R}^2$ is the velocity space (the space of possible velocities). $p_i(t, x, v)$ is the average activity of the population corresponding to position x and tuned to velocity v (Pinto et al, 1996).

The general neural equation for an activity based model is:

$$\begin{aligned} \frac{\partial \mathbf{p}}{\partial t}(t, r) = & -\Lambda \cdot \mathbf{p}(t, r) \\ & + \mathbf{S} \left(\int_{\Omega \times \mathcal{V}} \mathbf{W}(t, r, r') \mathbf{p}(t, r') dr' + \mathbf{K}(t, r) \right), \end{aligned} \quad (2)$$

where $\mathbf{p} = (p_1, p_2)^T$, $r = (x, v)$ characterises the population (position and velocity tuning), $\Lambda = \text{diag}(\lambda_1, \lambda_2)$ is a matrix describing the temporal dynamics of the membrane potential, and $\mathbf{S}(x) = (S_1(x_1), S_2(x_2))^T$ is a matrix of sigmoidal functions (defined by $S_i(s) = 1/(1 + e^{-s})$). \mathbf{K} is an external current that models external sources of excitations (in our case, $\mathbf{K} = (\lambda_1^f k_1, 0)^T$ since there is no external input to map p_2). More importantly, $\mathbf{W}(t, r, r')$ describes how the population r' (at position x' and tuned to the velocity v') influences the population r at time t .

In the right-hand side of equation (2), the first term denotes the passive activity decay (with rate $\lambda_{1,2}$) when the input features to the target population

is switched off. The second term denotes the cells activation functions ($S_{1,2}$), a non-linear transformation of the input.

Before giving more details on our model, let us mention three general interesting properties of the neural field formalism. First, the integral definition of the weights allows large extent connectivity. Second, the sigmoid provides a tool to study contrast-gain effects. Finally, delays can be incorporated (Veltz and Faugeras, 2010; Deco and Roland, 2010) to constrain the connectivity.

3.2.2 Local motion estimation

The initial stage of every motion processing system is the computation of local motion 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 and Sperling, 1985; Watson and Ahumada, 1985; Adelson and Bergen, 1985).

Starting from the input image sequence $I : (t, x) \in \mathbb{R}^+ \times \Omega \rightarrow I(t, x)$, we estimate the local motion k_1 using modified Reichardt detectors (Bayerl and Neumann, 2004) enhanced to support subpixel velocities estimation. Two filtered images are correlated to estimate population activity: Directional derivatives are used to filter the input:

$$c_1(t, x, \alpha) = \frac{I(t, x) \overset{x}{*} \partial_\alpha^2 G_\sigma}{\varepsilon + \sum_{\beta \in \mathcal{O}} |I(t, x) \overset{x}{*} \partial_\beta^2 G_\sigma| \overset{x}{*} G_\sigma},$$

where ε avoids division by zero, G_σ denotes a Gaussian kernel, σ 's are scaling constants, $\overset{x}{*}$ denotes the convolution operator in space and ∂_α^2 denotes the second order directional derivative in the direction $\alpha \in \mathcal{O}$.

From these filtered outputs, we defined the half detectors by correlation with another frame:

$$c_2^+(t, x, v) = \left(\sum_{\alpha \in \mathcal{O}} c_1(t, x, \alpha) c_1(t+1, x+v, \alpha) \right) \overset{x}{*} G_\sigma,$$

$$c_2^-(t, x, v) = \left(\sum_{\alpha \in \mathcal{O}} c_1(t+1, x, \alpha) c_1(t, x+v, \alpha) \right) \overset{x}{*} G_\sigma,$$

where σ 's are scaling constants. The half detectors are then combined by:

$$k_1(t, x, v) = \frac{|c_2^+(t, x, v)|_+ - \frac{1}{2}|c_2^-(t, x, v)|_+}{1 + |c_2^-(t, x, v)|_+},$$

where $|x|_+ = \max(0, x)$ is a positive rectification, for the activity of neurons is always positive.

3.2.3 Core equations

The core of our model is defined by the interaction between the two populations, p_1 and p_2 , as described in (2). More precisely, we propose the following model

$$\begin{aligned} \frac{\partial p_1}{\partial t}(t, r) = & -\lambda_1 p_1(t, r) \\ & + S_1 \left(k_1(t, r) (\lambda_1^f + \lambda^b p_2(t, r)) \right. \\ & \quad - \lambda_1^l G_{\sigma_1^l} \overset{x}{*} \int_{\mathcal{V}} p_1(t, x, w) dw \\ & \quad \left. + \lambda_1^d (G_{\sigma_1^d} \overset{x,v}{*} p_1(t, r) - p_1(t, r)) \right), \end{aligned} \quad (3)$$

$$\begin{aligned} \frac{\partial p_2}{\partial t}(t, r) = & -\lambda_2 p_2(t, r) \\ & + S_2 \left(\lambda_2^f G_{\sigma_2} \overset{x}{*} p_1(t, r) \right. \\ & \quad - \lambda_2^l G_{\sigma_2^l} \overset{x}{*} \int_{\mathcal{V}} p_2(t, x, w) dw \\ & \quad \left. + (G_{\sigma_2^d} \overset{x,v}{*} p_2(t, r) - p_2(t, r)) \right), \end{aligned} \quad (4)$$

where $r = (x, v)$ denotes the characteristic of the population (position and velocity tuning), λ 's and σ 's are constants and G_σ denote Gaussian kernels defined by

$$G_\sigma(x) = \frac{1}{2\pi\sigma^2} e^{-\frac{|x|^2}{2\sigma^2}},$$

when the convolution is in the spatial domain only, and

$$G_\sigma(x, v) = \frac{1}{4\pi^2\sigma_x^2\sigma_v^2} e^{-\frac{|x|^2}{2\sigma_x^2} - \frac{|v|^2}{2\sigma_v^2}}, \quad \text{with } \sigma = (\sigma_x, \sigma_v) \quad (5)$$

when the convolution is in the spatial and velocity domains.

One interesting property of this model is that it naturally performs a multiscale analysis of motion through the exchanges between the two populations: (i) The feed-forward input from the previous layer (p_1) is integrated at the level of p_2 using a Gaussian weighting function, thus implementing the V1-to-MT convergence of connectivities ($\lambda_2^f G_{\sigma_2} \overset{x}{*} p_1$), and (ii) the feedback signals are modulatory to yield a correlative enhancement when feed-forward and feedback coincide, while the feed-forward input is left unchanged when no feedback is delivered (term $k_1(\lambda_1^f + \lambda^b p_2)$). Note that the feedback is written in a multiplicative way (as in Bayerl and Neumann (2004)): We used a modulating feedback rather than driving feedback, similar to that found in studies of the motion processing system in primates (Sillito et al, 2006).

For both equations a selection mechanism is defined by the term $-\lambda G_\sigma \overset{x}{*} \int_{\mathcal{V}} p_i(t, x, w) dw$. Such short-range lateral inhibition, usually called recurrent inhibition, leads to a winner-take-all mechanism (Dayan and Abbott, 2001; Yuille and Grzywacz, 1989). Instead of a divisive inhibition found in some models (Nowlan and Sejnowski, 1994; Bayerl and Neumann, 2004), we implemented a subtractive inhibition in order to fit into the neural field formalism.

Finally, both equations incorporate a diffusion term defined by $G_{\sigma_i^d} \overset{x,v}{*} p_i - p_i$. As shown in Lemma 3.1 this term behaves asymptotically like a Laplacian operator.

Lemma 3.1. *Given a Gaussian kernel G_σ as defined in (5) with $\sigma = (\rho, \eta)$, let us denote*

$$A = G_\sigma \overset{x,v}{*} p(x, v) - p(x, v), \quad (6)$$

then we have

$$A = \frac{\rho^2}{2} \sqrt{\pi} D_x^2 p(x, v) + \frac{\eta^2}{2} \sqrt{\pi} D_v^2 p(x, v) + o(\rho^2, \eta^2, \rho\eta). \quad (7)$$

where D_x^2, D_v^2 denote the Laplacian operator in the physical space and the velocity space.

Proof. Rescaling inside the integral in (6) we get

$$A = \frac{1}{\pi^2} \int_{\mathbb{R}^4} e^{-|y|^2 - |w|^2} (p(x - \rho y, v - \eta w) - p(x, v)) dy dw.$$

Then using a Taylor expansion of p (and assuming that $p \in \mathcal{C}^3(\mathbb{R}^4)$), we obtain

$$\begin{aligned} A = & \frac{1}{\pi} \int_{\mathbb{R}^4} \exp(-|z|^2) \left[-\rho D_x p(x, v) \cdot x - \rho D_v p(x, v) \cdot v \right. \\ & + \frac{\rho^2}{2} D_x^2 p(x, v)(x, x) + \frac{\eta^2}{2} D_v^2 p(x, v)(x, x) \\ & + \frac{\rho\eta}{2} D_{xv}^2 p(x, v)(x, v) \\ & - \frac{\rho^3}{6} D_x^3 p(x - \rho\theta_1 y, v)(y, y, y) \\ & - \frac{\rho^2\eta}{6} D_x^2 D_v p(x - \rho\theta_2 y, v - \eta\theta_3 w)(y, y, w) \\ & - \frac{\rho\eta^2}{6} D_x D_v^2 p(x - \rho\theta_4 y, v - \eta\theta_5 w)(y, w, w) \\ & \left. - \frac{\eta^3}{6} D_v^3 p(x, v - \eta\theta_6 w)(w, w, w) \right] dy dw, \end{aligned}$$

where $z = (y, w)$, $\theta_i = \theta_i(x, v, \rho, \eta, y, w)$ belong to $(0, 1)$. But thanks to the moment conditions,

$$\begin{aligned} \int_{\mathbb{R}^2} \exp(-|z|^2) dz &= \pi, \\ \int_{\Omega} z_i \exp(-|z|^2) dz &= 0, \\ \int_{\mathbb{R}^2} z_i z_j \exp(-|z|^2) dz &= 0 \quad (i, j = 1, 2, i \neq j), \\ \int_{\mathbb{R}^2} z_i^2 \exp(-|z|^2) dz &= \frac{\pi\sqrt{\pi}}{2}, \end{aligned}$$

and we finally obtain (7). \square

\square

3.3 Existence and uniqueness of the solution

In order to study the well-posedness of our model (3)–(4), let us consider the results presented in Faugeras et al (2008), for neural field equations. Note

that in our case, the equations (3)–(4) do not exactly fit in the neural field formalism since the time-dependent input is used in a multiplicative way $k_1 p_2$ inside the sigmoid. As described in the previous section this term implements a modulating feedback diffusion. By applying the Cauchy-Lipschitz Theorem we show that the addition of such a multiplicative term to an activity-based neural field model maintains its well-posedness properties. First we check that the assumptions of the theorem are satisfied (Lemma 3.2 and Lemma 3.3). Then, since the theorem proves existence and uniqueness of the solution on an open and bounded time interval, we show that this interval can be extended to the full half real line using a continuity argument (Theorem 3.2)

Let \mathcal{F} be the set $\mathbf{L}^2(\Omega \times \mathcal{V})$ of square integrable functions defined on the product set $\Omega \times \mathcal{V}$ and taking their values in \mathbb{R} , and $\mathcal{F} = \mathcal{F} \times \mathcal{F}$. The basic idea is to rewrite (3)–(4) as a differential equation defined on the set \mathcal{F} . With a slight abuse of notation we can write $p_i(t)(x, v) = p_i(t, x, v)$ and note $\mathbf{p} : \mathbb{R} \rightarrow \mathcal{F}$ the function defined by the following Cauchy problem:

$$\mathbf{p}(0) = \mathbf{p}_0 \in \mathcal{F}, \quad (8)$$

$$\mathbf{p}' = -\Lambda \mathbf{p} + \mathbf{S}(\mathbf{W}(t) \cdot \mathbf{p} + \mathbf{K}(t)), \quad (9)$$

with $\mathbf{p} = (p_1, p_2)^\top$, $\mathbf{K} = (\lambda_1^f k_1, 0)^\top$, and $\mathbf{S}(x_1, x_2) = (S_1(x_1), S_2(x_2))$. The operator \mathbf{W} is the 2×2 connectivity matrix function defined by the four linear mappings from \mathcal{F} to \mathcal{F} :

$$W_{11} \cdot p = -\lambda_1^l G_{\sigma_1^l}^{x,v} * p + \lambda_1^d (G_{\sigma_1^d}^{x,v} * p + p),$$

$$W_{12} \cdot p = \lambda^b k_1 p,$$

$$W_{21} \cdot p = \lambda_2^f G_{\sigma_2^f} \delta_v^x * p,$$

$$W_{22} \cdot p = -\lambda_2^l G_{\sigma_2^l}^{x,v} * p + \lambda_2^d (G_{\sigma_2^d}^{x,v} * p + p).$$

Functionally W_{11} and W_{22} correspond to lateral interactions in maps v1 and MT, W_{12} denotes the backward connection from MT to v1, and W_{21} denotes the forward integration from v1 to MT. In the following we note f the mapping defined by the right-hand side of (9):

$$f(t, \mathbf{p}) = -\Lambda \mathbf{p} + \mathbf{S}(\mathbf{W}(t) \cdot \mathbf{p} + \mathbf{K}(t)).$$

Note that the time dependence in the definition of f arises solely from the function k_1 that occurs in W_{12} and in \mathbf{K} . We prove the existence and uniqueness of a solution to (9) by proving (i) that f maps $I \times \mathcal{F}$ to \mathcal{F} where I is an open interval containing 0 and (ii) that it is Lipschitz continuous with respect to the second variable. This allows us to apply the Cauchy-Lipschitz Theorem and to conclude that there is a unique maximal solution to (9), and that its interval of definition is an open interval $(-\alpha, \alpha)$ containing 0.

Lemma 3.2. *If $k_1(t)$ is measurable for all $t \in I$, f maps $I \times \mathcal{F}$ to \mathcal{F} .*

Proof. Let $\mathbf{p} = (p_1, p_2) \in \mathcal{F}$. If $k_1(t)$ is measurable for all $t \in I$, so is $W_{12} \cdot p_2$. All the other elements of $\mathbf{W} \cdot \mathbf{p}$ are simple or weighted sums (convolutions) of a measurable function \mathbf{p} and thus $\mathbf{W} \cdot \mathbf{p}$ is measurable. This implies that $\mathbf{S}(\mathbf{W}(t) \cdot \mathbf{p}(t) + \mathbf{K}(t))$ is in \mathcal{F} for all $t \in I$. \square \square

Lemma 3.3. *If $k_1(t)$ is measurable on $\Omega \times \mathcal{V}$ and bounded by \bar{k}_1 for all $t \in I$ the mapping f is Lipschitz continuous with respect to the second variable.*

Proof. We have

$$\begin{aligned} \|f(t, \mathbf{p}) - f(t, \mathbf{q})\| = \\ \| -\Lambda(\mathbf{p} - \mathbf{q}) + \mathbf{S}(\mathbf{W}(t) \cdot \mathbf{p} + \mathbf{K}(t)) - \mathbf{S}(\mathbf{W}(t) \cdot \mathbf{q} + \mathbf{K}(t)) \| \leq \\ \max(\lambda_1, \lambda_2) \|\mathbf{p} - \mathbf{q}\| + S'_m \|\mathbf{W}(t) \cdot (\mathbf{p} - \mathbf{q})\|, \end{aligned}$$

where S'_m is the maximum value taken by the derivatives of the sigmoids S_1 and S_2 . $\|\mathbf{W}(t) \cdot (\mathbf{p} - \mathbf{q})\|$ is upper-bounded by a constant times the sum of the four terms $\|W_{ij} \cdot (p_j - q_j)\|_{\mathcal{F}}$, $i, j = 1, 2$. Considering these terms we find two cases. The first case involves a convolution by a Gaussian is easily dealt with since:

$$\|G^{x,v} * p\|_{\mathcal{F}} \leq k \|p\|_{\mathcal{F}} \quad \forall p \in \mathcal{F},$$

where the constant k depends on the Gaussian kernel. The second case concerns the multiplication by $k_1(t)$ in $W_{12}(t)$. Because of the hypothesis $k_1(t)p_2$ belongs to \mathcal{F} for all $t \in I$ and $\|k_1(t)p_2\| \leq \bar{k}_1 \|p_2\|_{\mathcal{F}}$. This completes the proof that f is Lipschitz continuous with respect to the second variable. \square \square

Theorem 3.1. *If $k_1(t)$ is measurable on $\Omega \times \mathcal{V}$ and bounded by \bar{k}_1 for all $t \in I$ there exists an open interval $J = (-\alpha, \alpha) \subset I$ centred at 0 such that the Cauchy problem (8-9) has a unique solution, hence is in $\mathcal{C}^1(J, \mathcal{F})$.*

Proof. Thanks to Lemmas 3.2 and 3.3 the conditions of the Cauchy-Lipschitz Theorem are satisfied. \square \square

Then, thanks to the sigmoids, it is easy to show that this solution is bounded.

Proposition 3.1. *The solution described in Theorem 3.1 is bounded for all $t \in J$*

Proof. The variation of constant formula yields:

$$\mathbf{p}(t) = e^{-\Lambda t} \mathbf{p}_0(t) + \int_0^t e^{-\Lambda(t-s)} \mathbf{S}(\mathbf{W}(s) \cdot \mathbf{p}(s) + \mathbf{K}(s)) ds,$$

for $t \in J$, from which it follows that

$$\begin{aligned} \|\mathbf{p}(t)\| &\leq \|e^{-\Lambda t}\| \|\mathbf{p}_0\| + \\ &\quad \left\| \int_0^t e^{-\Lambda(t-s)} \mathbf{S}(\mathbf{W}(s) \cdot \mathbf{p}(s) + \mathbf{K}(s)) ds \right\| \\ &\leq e^{\max(\lambda_1, \lambda_2)\alpha} \|\mathbf{p}_0\| + \max\left(\frac{S_{1m}}{\lambda_1}, \frac{S_{2m}}{\lambda_2}\right) (e^{\max(\lambda_1, \lambda_2)\alpha} - 1) \\ &\leq e^{\max(\lambda_1, \lambda_2)\alpha} \left(\|\mathbf{p}_0\| + \max\left(\frac{S_{1m}}{\lambda_1}, \frac{S_{2m}}{\lambda_2}\right) \right), \end{aligned}$$

where S_{1m} and S_{2m} are the maximum values of the sigmoid functions S_1 and S_2 . \square \square

As in [Faugeras et al \(2008\)](#) we can extend this local result from $(-\alpha, +\alpha)$ to $(-\alpha, +\infty)$, assuming that the hypotheses on \mathbf{p}_0 in [Theorem 3.1](#) are satisfied for $t \in (-\alpha, +\infty)$. Indeed, either $+\alpha = +\infty$ and the result is proved or there exists $0 < \beta < \alpha$ such that \mathbf{p} is not bounded for all $\beta \leq t < \alpha$, thereby obtaining a contradiction.

We summarise these results in the following theorem:

Theorem 3.2. *If $k_1(t)$ is measurable on $\Omega \times \mathcal{V}$ and bounded by \bar{k}_1 for all $t \in (-\alpha, +\infty)$ the Cauchy problem (8-9) has a unique bounded solution, hence in $\mathcal{C}^1((-\alpha, +\infty), \mathcal{F})$.*

3.4 Relations to the state of the art in computer vision

One essential aspect of the neural field framework lies in the definition of interaction between populations through an integral form. Interestingly, under some assumptions, one can write relations between integral operators (acting in a neighbourhood) and differential operators (acting very locally). This question was investigated by [Degond and Mas-Gallic \(1989\)](#); [Edwards \(1996\)](#); [Cottet and Ayyadi \(1998\)](#) and further extended by [Viéville et al \(2007\)](#). In these papers, the authors show the correspondence between linear elliptic differential operators and their integral approximation. This idea has also been considered for nonlinear operators by [Buades et al \(2006\)](#); [Aubert and Kornprobst \(2009\)](#). Thus, one can see a direct relation between the neural field framework and PDE-based approaches.

As such, introducing the neural field framework for motion estimation can be related to the series of papers proposing PDE-based approaches for optical flow estimation, starting from [Horn and Schunck \(1981\)](#). In computer vision, this seminal work has been further improved by many authors such as [Enkelmann \(1988\)](#); [Black and Rangarajan \(1996\)](#); [Weickert and Schnörr \(2001\)](#); [Nir et al \(2008\)](#). Improvements concern mainly the definition of the regularisation term, which is how diffusion performs. In this class of approaches, since diffusion is defined by differential operators, the aperture problem is solved by local diffusion processes.

Here, using the neural field framework, we offer the possibility to define different kinds of connectivity patterns not necessarily corresponding to differential operators. More generally, for modelling in computer vision, the neural field formalism has two main advantages over PDE-based approaches: (i) The first advantage is that non-local interactions can be defined, which is not possible with classical PDE or variational approaches defining the interactions between neighbours through differential operators. (ii) The second advantage is to naturally describe interactions between several maps. In our article, the two maps correspond to two scales of analysis, thus providing a *dynamical* multiscale analysis.

4 Results

4.1 Implementation details

As far as implementation is concerned, spaces have to be discretised, including the velocity space \mathcal{V} . We chose $\mathcal{V} = \{-5, -4.5, \dots, 4.5, 5\}^2$ to sample the velocities on a grid of size 21×21 .

The model defined by equations (3)–(4) is fully specified by a set of fourteen parameters.¹ These parameters were tuned by matching the time scale dynamics of the simple translating bar stimulus presented latter in Figure 7. Integration was performed using a 4th order Runge–Kutta method with at most ten iterations between two frames.

Since our distributed motion representation can be hard to analyse, and to facilitate comparisons with computer vision approaches, we estimate an optical flow m_i by averaging at each position the population response across all velocities (Bayerl and Neumann, 2004):

$$m_i(t, x) = \frac{\sum_{v \in \mathcal{V}} p_i(t, x, v) v}{\sum_{v \in \mathcal{V}} p_i(t, x, v)}, \quad i \in \{1, 2\}. \quad (10)$$

Then the optical flow is represented either by arrows or by a colour coded image indicating speed and direction. We used the Middlebury colour code (Baker et al, 2007), which emerged as the *de facto* standard in the optical flow computer vision community. The direction of the velocity corresponds to the hue, for instance yellow for downward velocities, while the speed of the velocity is encoded in the saturation, whiter (or less saturation) for slower speeds. The colour code is illustrated in Figure 2a.

4.2 Natural scenes

We consider three classical videos from the computer vision (see Figure 2): (i) The Hamburg taxi sequence, a recorded sequence where three cars and a pedestrian are moving; (ii) Two videos for which the ground truth is available: the synthetic Yosemite sequence where the optical flow covers the whole spatial domain, and the rubber–whale sequence.

Results for the Hamburg taxi sequence are shown in Figure 3. We show how our model improves the initial noisy optical flow estimation obtained from the modified Reichardt detectors described in Section 3.2.2. The evolution of the optical flow between k_1 , p_1 and p_2 is shown in Figures 3a, 3b and 3c. Note that since the output of the Reichardt motion detectors is unable to handle the borders, we replace the information on the border by a small identical activity on all velocities. When the optical flow is computed, the replacement leads to a zero velocity on the borders represented in Figures 3a and 3b as a white frame. Due to the recurrent interactions in our model, those borders tend to be filled in when the motion is strong enough (such as the rightward moving car in Figure 3b). Such a filling in mechanism is particularly useful for dense optical flows since it nicely reconstructs the motion at the borders.

¹Parameters chosen for the experiments: $\lambda_1 = 2$, $\lambda_1^f = 1$, $\lambda_1^b = 24$, $\lambda_1^l = 4$, $\sigma_1 = 2$, $\lambda_2 = 2$, $\lambda_1^d = 6$, $\lambda_2^f = 16$, $\lambda_2^l = 4$, $\sigma_2 = 2$, $\sigma_2^f = 8$, $\lambda_2^d = 10$, $\sigma_1^d = 2$, $\sigma_2^d = 10$.

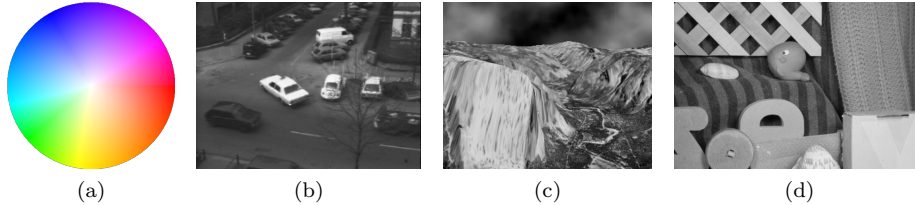


Figure 2: Experimental protocol. (a) Middlebury colour disk mapping motion direction to hue and speed to saturation, and tested videos: (b) Hamburg taxis sequence with three cars moving; (c) Yosemite sequence with clouds (Barron et al, 1994); (d) Rubber-whale sequence from the Middlebury database (Baker et al, 2007).

Results for the Yosemite sequence are shown in Figure 4. The optical flow m_2 estimated from p_2 is shown in Figure 4a, with its associated angular error in Figure 4b. The angular error is defined by:

$$\arccos \left(\frac{u_1 v_1 + u_2 v_2 + 1}{\sqrt{u_1^2 + u_2^2 + 1} \sqrt{v_1^2 + v_2^2 + 1}} \right)$$

where $u = (u_1, u_2)$ is the correct velocity and $v = (v_1, v_2)$ is the estimated velocity. In our case, the average error is 3.97° which is in the range of baseline results from Baker et al (2007). Note that this error evolves in time (see Figure 5) and that this average was estimated at convergence. Then it is important to mention that a large region of high angular error is located on subpixel velocities. One can explain such an error by the peculiar velocity space which offers poor angular resolution at low velocities. A coarser velocity space such as $\mathcal{V}' = \{-5, -4, \dots, 4, 5\}^2$ used in Tlapale et al (2010b) would lead to a worse angular error as shown in Figure 4d (average angular error 6.48°). To explain this, we show in Figure 4c the norm of the ground truth for the Yosemite sequence: A subpixel definition of the velocity space is necessary because of the continuity of the optical flow. The influence of the diffusion term on the smoothness of the solution is illustrated in Figure 4e where we set the diffusion to zero ($\lambda_{1,2}^d = 0$). Removal of the diffusion leads to spatial patches of selected velocities which increase the angular error. Motion within individual patches has the tendency to represent one selected motion direction. Consequently, the angular error is high inside those patches where only one velocity is activated, but low at their borders where multiple velocities simultaneously exist (see Figure 4f).

In Figure 5 we show the evolution of the average angular error (AAE) for each frame of the Yosemite sequence. We observe that the convergence is not reached between the first pair of frames (as in classical computer vision methods) simply because we limited the number of iterations between two frames. Indeed, we designed our model to reproduce motion integration dynamics and this exponential decay of the error is very important in psychophysics (see Section 4.3).

Results for the rubber-whale sequence are shown in Figure 6. In this sequence we obtain an average angular error of 10.40° (median 4.00°). The highest errors appear at occlusions, in particular inside the hole of the e -shaped object (see red arrow) where a maximal error of 105° is reached.

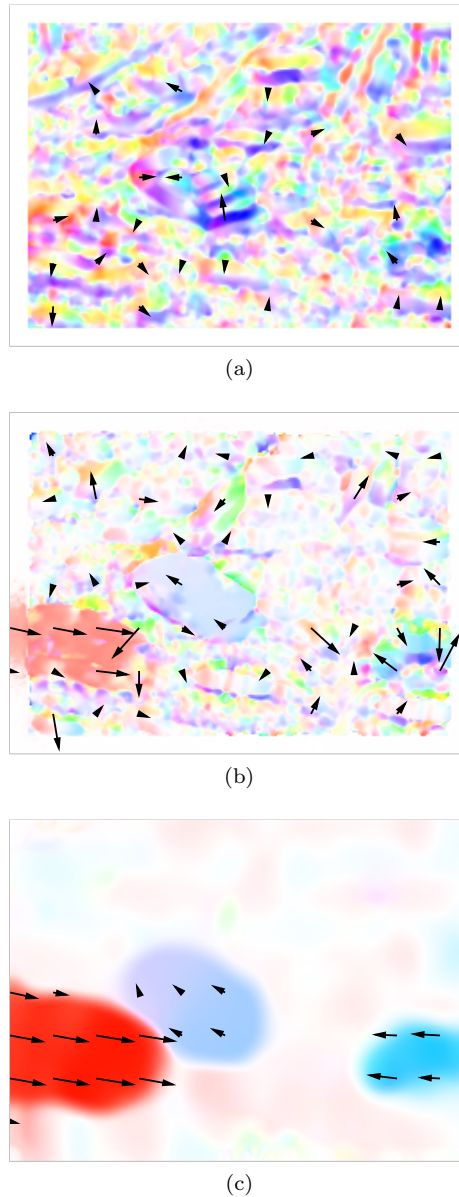


Figure 3: Result for the Hamburg taxi sequence. (a) Optical flow estimated from the Reichardt detectors k_1 . (b) Optical flow m_1 . (c) Optical flow m_2 . Note the filling in of the left margin due to the activity evoked by the leftward moving car.

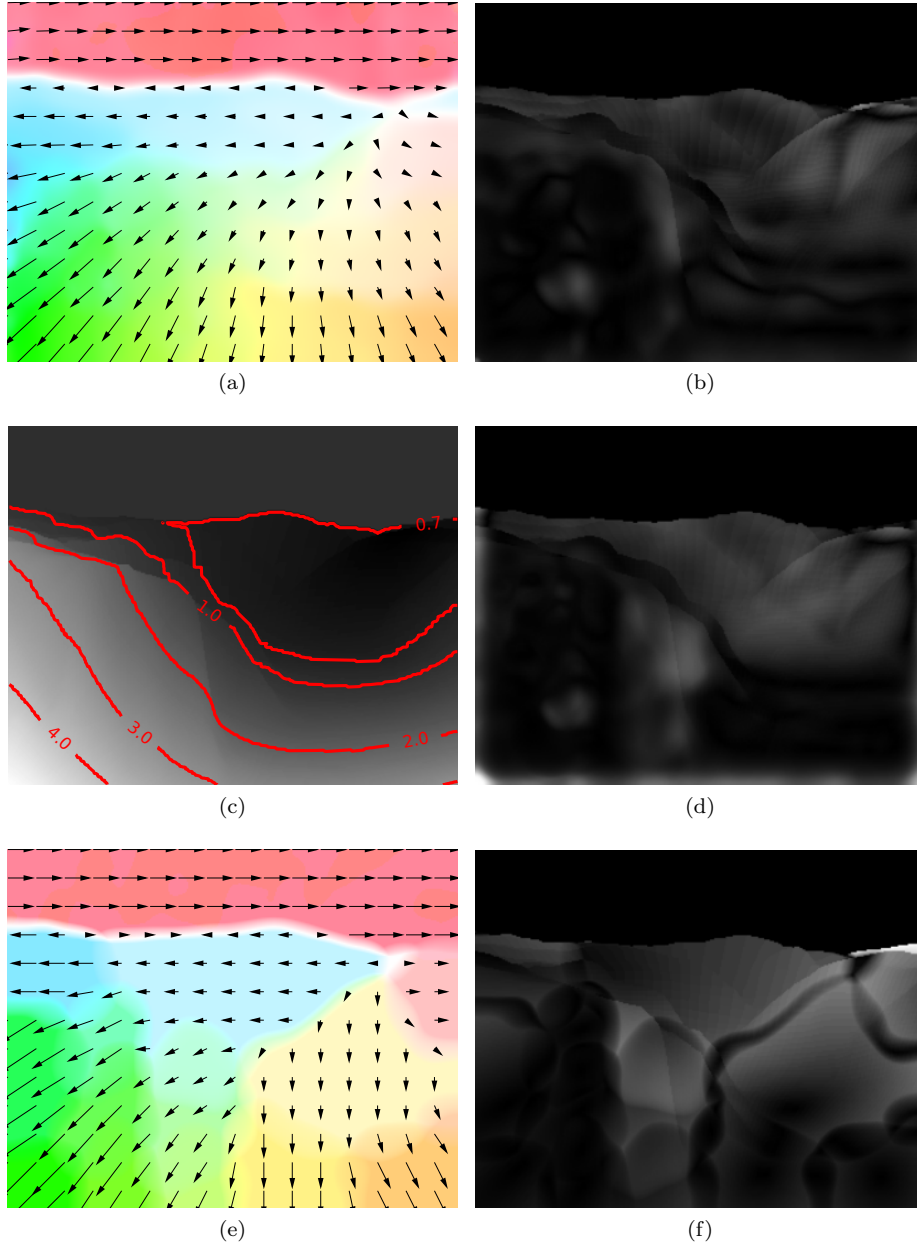


Figure 4: Results for the Yosemite sequence. (a) Optical flow m_2 . (b) Associated angular error. (c) Norm of the ground truth with iso-contours plotted for some values. (d) Same as b but with the coarser discrete velocity space \mathcal{V}' . (e) Same as b but when diffusion terms are removed: patches appear. (f) Angular error corresponding to the optical flow shown in e : the error is low at the borders of the patches.

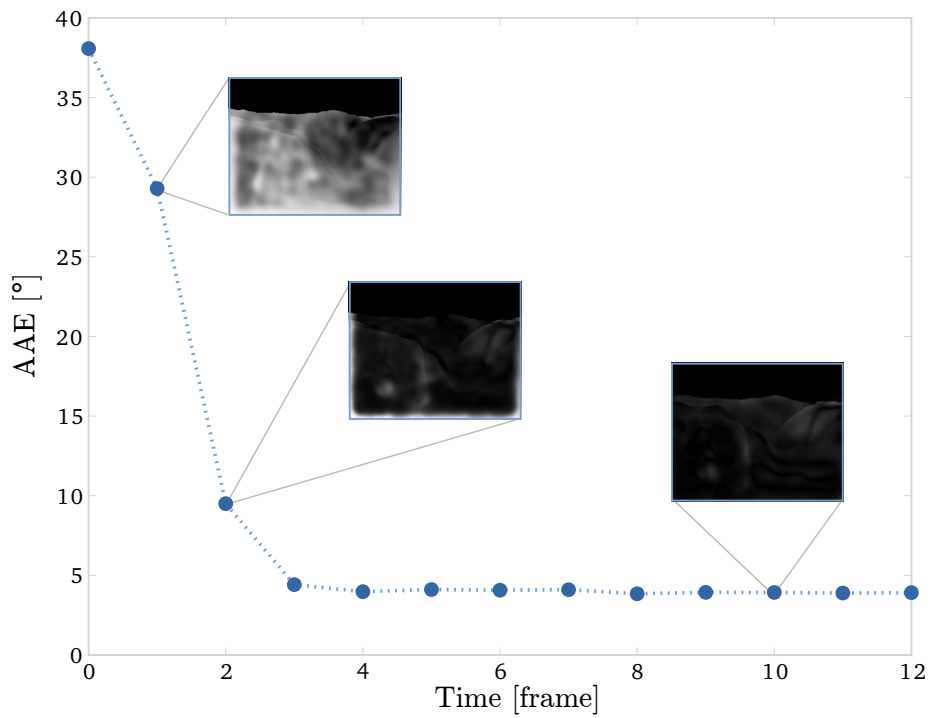


Figure 5: Dynamical evolution of the average angular error (AAE) on the Yosemite sequence.

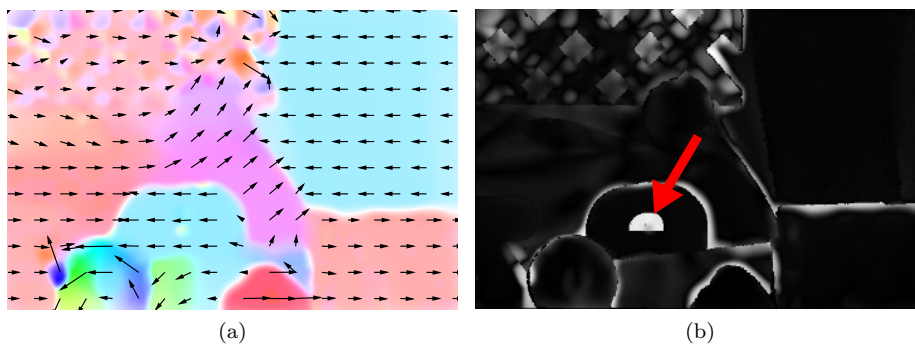


Figure 6: Rubber whale sequence. (a) Activity response in our p_2 area. (b) Angular error in p_2 (average is 10.40°).

4.3 Psychophysical stimulus

The dynamics of motion integration can be well characterised by a simple translating bar stimulus (see Figure 7a). It has been shown experimentally that the perceived direction is initially biased towards the direction orthogonal to the bar orientation, and that this perceptual bias is corrected for longer durations (Castet et al, 1993). This early bias was also shown in smooth pursuit for humans (Masson and Stone, 2002; Wallace et al, 2005) and monkeys (Born et al, 2006).

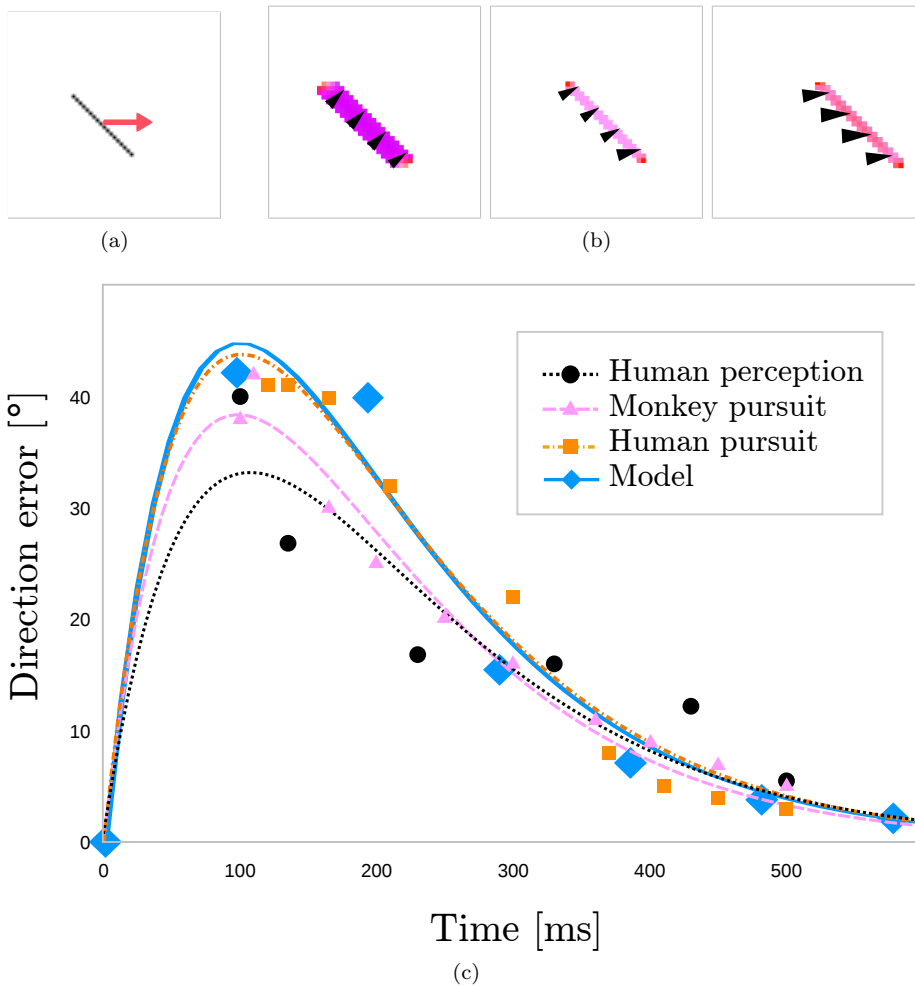


Figure 7: Results for the translating bar. (a) The stimulus is a rightward translating bar, tilted at 45° . (b) Optical flow m_2 at different time (100, 200 and 600 ms). (c) Temporal dynamics of the observed direction error for human perceived direction, human and macaque tracking direction, as well as our model. Data is reproduced from Lorenceau et al (1993); Wallace et al (2005); Born et al (2006). Both discrete measurements and best fit are shown (each data set was fitted by $f(t) = A \frac{x}{\tau} \exp(-\frac{x}{\tau})$).

In order to compare psychophysical dynamics to our results, let us define the perceived direction (from the population activity) as a read-out. The perceived direction $w(t) \in \mathbb{R}^2$ is defined as a unique velocity per frame and corresponding to the global motion. In this article, we defined it simply by averaging the activity of the population p_2 , with a temporal smoothing defined by the following dynamical equation:

$$\frac{dw}{dt}(t) = \lambda \left(\sum_{x \in \Omega} m_2(t, x) - w(t) \right), \quad (11)$$

where m_2 is defined by (10).

In Figure 7b, we show the optical flow m_2 that we obtained at different times (here we assume that the images are sampled every 100 ms). Then, using formula (11), we show in Figure 7c the estimated direction error defined by the angular difference between the perceived direction and the true direction of the object. After a short period of time where the direction error reaches 45° , the perceived direction converges to the true direction with an exponential decay. The dynamics that we observe in our model closely reproduce the experimental data measured for both pursuit and perception.

Note that the dynamics we observe here is not reproduced by any classical computer vision approach. Since their goal is generally to estimate the motion between two consecutive frames, the result is the same independently of the time at which the optical flow is estimated.

5 Towards a bio-inspired benchmark

5.1 Motivation

Following the results shown in Section 4.3 where we compared our results to biological data, we investigate how to set up a new kind of benchmark to evaluate models against visual system performance. Indeed, evaluating motion models by testing them only on realistic scenes (such as in Baker et al (2007)) is certainly not satisfactory if one claims that a model is bio-inspired. Evaluation methodology needs be reconsidered, in the light of available experimental data in neuroscience, in order to evaluate the bio-plausibility of a given model.

Such an evaluation methodology is very different from classical computer vision benchmarks where only flow fields are compared if possible against a ground truth. In the biological context, the notion of local motion does not make a lot of sense when considering the visual system performance since the purpose of the visual system is not to estimate a dense flow field. In addition, if we consider the class of all motion estimation models, there is a wide variety of possible motion representations: For example, the output can be described by global velocity likelihoods, velocity distributions at every position, filter responses, time-correlated spike trains, or 2D flow fields. Thus, we need to define a global readout such as the perceived motion $w(t)$ defined in (11), which is a suitable indicator in order to (i) compare output from models with observable quantities measured in neuroscience experiments and (ii) have a common representation to compare models.

Finally, one major difficulty to establish a benchmark based on human performance is the lack of ground truth. Contrary to computer vision where the

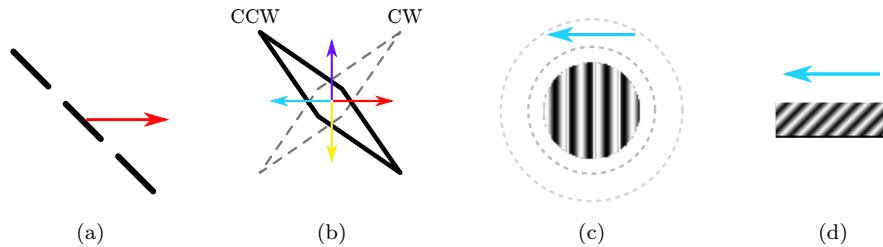


Figure 8: Bio-inspired benchmark: Database design. The proposed stimuli fit into two classes: line drawings and gratings. (a) Translating bar. (b) Translating diamond. (c) Grating size. (d) Barber pole.

ground truth is defined by the true velocity field, in psychophysical studies the notion of ground truth is impossible to define in a strict sense. For example, one has to handle the great variability between subjects or between trials for a single subject. The concision of data reported in the literature, often a mean and a standard deviation, does not allow the extraction of the statistical laws underlying the data. In addition, the set of experimental stimuli studied in neuroscience provides results at different levels. Given the diversity of the neuroscience experiments, capturing the main properties and results of motion estimation appears to be a complex task. For this reason we need to restrict our study to a set of fundamental questions.

Here we present the main ideas leading to such a bio-inspired benchmark and we refer the reader to Tlapale et al (2010a) for more details. Stimuli, scoring procedure and baseline results are also available online on the associated website:

<http://www-sop.inria.fr/neuromathcomp/psymotionbench>

5.2 Database Design

5.2.1 The two fundamental questions

In the proposed benchmark, we focus on two fundamental aspects of motion integration: (i) the respective influence between 1D *versus* 2D cues;² (ii) the dynamics of motion integration. We selected stimuli for which smooth pursuit eye movements and motion perception data were available to allow quantitative comparisons. The stimuli are shown in Figure 8 and described in the following section.

In this article we only present the static evaluation based on the solution at convergence and ignoring the dynamics. Indeed, studying the dynamics of motion integration is a criterion related only to biology, and to our knowledge there is no computer vision approach trying to reproduce the dynamical properties of motion integration. For more details about the dynamical evaluation, we refer the interested reader to Tlapale et al (2010a).

²1D cues refer to locations where the aperture problem cannot be solved (e.g., motion of straight edges) while 2D cues refer to locations where motion can be estimated unambiguously (e.g. motion of corners)

5.2.2 Line-drawing objects

Translating bars In [Lorenceanu et al \(1993\)](#); [Biber and Ilg \(2008\)](#) the authors consider tilted translating bars. Pursuing a translating bar whose true motion direction is not normal to its orientation leads to initial deviation in the *smooth pursuit* eye movement direction. To obtain a model evaluation procedure, the slope of the directional errors could be analysed with respect to bar length or number of bar tiles. Indeed, as the bar length is increased it becomes more complicated to recover its true direction. Likewise it is easier to pursue one long bar, if it is tiled into several sub bars ([Lorenceanu et al, 1993](#); [Biber and Ilg, 2008](#)). In the proposed experiment, we investigate the effect of the number of bars on the final percept.

Translating diamonds In [Masson and Stone \(2002\)](#) the authors consider diamond stimuli translating either vertically or horizontally. Due to the local orientations of the diamonds edges with respect to the translating direction, these stimuli mimic type II plaids. Indeed the vector average of the edge motions is biased $\pm 45^\circ$ away from the object's direction. The stimuli thus provide an interesting example to study the influence of 1D and 2D cues on motion integration.

Changing the configuration of the stimulus, by using clockwise (CW, stimulus main orientation is 45°) or counter-clockwise (CCW, stimulus main orientation is -45°) stimuli, or by varying the direction of the translation, does not influence the ability to pursue the translating diamonds. In all the cases, the initial pursuit direction as well as the fastest perceptual estimates are biased towards the vector average of the edge motions.

5.2.3 Gratings

Gratings sizes In [Barthélemy et al \(2006\)](#) the authors use a drifting grating viewed through a circular aperture. The orientation of the grating is constant and orthogonal to its drifting direction, but the diameter of the circular aperture varies among the stimuli. The authors quantify the change in eye direction during several time windows with respect to the diameter of the aperture (see dotted circles in [Figure 8c](#)). It is possible to look at the perceptual effects of such stimuli: varying sizes of grating patches affect motion detection as well as motion after effect. Many psychophysical studies have been conducted following the perceptual consequences of the centre-surround interactions in early visual areas (see [Seriès et al \(2001\)](#) for a review) and it becomes possible to compare these results for the properties of neuronal receptive fields in area V1, MT or MST in macaque monkeys.

Barber pole In the classical barber pole illusion, a translating grating is viewed through a rectangular aperture, leading to two orthogonal sets of 2D cues [Wallach \(1935\)](#). The larger set of 2D cues originates from the longest side of the rectangular aperture, while the smaller set of 2D cues originates from the shortest side. According to psychophysical experiments, as well as neurobiological data, the final perceived motion direction is the same as the orientation of the elongated side of the aperture, after an initial direction orthogonal to the





Approach	Avg.				
Our approach	1.00	1.00	1.00	1.00	1.00
SRDB-10	0.86	1.00	1.00	0.65	0.78
BM-10	0.74	1.00	1.00	0.00	0.98
BN-04	0.68	1.00	1.00	0.36	0.38
LK-81	0.45	0.81	0.00	0.99	0.00
HS-81	0.39	0.52	0.00	1.00	0.03
BMOCV	0.19	0.00	0.32	0.44	0.00

Table 1: Baseline results for our bio-inspired benchmark. Scores range between 0 (for low performance) and 1 (for high performance). Approaches are ranked according to their average score. We show the results obtained for our approach and then [Sun et al \(2010\)](#), [Brox and Malik \(2010\)](#), [Bayerl and Neumann \(2004\)](#), [Lucas and Kanade \(1981\)](#), [Horn and Schunck \(1981\)](#) and [Bradski \(2000\)](#).

grating orientation [Masson et al \(2000\)](#). The perceived motion direction thus corresponds to the 2D cues with the greater number of occurrences.

Again, similar observations are available at both psychophysical ([Castet et al, 1999](#); [Kooi, 1993](#)) and neuronal ([Pack et al, 2004](#)) levels. It is thus possible to compare model output with a global readout such as time-dependent ocular pursuit but also to compare the dynamics of single model neurons with that of V1 and MT neurons.

5.3 Baseline results

In Table 1 we present the baseline results with scores between 0 (for low performance) and 1 (for high performance). The approaches are ordered by their average score reported on the *Avg.* column. The full scoring procedure for each experiment is available online.

For example, let us explain how the score is obtained for the translating diamonds experiment. We start from the final perceived motions $w_s \in [0, 2\pi)$ estimated for each configuration $s \in \mathcal{S}$ with $\mathcal{S} = \{\text{up, down, left, right}\} \times \{\text{CW, CCW}\}$. Since the veridical motion \hat{w}_s is known, and because the experimental article provides quantitative data, we defined the score by

$$\text{score} = \frac{1}{|\mathcal{S}|} \sum_{\mathcal{S}} G_{\sigma}(w_s - \hat{w}_s),$$

where σ is defined by experimental data. The score is similarly defined for the other stimuli, considering final states with different stimuli configurations.

We applied our evaluation methodology to both biologically inspired artificial vision models ([Bayerl and Neumann, 2004](#); [Tlapale et al, 2010b](#)) and computer vision models ([Horn and Schunck, 1981](#); [Lucas and Kanade, 1981](#); [Sun et al, 2010](#); [Brox and Malik, 2010](#)) by running either the original implementation from the authors or the code that was available in the OpenCV library ([Bradski, 2000](#)). A single set of parameters was experimentally tuned in order to achieve the overall best score across all experiments.

As a general comment, it is interesting to remark that models performance somewhat follows research evolution. For example the seminal approaches for optical flow proposed by [Horn and Schunck \(1981\)](#) or [Lucas and Kanade \(1981\)](#) show quite a poor performance on most stimuli. The fact that these approaches are differential and not multi-scale largely explains this performance. Being differential, the optical flow is estimated based on the brightness consistency assumption, which is a local indication. Thus, when there is a majority of 1D cues, the input to differential algorithms is not very informative and leads to an aperture problem that is hard to solve numerically.

Considering multi-scale approaches is today one classical method to solve the aperture problem more efficiently. This solution is now used by most current models, such as the recent models by [Sun et al \(2010\)](#); [Brox and Malik \(2010\)](#), which are now among the best computer vision models (see the latest results online from [Baker et al \(2007\)](#)). Interestingly, those models also perform very well for most of our experiments.

Biologically inspired artificial vision models ([Bayerl and Neumann \(2004\)](#) and our approach) show high performance. In particular the model obtains the maximum score. The major strength of our model is that its design is naturally multi-scale as it is inspired from the multi-layer architecture of the brain cortical areas (V1 and MT) with proper connectivity patterns. This is one important conclusion of this evaluation methodology because it shows that taking biology into account can lead to improved performance.

6 Conclusion

In this article, our goal was to show how biology can be a source of inspiration, focusing on the three following questions: (i) How to design a model taking into account biology? (ii) What could be a suitable mathematical framework? (iii) How to evaluate an approach against the visual system performance?

We showed how to start from the state-of-the-art in neuroscience concerning the motion pathway, as the source of inspiration to define our model. So far, it is not possible to retain the full complexity of the cortical architecture, but our goal was to identify which key features should be incorporated into our model in order to reach a good compromise between bio-inspiration and computational efficiency.

Then, we chose a suitable mathematical formalism, namely the neural field formalism, in order to write the core equations of our model. Choosing this formalism has a biological interpretation: We focus on the macroscopic scale by defining the cortical activity at the population level. Since the number of neurons and synapses even in a small piece of cortical area is immense, we can assume that the relevant observable quantity is at the level of the population and not at the single cell level. From a computer vision point of view, the proposed neural field formalism, based on integral equations, has some interesting relationships with PDE-based approaches. Here we show that the neural field framework can successfully handle complex computer vision problems like motion estimation, and at the same time it offers a new well-posed framework, which can bring new ideas into the community.

Finally, considering that evaluating motion models only on real scenes is certainly not enough if one claims that a model is bio-inspired, we set the basis for

a new kind of benchmark based on human visual performance. Baseline results are provided on both bio-inspired and computer vision models. Interestingly, the recent computer vision approaches, which were not designed to match visual system performance, do perform very well on the stimuli presented here. In addition, the results show that our approach obtains the maximum score for the stimuli presented here. This is one important conclusion of this evaluation methodology as it shows that taking biology into account in models can lead to improved performance.

Perspectives around this work are promising and cover (i) improving the model, (ii) using the power of the neural field formalism more effectively, and (iii) further extending our benchmark which amounts to better characterise the visual system performance.

First, the proposed model can of course be extended in order to explain other biological phenomena. Among the possible extensions, we expect that the addition of delays (Faye and Faugeras, 2009; Deco and Roland, 2010) will be essential to constrain the propagation speed and account for more experimental data.

Second, the theoretical analysis of such equations is expected to have profound impact on our knowledge of the human visual system. For instance, delay equations allow one to distinguish between intra-cortical and extra-cortical interactions, due to their different speeds, and allows one to gain new insights into the mechanisms of the primate cortex. Another example is the application of the bifurcation theory to the proposed models in order to explain multi-stable percepts.

Finally, the proposed evaluation methodology can be further developed. For example, as mentioned above, it is well known in the literature that most of the motion stimuli are multi-stable. In the case of drifting plaids, one can perceive either two gratings with different velocities, or one single plaid motion (Hupé and Rubin, 2003). Incorporating this multi-stability in models is still only at the sketch level (Giese, 1998; Veltz and Faugeras, 2010; Tlapale et al, 2010b), and mostly ignored in motion benchmarks. Also, among the considered stimuli, various properties affecting the motion integration mechanisms were ignored. For example, contrast variations and disparity used in binocular experiments are missing, although their role in the dynamics and perception is significant.

Acknowledgements

This research work received funding from the Région PACA, the CNRS, the European Community (through FACETS, IST-FET, Sixth Framework, No 025213, and SEARISE, Seventh Framework, No 215866), the ERC grant No 227747 (NERVI) and the Agence Nationale de la Recherche (ANR, NATSTATS).

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
- Amari SI (1977) Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics* 27(2):77–87

- Angelucci A, Bullier J (2003) Reaching beyond the classical receptive field of v1 neurons: horizontal or feedback axons? *Journal of Physiology - Paris* 97(2-3):141-154
- Aubert G, Kornprobst P (2009) Can the nonlocal characterization of Sobolev spaces by Bourgain et al. be useful to solve variational problems? *SIAM Journal on Numerical Analysis* 47(2):844-860, DOI 10.1137/070696751
- 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(1):43-77
- Barthélemy F, Vanzetta I, Masson G (2006) Behavioral receptive field for ocular following in humans: Dynamics of spatial summation and center-surround interactions. *Journal of Neurophysiology* 95(6):3712-3726
- Bayerl P, Neumann H (2004) Disambiguating visual motion through contextual feedback modulation. *Neural Computation* 16(10):2041-2066
- Biber U, Ilg U (2008) Initiation of smooth-pursuit eye movements by real and illusory contours. *Vision Research* 48(8), DOI 10.1016/j.visres.2008.01.021
- Black M, Rangarajan P (1996) On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *The International Journal of Computer Vision* 19(1):57-91
- Born R (2001) Visual processing: parallel-er and parallel-er. *Current Biology* 11(14):R566-R568
- 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
- Bradley D, Goyal M (2008) Velocity computation in the primate visual system. *Nature Reviews Neuroscience* 9(9):686-695
- Bradski G (2000) The OpenCV Library. *Dr Dobb's Journal of Software Tools*
- Brox T, Malik J (2010) Large displacement optical flow: descriptor matching in variational motion estimation. *pami*
- Buades A, Coll B, Morel JM (2006) Neighborhood filters and PDE's. *Numerische Mathematik* 105(1):1-34
- Carandini M, Demb JB, Mante V, Tollhurst DJ, Dan Y, Olshausen BA, Gallant JL, Rust NC (2005) Do we know what the early visual system does? *Journal of Neuroscience* 25(46):10,577-10,597

- 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
- Castet E, Charton V, Dufour A (1999) The extrinsic/intrinsic classification of two-dimensional motion signals with barber-pole stimuli. *Vision Research* 39(5):915–932
- Chey J, Grossberg S, Mingolla E (1997) Neural dynamics of motion processing and speed discrimination. *Vision Res* 38:2769–2786
- Churchland P, Sejnowski T (1992) *The computational brain*. MIT Press
- Coombes S (2005) Waves, bumps, and patterns in neural fields theories. *Biological Cybernetics* 93(2):91–108
- Cottet GH, Ayyadi ME (1998) A Volterra type model for image processing. *IEEE Transactions on Image Processing* 7(3)
- Dayan P, Abbott L (2001) *Theoretical Neuroscience : Computational and Mathematical Modeling of Neural Systems*. MIT Press
- Deco G, Roland P (2010) The role of multi-area interactions for the computation of apparent motion. *NeuroImage* 51(3):1018–1026
- Degond P, Mas-Gallic S (1989) The weighted particle method for convection-diffusion equations. *Mathematics of Computation* 53(188):485–525
- Edwards R (1996) Approximation of neural network dynamics by reaction-diffusion equations. *Mathematical Methods in the Applied Sciences* 19:651–677
- Enkelmann W (1988) Investigation of multigrid algorithms for the estimation of optical flow fields in image sequences. *Computer Vision, Graphics, and Image Processing* 43:150–177
- Ermentrout B (1998) Neural networks as spatio-temporal pattern-forming systems. *Reports on Progress in Physics* 61:353–430
- Escobar MJ (2009) *Bio-inspired models for motion estimation and analysis: Human action recognition and motion integration*. PhD thesis, Université de Nice Sophia-Antipolis
- Escobar MJ, Masson GS, Vieville T, Kornprobst P (2009) Action recognition using a bio-inspired feedforward spiking network. *International Journal of Computer Vision* 82(3):284
- Faugeras O, Grimbert F, Slotine JJ (2008) Absolute stability and complete synchronization in a class of neural fields models. *SIAM Journal of Applied Mathematics* 61(1):205–250
- Faye G, Faugeras O (2009) Some theoretical and numerical results for delayed neural field equations. *Physica D* 239(9):561–578, special issue on Mathematical Neuroscience.

- Giese M (1998) *Dynamic Neural Field Theory for Motion Perception*. Springer
- Grinvald A, Lieke EE, Frostig RD, Hildesheim R (1994) Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. *Journal of Neuroscience* 14(5):2545–2568
- Grossberg S, Mingolla E (1985) Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychological review* 92(2):173–211
- Grzywacz N, Yuille A (1991) *Theories for the visual perception of local velocity and coherent motion*, The MIT Press, chap 16, pp 231–252. Bradford Books
- Hildreth E (1983) *The measurement of visual motion*. PhD thesis, MIT
- Horn B, Schunck B (1981) Determining Optical Flow. *Artificial Intelligence* 17:185–203
- Hupé J, Rubin N (2003) The dynamics of bi-stable alternation in ambiguous motion displays: a fresh look at plaids. *Vision Research* 43(5):531–548
- Hupé J, James A, Payne B, Lomber S, Girard P, Bullier J (1998) Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature* 394:784–791
- Kooi T (1993) Local direction of edge motion causes and abolishes the barber-pole illusion. *Vision Research* 33(16):2347–2351
- Löffler G, Orbach H (1998) Computing feature motion without feature detectors: A model for terminator motion without end-stopped cells. *Vision Research* 39(4):859–871
- Lorenceau 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
- Lucas B, Kanade T (1981) An iterative image registration technique with an application to stereo vision. In: *International Joint Conference on Artificial Intelligence*, pp 674–679
- Masson G, Stone L (2002) From following edges to pursuing objects. *Journal of neurophysiology* 88(5):2869–2873
- Masson G, Rybarczyk Y, Castet E, Mestre D (2000) Temporal dynamics of motion integration for the initiation of tracking eye movements at ultra-short latencies. *Visual Neuroscience* 17(05):753–767
- Montagnini A, Mamassian P, Perrinet L, Castet E, Masson G (2007) Bayesian modeling of dynamic motion integration. *Journal of Physiology – Paris* 101(1-3):64–77
- Movshon J, Adelson E, Gizzi M, Newsome W (1985) The analysis of visual moving patterns. *Pattern recognition mechanisms* pp 117–151

- Nakayama K, Silverman G (1988) The aperture problem. II. spatial integration of velocity information along contours. *Vision Research* 28(6):747–753
- Nassi J, Callaway E (2009) Parallel processing strategies of the primate visual system. *Nature Reviews Neuroscience* 10(5):360–372
- Nassi JJ, Callaway EE (2006) Multiple circuits relaying primate parallel visual pathways to the middle temporal area. *Journal of Neuroscience* 26(49):12,789–12,798
- Nir T, Bruckstein AM, Kimmel R (2008) Over-parameterized variational optical flow. *International Journal of Computer Vision* 76(2):205–216
- Nowlan S, Sejnowski T (1994) Filter selection model for motion segmentation and velocity integration. *J Opt Soc Am A* 11(12):3177–3199
- 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, Gartland A, Born R (2004) Integration of contour and terminator signals in visual area MT of alert macaque. *The Journal of Neuroscience* 24(13):3268–3280
- Pinto D, Brumberg J, Simons D, Ermentrout G, Traub R (1996) A quantitative population model of whisker barrels: re-examining the wilson-cowan equations. *Journal of Computational Neuroscience* 3(3):247–264
- 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(11):2941–2950
- Reichardt W (1957) Autokorrelationsauswertung als Funktionsprinzip des Zentralnervensystems. *Zeitschrift für Naturforschung* 12:447–457
- Rockland K, Knutson T (2000) Feedback connections from area MT of the squirrel monkey to areas V1 and V2. *The Journal of Comparative Neurology* 425(3):345–368
- Rust N, Mante V, Simoncelli E, Movshon J (2006) How MT cells analyze the motion of visual patterns. *Nature Neuroscience* 9:1421–1431
- Seriès P, Georges S, Lorenceau J, Frégnac Y (2001) A network view of the structure of center/surround modulations of v1 receptive field properties in visual and cortical spaces. *Neurocomputing* 38:881–888
- Shiffrar M, Lorenceau J (1996) Increased motion linking across edges with decreased luminance contrast, edge width and duration. *Vision Research* 36(14):2061–2067
- Sillito A, Cudeiro J, Jones H (2006) Always returning: feedback and sensory processing in visual cortex and thalamus. *TRENDS in Neurosciences* 29(6):307–316
- Simoncelli E, Heeger D (1998) A model of neuronal responses in visual area MT. *Vision Research* 38:743–761

- Smith M, Majaj N, Movshon A (2005) Dynamics of motion signaling by neurons in macaque area MT. *Nature Neuroscience* 8(2):220–228
- Sun D, Roth S, Darmstadt T, Black M (2010) Secrets of optical flow estimation and their principles. *cvpr*
- Tlapale É, Kornprobst P, Bouecke J, Neumann H, Masson G (2010a) Towards a bio-inspired evaluation methodology for motion estimation models. Research Report RR-7317, INRIA
- Tlapale É, Masson G, Kornprobst P (2010b) Modelling the dynamics of motion integration with a new luminance-gated diffusion mechanism. *Vision Research* 50(17):1676–1692, DOI 10.1016/j.visres.2010.05.022
- Van Santen J, Sperling G (1985) Elaborated reichardt detectors. *Journal of the Optical Society of America A* 2(2):300–320
- Veltz R, Faugeras O (2010) Local/global analysis of the stationary solutions of some neural field equations. *SIAM Journal on Applied Dynamical Systems* 9(3):954–998, DOI 10.1137/090773611
- Viéville T, Chemla S, Kornprobst P (2007) How do high-level specifications of the brain relate to variational approaches? *Journal of Physiology - Paris* 101(1-3):118–135
- Wallace J, Stone L, Masson G (2005) Object motion computation for the initiation of smooth pursuit eye movements in humans. *Journal of Neurophysiology* 93(4):2279–2293
- Wallach H (1935) Über visuell wahrgenommene Bewegungsrichtung. *Psychological Research* 20(1):325–380
- Watson A, Ahumada A (1985) Model of human visual-motion sensing. *J Opt Soc Am A* 2(2):322–342
- Weickert J, Schnörr C (2001) Variational optic flow computation with a spatio-temporal smoothness constraint. *Journal of Mathematical Imaging and Vision* 14(3):245–255
- Weiss Y, Adelson E (2000) Adventures with gelatinous ellipses – constraints on models of human motion analysis. *Perception* 29:543–566
- Weiss Y, Adelson EH (1998) Slow and smooth: A Bayesian theory for the combination of local motion signals in human vision. Center for Biological and Computational Learning Paper
- Wilson H, Cowan J (1972) Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys J* 12:1–24
- Wilson H, Cowan J (1973) A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Biological Cybernetics* 13(2):55–80
- Wilson H, Ferrera V, Yo C (1992) A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience* 9(1):79–97

- Yo C, Wilson H (1992) Perceived direction of moving two-dimensional patterns depends on duration, contrast and eccentricity. *Vision Research* 32(1):135–47
- Yuille A, Grzywacz N (1989) A winner-take-all mechanism based on presynaptic inhibition feedback. *Neural Computation* 1(3):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