Modelling the dynamics of contextual motion integration in the primate

Émilien Tlapale

▶ **To cite this version:**

Émilien Tlapale. Modelling the dynamics of contextual motion integration in the primate. Modeling and Simulation. Université Nice Sophia Antipolis, 2011. English. NNT : . tel-00850265

HAL Id: tel-00850265
https://theses.hal.science/tel-00850265

Submitted on 12 Aug 2013

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.

University of Nice Sophia Antipolis — Faculty of Science

Doctoral School STIC (Information and Communication Sciences)

THESIS

submitted in partial fulfillment of the requirements for the degree of

Doctor of Science

of the University of Nice Sophia Antipolis

Field of study

Control, signal and image processing

Modelling the dynamics of contextual motion integration in the primate

Émilien Arturo Laurent Tlapale Olivier

Defence expected on 25 January 2011

Supervisors	Pierre Kornprobst	Inria Sophia Antipolis
	Guillaume S. Masson	Cnrs Marseille
Examiners	Frédéric Alexandre	Inria Nancy
	James A. Bednar	University of Edinburgh
	Ennio Mingolla	Boston University
Jury members	Olivier Faugeras	Inria Sophia Antipolis
	Jean Lorenceau	Cnrs Paris
	Gabriel Peyré	Université Paris-Dauphine

Université Nice Sophia Antipolis UFR Sciences

École Doctorale STIC (Sciences et Technologies de l'Information et de la Communication)

THÈSE

pour obtenir le titre de

Docteur en Sciences

de l'Université Nice Sophia Antipolis

Mention

Automatique, traitement du signal et des images

Modélisation des dynamiques d'intégration contextuelle du mouvement chez le primate

Émilien Arturo Laurent Tlapale Olivier

Soutenance prévue le 25 janvier 2011

Directeurs	Pierre Kornprobst	Inria Sophia Antipolis
	Guillaume S. Masson	Cnrs Marseille
Rapporteurs	Frédéric Alexandre	Inria Nancy
	James A. Bednar	University of Edinburgh
	Ennio Mingolla	Boston University
Examinateurs	Olivier Faugeras	Inria Sophia Antipolis
	Jean Lorenceau	Cnrs Paris
	Gabriel Peyré	Université Paris-Dauphine

Émilien A.L. Tlapale O., *Modelling the dynamics of contextual motion integration in the primate*, PhD Thesis,
© January 2011. Prepared at the inria Sophia Antipolis Méditerranée.

website
http://emilien.tlapale.com

e-mail
emilien@tlapale.com

Content

Abstract

This thesis addresses the study of motion integration in the primate. Based on anatomical and functional knowledge of two cortical areas involved in motion perception, namely V1 and MT, we explain various perceptual and oculo-motor responses found in the literature. First, we build a recurrent model of motion integration where a minimal number of cortical interactions are assumed. Proposing a simple readout mechanism, we are able to reproduce not only motion perception but also the dynamics of smooth pursuit eye movements on various line figures and gratings viewed through different apertures. Second, following perceptual studies concerning motion integration and physiological studies of receptive fields, we construct another dynamical model where motion information is gated by form cues. To this end, we postulate that the visual cortex takes advantage of luminance smoothness in order to gate motion diffusion. Such an elementary diffusion mechanism allows to solve various contextual problems where extrinsic junctions should be eliminated, without relying on complex junction detectors or depth computation. Finally, we rewrite the initial dynamical model into the neural fields formalism in order to mathematically analyse its properties. We incorporate the multiplicative feedback term into the formalism, and prove the existence and uniqueness of the solution. To generalise the comparison against visual performance, we propose a new evaluation methodology based on human visual performance and design a database of image sequences taken from biology and psychophysics literature. Indeed, offering proper evaluation methodology is essential to continue progress in modelling the neural mechanisms involved in motion processing. To conclude, we investigate the performances of our neural fields model by comparison against state of the art computer vision approaches and sequences. We find that, despite its original objective, this model gives results comparable to recent computer vision approaches of motion estimation.

KEYWORDS bio-inspired models · dynamics · luminance gating · modelling · motion integration · neural fields · perception

Résumé

Dans cette thèse, nous étudions l'intégration du mouvement chez le primate. En se basant sur les connaissances actuelles concernant l'anatomie et les fonctions de deux aires corticales impliquées dans le mouvement, VI et MT, nous expliquons un certain nombre de réponses perceptuelles et oculo-motrices rapportées dans la littérature. Tout d'abord, nous construisons un modèle récurrent d'intégration du mouvement se basant sur un nombre minimal d'hypothèses concernant les interactions corticales. En proposant un simple mécanisme de « lecture », nous sommes capable de reproduire non seulement la perception, mais aussi les dynamiques oculaires de poursuite sur des stimuli de type ligne ou grille. De là, en se basant des études psychophysiques sur l'intégration du mouvement et sur des études physiologique concernant les champs récepteurs, nous construisons un deuxième modèle dynamique dans lequel l'information concernant le mouvement est dirigée par un signal de forme. Pour cela, nous postulons que le cortex visuel utilise la régularité de la luminance pour diriger la diffusion du mouvement. Un tel mécanisme élémentaire de diffusion permet de résoudre des problèmes contextuels, dans lesquels les jonctions extrinsèques doivent être ignorées, sans avoir besoin d'utiliser des mécanismes plus complexes tels que les détecteurs de jonctions ou le calcul de profondeur. Enfin, nous reformulons le modèle initial dans le cadre du formalisme des champs neuronaux afin d'analyser mathématiquement ses propriétés. Nous incorporons la rétroaction multiplicative dans le formalisme et prouvons l'existence et l'unicité de la solution. Afin de généraliser les comparaisons aux performances du système visuel, nous proposons une nouvelle méthodologie d'évaluation basée sur les performances du système visuel humain, accompagnée d'une série de vidéos issues de la littérature biologique et psychophysique. En effet, une méthodologie d'évaluation adaptée nous semble essentielle afin de continuer les progrès en modélisation des mécanismes neuraux impliqués dans le traitement du mouvement. Pour conclure, nous analysons les performances de notre modèle d'intégration du mouvement en l'appliquant à des problèmes classiques et récents issus de la vision par ordinateur. En dépit de son objectif initial, notre modèle est capable de donner des résultats comparables aux récentes approches proposées en vision par ordinateur au niveau de l'estimation du mouvement.

MOTS-CLÉS champs neuronaux · dynamiques · luminance · modèles bio-inspirés · modélisation · intégration du mouvement · perception

Introduction

MOTION AND THE BRAIN

Perception

Perceiving motion is being able to see that an object moves and to know in which direction and at which speed.

The famous case of M.P. described by Zihl, von Cramon and Mai[266] allowed a better understanding of motion perception. Indeed a cardiovascular accident damaged the cortical areas involved with motion perception in this patient. From that point in time M.P. has been unable to achieve simple acts such as filling a cup of tea or crossing a street. While still perceiving the level of the liquid or the position of cars, M.P. cannot estimate the overflowing time or distinguish between stopped and moving cars.

The state of M.P., known as *akinetopsia*, did not decrease her abilities to recognise shapes or faces, and her colour perception remained intact. Such a disability suggests that motion information flows through a specific pathway in the brain.* This division between different visual pathways is confirmed by the existence of other similar conditions such as *cerebral achromatopsia*, colour blindness originating from brain malfunction.[263]

* At least a part of motion information. Recognition of biological motion or structure from motion seem not affected by those damages.[235]

Yet divisions between visual pathways, as well as between visual and non-visual pathways, are porous. For instance colour and luminance information[191,177] as well as auditory motion[213] can improve motion perception. An extreme case of this porosity is *synesthesia* in which two more senses are coupled.[97]

Understanding the mechanisms involved in motion perception is the first step towards a hope for cure in patients with damages to motion areas.

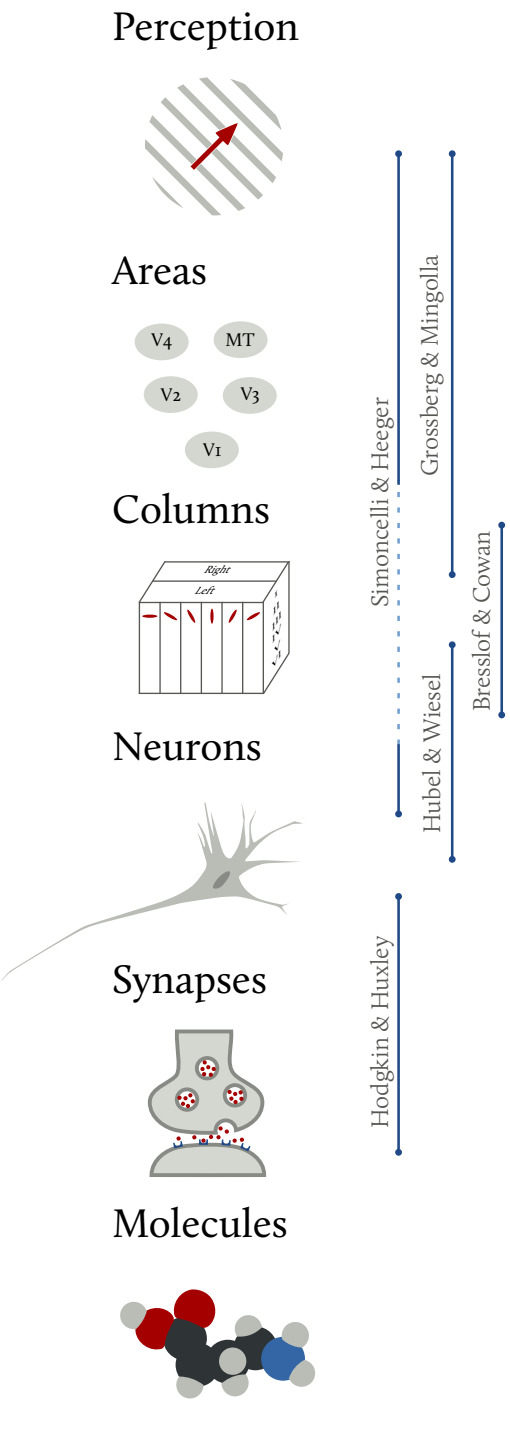


Fig. 1 Some possible modelling levels in visual neuroscience.

Eye movements

Partly due to the peculiar nature of our retina, in which the central part—the *fovea*—is the only precise and colour-aware region, our eyes are constantly in motion. Various type of eye movements exist, such as *saccades* to centre the projection of an object on the retina, the *vestibulo-ocular reflex* to compensate head movements, or *smooth pursuit* to follow moving objects.

Smooth pursuit movements are tightly linked to motion perception. As it is the case in perception, following an object requires to know in which direction and at which speed it is moving. It has been shown that both pursuit and perception share common mechanism of motion integration. Since eye movements are fundamentally dynamical, they give us essential information on the involved mechanisms.

Physiology

In the case of patient M.P., damages occurred in the cortical area MT—also named V5. This cortical area happen to be specialised in motion integration,[69,5,264] and is involved both in motion perception[44] and smooth pursuit.[140,118] The area MT is only one of the numerous cortical areas of the visual system. Each cortical area is tuned for specific features—colour, motion, disparity, etc.—and is composed of large number of interconnected neurons. Various techniques were developed to explore the function and anatomy of the visual cortex, and those techniques allow us to understand how the processing of visual information works.

Modelling

Dayan and Abbott[64] classify models in three categories depending on which of the *what*, *how* or *why* questions they answer. *Descriptive models* (or phenomenological models) aim at accurately describing large amount of experimental data. *Mechanistic models* address the question of how systems operates in a bottom-up approach. *Interpretive models* (or optimal models) try to understand the behaviour of a system with a top-down approach focusing on the functional role of a phenomenon.

Mechanistic models link experimental recordings across different levels to form a single system (see Fig. 1). For instance the model of Hodgkin and Huxley,[109] describes the ionic interactions inside the giant axon of the squid to explain the emission and propagation of action potentials, also called *spikes*.

In this thesis we propose mechanistic models linking the neural activity inside the cortical areas processing motion, and perception. Several models were proposed to explain motion perception based on the anatomy of the cortical areas V1 and MT.[208,24,32] With various granularity of details, they manage to explain some important experiments recorded in psychophysics.

THESIS ORGANISATION

The manuscript is divided in three parts. In Part I we review the experimental data concerning motion at perceptual, behavioural and physiological levels. In Part II we propose different versions of a recurrent dynamical architecture reproducing the dynamics of motion integration. In Part III we describe an evaluation methodology for motion estimation based on human visual performance, and we evaluate our neural fields model against computer vision performance.

SEEING MOTION — Part I

Perception — *Chapter 1*

We begin by presenting the fundamental problems of motion integration in the case of motion perception. In particular we put the *aperture problem* as the centre of our study in various line-segments and gratings configurations.

Tracking — *Chapter 2*

Smooth pursuit and motion perception are two intertwined processes of the visual system. We discuss the similarities between pursuit and perception in the case of the aperture problem and show how the dynamics of eye movements can reveal the mechanisms underlying motion integration.

Physiology Chapter 3

To understand how motion perception and smooth pursuit are controlled, we detail the physiology of the visual cortex focusing on two areas, V1 and MT. Indeed, V1 is the major entry point of the visual cortex, and MT is a cortical area shown to be involved with both perception and pursuit.

Models Chapter 4

Several models of motion perception and smooth pursuit are described in the literature. We selected three models of motion perception describing the connectivity between V1 and MT. Those models served as inspiration for our work.

MODELLING MOTION Part II

Model architecture Chapter 5

Based on current knowledge on visual system interactions, we define a two-layer model architecture for motion integration. Each layer corresponds to the activity of neuronal populations in cortical areas V1 and MT. Lateral, feed-forward and feedback connections are proposed to mimic cortical interactions. We also propose a simple readout mechanism able to control direction and speed of smooth pursuit.

Dynamics of integration Chapter 6

We propose a model able to explain not only motion perception on a large range of stimuli, but also to reproduce the ocular dynamics of smooth pursuit. The model postulates the existence of a small number of neural mechanisms and we justify them with anatomical data and neural recordings of the two cortical areas V1 and MT.

Luminance gating Chapter 7

Then we introduce a novel mechanism of motion integration gated by luminance information coming from another cortical area, V2. This contextual information allows the visual system to better analyse the visual scene by considering that different

objects have different luminance. It also explains how contextual information can influence motion integration.

Neural fields *Chapter 8*

Finally we study the mathematical and computational properties of the proposed mechanism. To this end we fit our model into the *neural fields* formalism.[254,7] This rewriting allows us to prove the existence and uniqueness of the solution given by our model. We also study the computational implications of the selection mechanism and of lateral diffusion.

EVALUATING MODELS Part III

Comparison to the visual system *Chapter 9*

Generalising our comparisons, we design a benchmark for models of motion perception and motion integration. We consider several stimuli and suggest evaluation criteria to compare the output of motion integration models with behavioural and perceptual results.

We divided our benchmark into two kinds of evaluations.

The *static evaluation* only considers the final percept induced by a stimulus. This evaluation mostly use psychophysical results and can be applied not only to the output of motion models but also to computer vision approaches. Indeed the wide range of spatial scales or frequencies used in psychophysics is able to challenge several single-scale computer vision approaches.

The *dynamic evaluation* considers the time course of motion integration, and is thus linked to smooth pursuit or perception dynamics.

Comparison to computer vision *Chapter 10*

In another field of study, understanding motion perception can improve the design of efficient robotic applications. Indeed the human visual system outperforms computer vision methods in a wide range of applications, despite the rapidly increasing hardware capabilities. We show that even our model of motion

perception give results in the range of current computer vision algorithms.

APPENDICES — Part IV

European project SEARISE — Chapter A

In the first appendix we describe our work in the context of the European project SEARISE. The goal of SEARISE was to developed a trinocular active cognitive visual system, *Smart-Eyes*, for detection, tracking and categorisation of salient events and behaviours. In this context our main contribution was to design motion integration algorithms and implement them using GP-GPU technologies.

Reading with low-vision — Chapter B

In this appendix we propose a reading aid software for low vision patients. This work started as a postgraduate fellowship in a collaboration with Éric Castet and Jean-Baptiste Bernard from the CNRS. It was extended during the PhD thesis, although not being its main objective.

Million of people suffer from low-vision, a disability tightly linked to age. In cases such as age-related macular degeneration (AMD), the visual acuity decreases, and patients read more easily with uncluttered large characters. In cases such as glaucoma, the periphery of the visual field is impaired, decreasing the ability to navigate in complex documents.

In this appendix we propose a system facilitating reading for low-vision patients by analysing complex documents in order to ease navigation and allow custom text display. To validate our approach, 26 subjects compared our software to existing reading aid under the supervision of orthoptists. Promising results validate our approach which allows patients to continue reading, in comparison with more autonomous systems such as voice synthesisers.

Part I

Motion integration in the primate

Part 1 Motion integration in the primate

Sharing common neural substrate, motion perception and eye movements are two intertwined processes of the primate visual system. We start by presenting the fundamental problems of motion integration and their relation to human perception in Chapter 1. We investigate the solution of the aperture problem and the discrimination between intrinsic and extrinsic junctions and conclude the chapter with temporal dynamics in perception.

This transition to dynamics allows comparison between perception and one class of eye movements, smooth pursuit, and we investigate the links between the two in Chapter 2. As the objective of smooth pursuit is to track moving objects, a direction need to be computed and the initial direction error reveals the inner mechanisms underlying the tracking.

Deeper into the brain, anatomical data and physiological recordings divulges the machinery behind motion integration, providing the chemical and cellular substrate of the various computations hinted by perception and behaviour. In Chapter 3 we discuss the visual information stream in the primate visual system and its relation with motion perception and smooth pursuit.

We conclude our motion integration panorama in Chapter 4 with the description of several models of motion computation. The architecture of the described models follows the observations made on the visual cortex and focus on motion perception. The models we propose in the second part of this thesis are inspired by those work.

Chapter 1 Perceiving motion

In this chapter we introduce motion integration from a perceptual perspective. We start by describing the fundamental aperture problem *in Section 1.1. Then we review some classical mechanisms and terminology in Section 1.2. In Section 1.3 we discuss contextual influences and cross-modality. We end the chapter in Section 1.4 by showing that perception is not stable but changes in time.*

1.1 THE APERTURE PROBLEM

Natural scenes present many sources of ambiguities, that are to be solved in order to extract reliable information to control behaviour. One example of such ambiguity is the *aperture problem* in motion perception, described by Wallach.[243]

When a translating bar is viewed behind an occluder masking its extremities, its translation direction cannot be recovered. Indeed an infinite number of constrained translations are possible. In Fig. 1.1 we show such a translating bar in the first row, at three different times. The second and last rows show two different translations coherent with the stimulus.

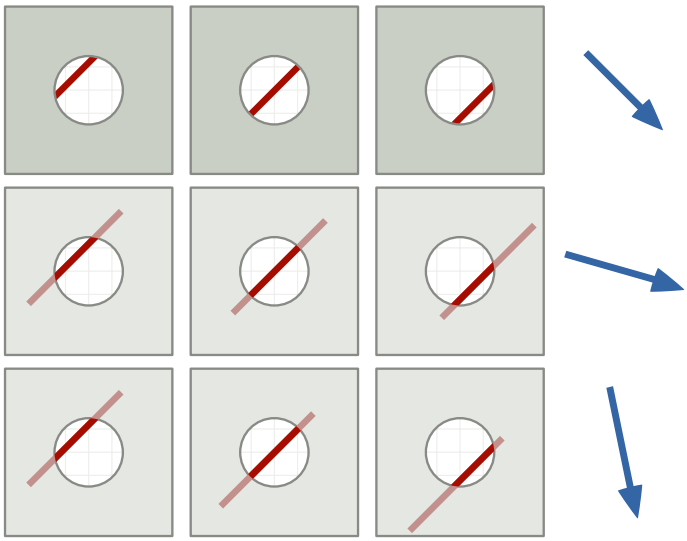


Fig. 1.1 Aperture problem in motion detection. When an edge is seen through an aperture masking its extremities, its motion direction is perceived as orthogonal to its orientation, a choice between the infinity of possible motion directions.

Interestingly when only the central part of the bar is visible,

the perceived motion is the translation orthogonal to the bar orientation.

The aperture problem is fundamental in the visual cortex. Indeed it appears at several levels of the visual system, including at the oculo-motor (see Chapter 2) and at the neuronal levels (see Chapter 3). More generally the aperture problem can be found in non-motion visual processing such as stereopsis.[160]

1.2 MOTION INTEGRATION

In order to compute the global motion of an object embedded in a complex surrounding, the visual system takes local motion estimates as input. As a consequence, it must deal with numerous *1D features*, corresponding to edges and, generally fewer, *2D features* such as corners or line-endings. Indeed the 1D features—also called 1D cues—only allow to recover the motion orthogonally to the bar, whereas 2D features—or 2D cues—allow to recover the complete 2D motion.

After several decades of intensive research on 2D motion perception and its neural substrates,[152] 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.[41] We review some classical explanations of motion integration in the next sections.

1.2.1 *1D motion integration*

In order to describe visually the computation rules suggested in the literature, the set of possible velocities can be shown in their velocity space. In Fig. 1.2 we show some of the possible velocities occurring at the centre of a bar and represent them in the associated velocity space. To explain that perception and ocular responses output the motion orthogonal to the bar orientation, one can postulate the existence of a prior on small velocities.[248]

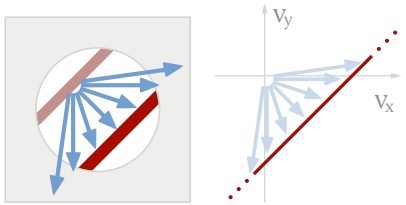


Fig. 1.2 Velocity space representation. *Left*: some of the possible velocities in the aperture problem. *Right*: All the possible velocities are aligned when represented in a velocity space.

To unravel the computational rules involved in motion perception more complex stimuli were proposed. By increasing the number of translating bars a *grating* is generated. The superposition of two gratings with different orientations is called a

plaid (see Fig. 1.3). Each of the gratings is represented by a line in the velocity space and perceived as moving in a direction orthogonal to its orientation. When both gratings are presented simultaneously as a plaid pattern, the perceived motion direction corresponds to the intersection of the lines in the velocity space—purely horizontal in the example. This geometrical solution is named the *intersection of constraints* (IOC)[78] and several studies have proposed that the primate visual system uses a similar computation.[2]

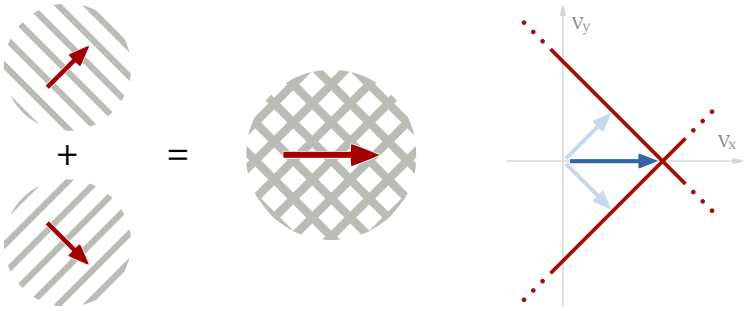


Fig. 1.3 Plaid pattern. *Left*: The superposition of two translating gratings with different orientations is called a plaid. *Right*: Superposition of the associated velocity spaces.

It remains however unclear how the visual system can implement the IOC rule. Moreover, the fact that perceived direction does not always correspond to the IOC solution, at least for short stimulus durations,[260] has supported alternative models. Among them, the *vector average* (VA) is defined by averaging the gratings motions. In the plaid presented in Fig. 1.3 the vector average has the same direction as the IOC, but a slower velocity. By changing the properties of the gratings separately, one can create *type II* plaids where the motion direction of the IOC and the VA are different.

1.2.2 *2D motion integration*

As pointed out by Wallach[243] a spatial integration of 1D features can be used to reconstruct the translation of a moving object. But 2D features can also be extracted as their motion seen through the same aperture size is not ambiguous. In the stimuli presented in the previous section, we only considered and plotted the 1D motion, and ignored 2D features coming from the edges of the aperture or the gratings intersections.

To investigate the effect of the number of 2D cues on mo-

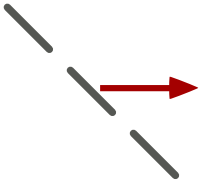


Fig. 1.4 Cut translating bar.

tion one can cut a translating bar into smaller ones increases the number of 2D cues (see Fig. 1.4). Lorenceau and colleagues[144] investigated the effect of increased 2D cues by displaying translating bars of different length for a short period of time and asking the subject to choose a motion direction. Their results show that increasing the numbers of 2D cues decreases the perceived direction error.

For certain stimuli such as the barber pole illusion where a grating is viewed through a rectangular aperture, considering 2D features is fundamental. In Fig. 1.5 we show two identical gratings behind rectangular apertures with different orientations. With such configurations, the perceived motion direction is biased towards the elongated border of the aperture. Yet a pure 1D motion integration analysis would consider both configurations as equivalent since the grating motion direction is the same in both cases.

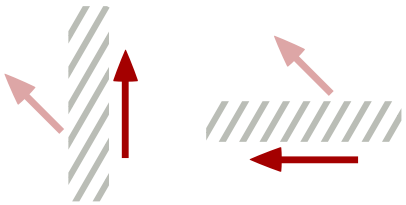


Fig. 1.5 Barber pole illusion. In both stimuli an identical grating translating in a diagonal direction (*light arrow*) is viewed through a rectangular aperture. Yet the perceived direction (*dark arrow*) depends on the aperture orientation.

From the configurations of the barber pole illusions two sets of 2D cues arise with directions collinear to the short and long edges of the aperture. Wallach[242] proposed that the greater number of terminators on the longer borders explains the perceived motion. This corresponds to a higher peak in the velocity space representation of Bayesian models.[248] Other models propose the unambiguous information to propagate on the ambiguous line.[108]

1.3 CONTEXTUAL INTEGRATION

1.3.1 *Intrinsic versus extrinsic junctions*

Moving objects constantly mask and unmask other objects, creating occlusions. For instance when two squares are translating, 1D motion cues can be extracted from their edges, and 2D motion cues can be extracted from their angles. If the two squares are overlapping, new 2D motion cues appears as shown in Fig. 1.6.

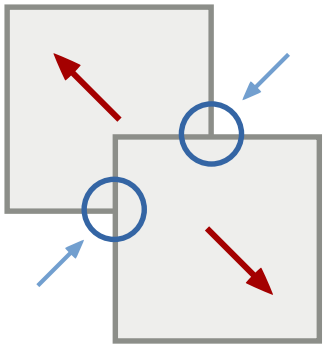


Fig. 1.6 When two squares are crossing extrinsic junctions appear at the occlusions (*circles*) leading to erroneous 2D signals (*small arrows*)

.

In such cases the visual system makes certain assumptions to eliminate 2D features created by the occlusions, and to keep the veridical 2D features.[206] The 2D features to be ignored are called *extrinsic junctions*, whereas the true 2D features are called *intrinsic junctions*.

In the chopstick illusion presented in Fig. 1.7 two translating bars are presented. Despite one configuration having occluders masking the extremities of the bars, both stimuli have the same set of 1D and 2D features. However the perceived motion is each case is very different. As such the chopstick illusion provides an example where pure motion processing is not sufficient to explain the percept. Similarly, adding occluders to the borders of barber pole stimuli changes the motion percept.

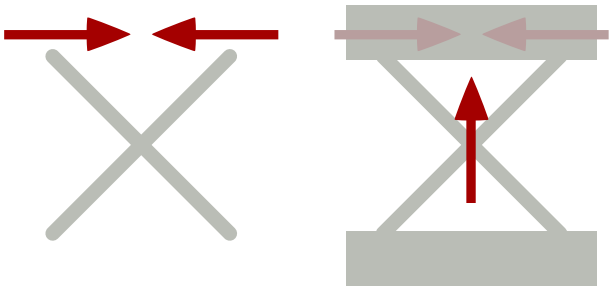


Fig. 1.7 In both stimuli two translating bars are crossing leading to the same set of 1D and 2D features. Yet the two translating bars are perceived as a single upward moving object when their extremities are occluded.

To account for perception several proposals model an explicit junction segregation through features detectors. In their seminal work Shimojo and colleagues[206] suggested T-junctions to act as monocular depth cues and should thus be classified as extrinsic. Indeed looking at the content of the circles—the extrinsic junctions—in Fig. 1.6 one can see rotated T created by the squares borders. Inhibition of T-junctions was confirmed by Lorenceau and Zago,[145] at least for low spatial frequencies and small stimuli sizes. Several models implement this idea: Weiss and Adelson mark three-labelled regions as extrinsic;[249] Grossberg and Mingolla[139,96] implement a T-junction detector in their FACADE model; Bayerl and Neumann[23] perform a similar feature detection but use it in an excitatory way.

1.3.2 *Multi-aperture stimuli*

The diamond stimuli proposed by Lorenceau and colleagues[143] provide an interesting set of experiments to understand the mechanisms of contextual motion perception. Their basic stimulus is a diamond viewed behind four apertures that only show the edges. Depending on the orientation of the apertures, the IOC and the VA solutions are not always compatible with the true diamond motion (see Fig. 1.8)

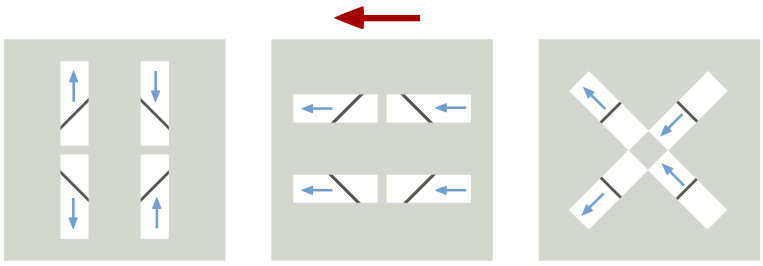


Fig. 1.8 If a translating diamond is viewed behind four rectangular apertures showing only its edges, IOC and VA solutions are not always compatible with the true translation (*big arrow*).[143]

In the experiments of Lorenceau and Shiffrar[143], a diamond is translating along a circular path. Perception of a moving rigid object is stronger when the aperture is clearly delimited (as in Fig. 1.8), for configurations for which the IOC solution is valid, when the stimulus is presented in the periphery of the visual field, or when the terminators noise is increased.

A later set of stimuli by Lorenceau and Alais[142] provided clues on how geometry influences motion integration. By shifting the edges of the diamond, the authors create an extensive set of geometric objects to study perception (see Fig. 1.9). Results show that form has a critical role in motion integration. In particular global motion is perceived more easily if the object is a diamond than if the object is an arrow or a cross.

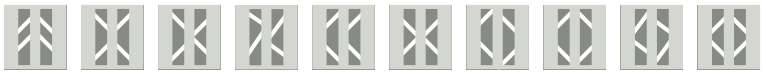


Fig. 1.9 Diamonds and arrows from Lorenceau and Alais.[142]. All geometric figures are derived from a diamond but have the edges shifted differently.

1.3.3 *Cross-modal influences*

The hypothesis underlying the classification of T-junctions as extrinsic comes from the statistical observation of occluded objects. It relies—at least in models—on a non-motion information originating from the junction detectors in the *form pathway*. Yet several other multi-features influences are reported in the literature.

For instance varying the luminance at the intersection of superposed gratings, Ramachandran and colleagues[218,185] found that luminance compatible with transparency elicit the percept of two gratings, and not one single plaid (see Fig. 1.10). Thus

perception can switch between one single plaid—*pattern motion*—or two gratings moving in different directions—*component motion*.[218,117]

Several other examples of influence between features exist. Obviously, disparity information from binocular stimuli greatly improves motion perception. For monocular stimuli, luminance or colour information[191,177] as well as auditory motion[213] were shown to influence motion perception.

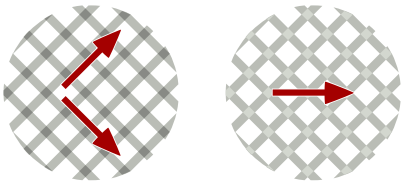


Fig. 1.10 Transparent motion perception. *Left*: When the luminance at the intersection of the gratings is compatible with transparency, two motion are perceived. *Right*: When the luminance is incompatible with transparency, a single plaid motion is perceived.

1.4 Dynamics

1.4.1 *Multi-stability*

Most, if not all, the stimuli used in psychophysics are multi-stable, as they can be perceived differently from trial to trial, or in time within a single experiment. For most of the stimuli the alternative perception are marginal and only the predominant percept is studied in the literature. However some stimuli have been designed to study those changes in percept.

A crossed barber pole can be obtained by combining the two barber poles of Fig. 1.5, i.e. by viewing a drifting grating through a cross-shaped aperture (see Fig. 1.11). In this case, the percept is strongly multi-stable and observers perceive either the grating translation or two orthogonal motion.[52]

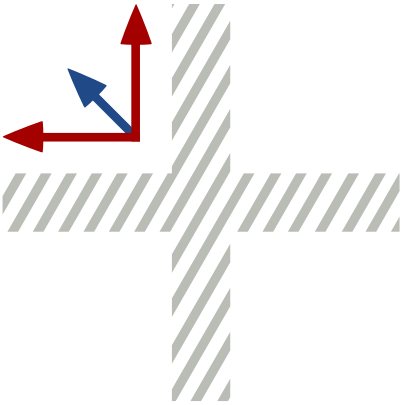


Fig. 1.11 Crossed barber pole. Observer perception shifts between grating motion (*blue*) and barber poles (*red*).

1.4.2 *Presentation time*

Perception changes over time in multi-stable stimuli, but perception also changes depending on the presentation time of a given stimulus. For instance, when a simple translating object such as a slanted bar is presented to an observer, his visual system is fooled by local ambiguous cues for several hundreds of milliseconds.

For short durations the subject perceives a motion direction orthogonal to the bar orientation instead of the true motion direction[51,144,205] (see Fig. 1.12). Such a perceptual bias is corrected for longer durations.

Therefore the visual perception system seemed to be fooled by the aperture problem for short durations, as if 2D features from bar extremities are delayed in comparison with 1D fea-

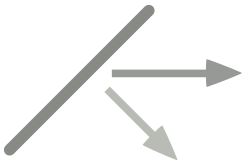


Fig. 1.12 Translating bar stimulus. A slanted bar is translating horizontally (*darker arrow*), but its perceived direction is initially biased towards the direction orthogonal to its orientation (*lighter arrow*).

tures from the middle region. Varying the presentation time of the stimulus, one can plot the temporal evolution of the direction error. In Fig. 1.13 we show such a time course for a translating bar stimulus as recorded by Lorenceau and colleagues.[144]

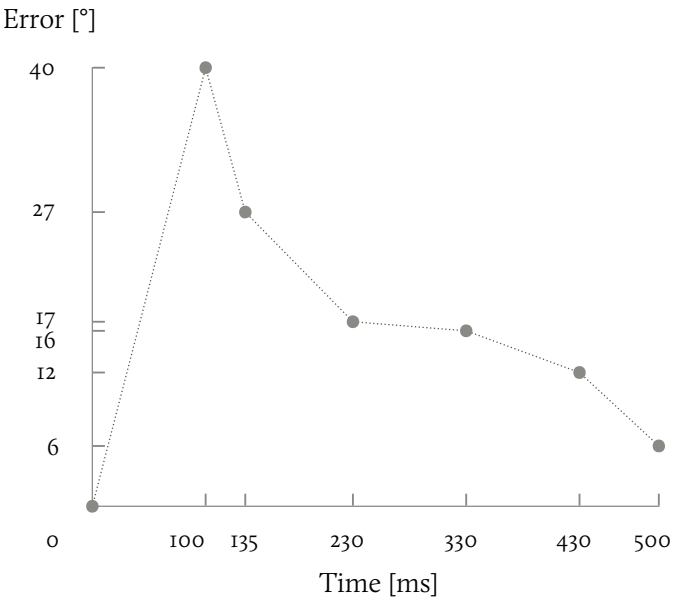


Fig. 1.13 Perception dynamics of the translating bar. The angular error of perceive direction *versus* true bar direction is plotted as a function of stimulus duration (from Lorenceau and colleagues[144]).

In the next chapter, we introduce a more natural approach to explore motion integration dynamics, recording *smooth pursuit* eye movements.

Chapter 2 Tracking objects

In this chapter we introduce smooth pursuit eye movements as a tool to understand the mechanisms underlying motion integration. We start in Section 2.1 by describing what is smooth pursuit and how it is recorded. In Section 2.2 we show that pursuit is also influenced by the aperture problem for short duration. In Section 2.3 we discuss the spatial integration leading to smooth pursuit.

2.1 SMOOTH PURSUIT

2.1.1 *Eye movements*

In the previous chapter we described 2D motion integration from a psychophysics point of view. As such we considered perception of the subjects at the end of an experiment and building up a time analysis required stacking experiments of different durations (see Section 1.4.2). In order to consider more dynamical data this chapter describes an alternative view on motion integration: the oculo-motor level.

Each of the primate eye is controlled by three pairs of antagonistic muscles (see Fig. 2.1) which are able to produce several types of movement to shift of stabilise gaze (see Krauzlis, 2008,[126] for a review).

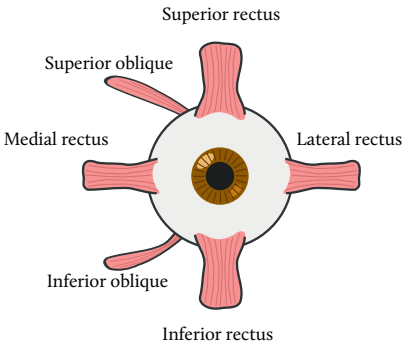


Fig. 2.1 Eye muscles controlling the eye.

The *vestibulo-ocular reflex* (VOR) and the *optokinetic response* (OKR) both stabilise gaze to compensate for head movements, although they operate at different time scales and use different sensors. Among the gaze shifting movements, *saccades* rapidly—hundreds of degrees per second—shift gaze to specific locations in the visual field, whereas *smooth pursuit* slowly—tens of degrees per second—follows moving objects to minimise the blurring of the target. *Vergence* adjusts the eye to the depth of an object. Finally *microsaccades*, interspersed among *drifts*, contribute to fixation and high visual acuity tasks.[123]

In this study we only consider *pursuit movements* as it is closely related to 2D motion integration. Indeed smooth pur-

suit eye movements are used to follow moving objects and are less consciously influenced than saccades for instance—it is really difficult to start a smooth pursuit without a moving object. Yet saccades and smooth pursuit appear to share a common functional architecture[127], and studying one can help understanding the other.

2.2 FOLLOWING LINE DRAWINGS

2.2.1 *Tracking bars*

Smooth pursuit movements were recorded using stimuli similars to the one described in the previous section. In the motor experiment associated to the translating bar, subjects are asked to follow the centre of the bar, while macaques are trained to do so.

The dynamics of the eye movements are now recorded within a single experiment and they show striking similarities, both in term of time scale and angular tracking error, with perceptual results.[241] In Fig. 2.2 we show angular tracking error for the translating bar in both human[241] and macaques,[35] and redraw the results of Fig. 1.13 for perception.

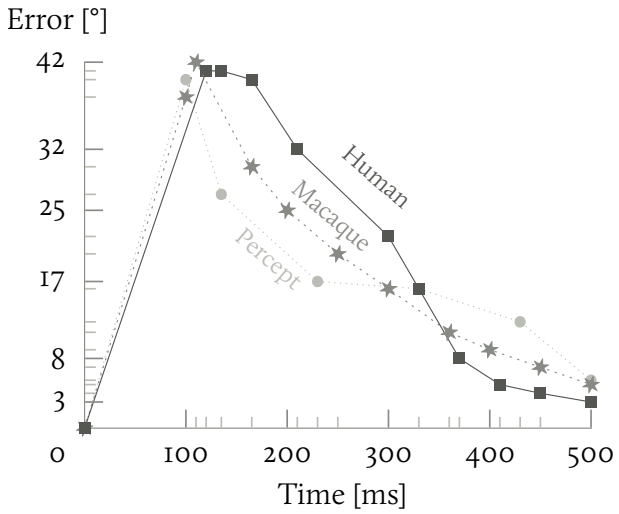


Fig. 2.2 Tracking error and perception on a translating bar. The angular error tracking error is plotted for human subjects (*squares*) and macaques (*stars*). We re-plotted results for perception (*disks*).

Again the oculo-motor system is initially biased in a direction orthogonal to the bar, and we can assume that it is subject to the aperture problem which is solved in a few hundred milliseconds.

One can assume a common mechanism at both perceptual and motor level that takes local motion cues as input. Interestingly this initial bias does not decrease if the subject knows the true direction of the moving bar before the experiment. Thus smooth pursuit eye movements provide us an easier reproducible mechanism to unravel the computations underlying motion integration.

2.2.2 *Diamonds*

In another line drawings experiments, Masson and Stone considered diamonds translating[154] diamond stimuli translating either vertically or horizontally (see Fig. 2.3). 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 44° 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.

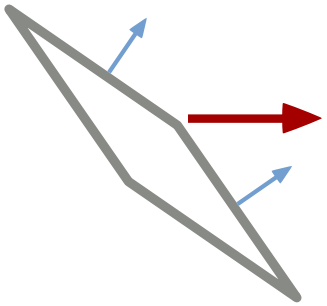


Fig. 2.3 In the translating diamonds stimulus,[154] the motion cues given by the edges of the diamonds (*blue arrows*) do not average to the real translating direction (*red arrow*), as in the type II plaids.

Changing the configuration of the stimulus, by using clockwise (CW) or counter-clockwise (CCW) stimuli, or by varying the direction of the translation, does not influence the ability to pursuit 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. It is only after a few hundred milliseconds of exponential direction error decay that the eyes correctly track the object or that human subjects report the correct direction of motion.

2.3 SPATIAL INTEGRATION

Several clues indicates that smooth pursuit involves a spatial integration of motion cues, as it is the case for motion perception (see Section 1.2.2).

Barthélemy and colleagues[20] used a drifting grating viewed through circular aperture with different sizes to investigate spatial integration. The orientation of the grating is constant and orthogonal to its drifting direction, but the diameter of the circular aperture varies among the stimuli (see Fig. 2.4).

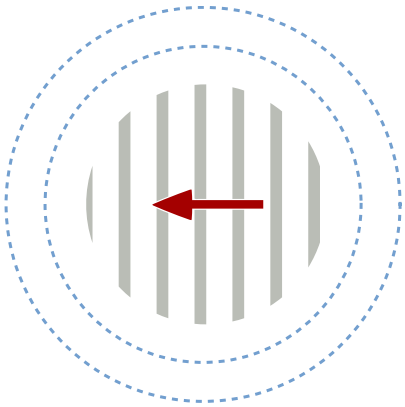


Fig. 2.4 Drifting grating viewed through different aperture sizes.

The authors quantify the change in eye direction during

several time windows with respect to the diameter of the aperture. Their goal is to provide a quantitative measure of the spatial summation area, i.e. the smallest diameter leading to the strongest change in eye position. Such spatial summation functions can be seen as a global readout of the motion integration performed in cortical area MT.

It is however also 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 on the perceptual consequences of the centre-surround interactions in early visual areas.[202] and it becomes possible to compare these results for the properties of neuronal receptive fields in various cortical areas.

In the following chapter we will study the biological substrate for the various phenomena reported in the literature.

Chapter 3 Neural architecture

In this chapter we discuss the anatomical and physiogical substrate of motion integration. Section 3.1 briefly review the neural architecture of the visual system, in particular its organisation in areas. Section 3.2 describe the functional properties of neurons in the visual cortex. Section 3.3 provides some hints on how motion is processed.

3.1 BRAIN ANATOMY

3.1.1 *Retinal input*

At the entrance of the visual system is the retina which transduces the received light into electrical impulses. The retina itself is made up several layers of neurons, from the photoreceptors catching the photons to the ganglion cells sending the signal to the rest of the visual system. Measuring and interpreting both retinal responses and the retinal computations are currently very active fields of research.[91,256,81]

The visual signals can take various paths starting from the retina. In this study we consider the retino-geniculo-cortical pathway, since we are interested in the cortical interactions and their relation to perception and dynamics. In this pathway, the signal emitted by the retinal ganglion cells is passed through the lateral geniculate nucleus (LGN), a thalamic region of the brain, and forwarded to the primary visual cortex (V1) (see Fig. 3.1). V1 is the cortical area at the entrance of the visual cortex.

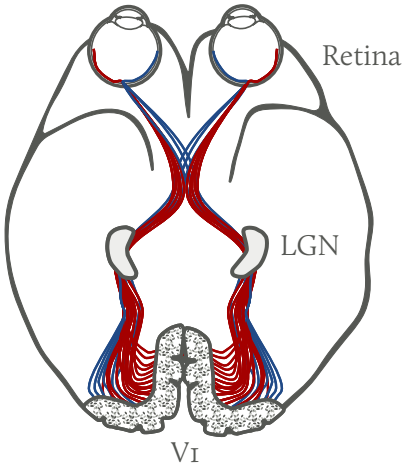


Fig. 3.1 Pathway to the visual cortex. Light received by the retina is transformed into an electrical impulse passed through the LGN and forwarded to V1.

Despite not being the focus of this study, other visual pathways can also provide useful information leading to changes in perception or oculo-motor dynamics. For instance information from the retina also flows into the *superior colliculus*, a region of the midbrain closely involved with eye movements (also called tectum in other species). Connections between the superior colliculus and other regions of the brain which are supposed to be beyond the primary visual cortex explain why some those other regions can still be activated even if the primary

visual cortex is disabled.[195,90]

3.1.2 *Cortical areas*

The cerebral cortex of primates can be divided in anatomically and functionally distinct areas called the cortical areas. Focusing on the visual cortex, one can find cortical areas dedicated to colour, motion perception, object recognition, or eye movements.

Although the separation is not so strict, a classical view is to consider Cortical areas in the visual system to be grouped in two main pathways. Areas in the *form pathway* mostly process static information such as colour or shape, and included areas dedicated to shape recognition. Areas in the *motion pathway* analyse changes in the image, integrate motion, and initiate eye movements. Both pathways have anatomically distinct cells, called *parvocellular* in the form pathway, and *magnocellular* in the motion pathway. The neurons not classified in one of those groups are known as *koniocellular*.

3.1.3 *Connections*

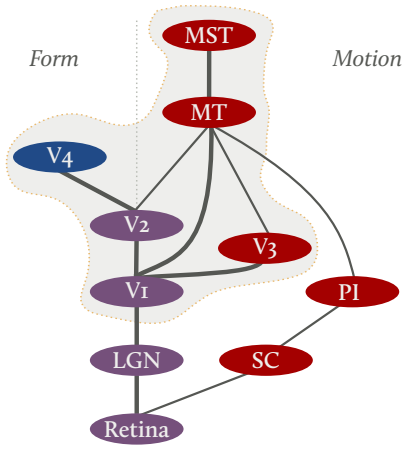


Fig. 3.2 Major areas involved with motion in the primate visual system (adapted from Felleman and van Essen, 1991[77]). Dotted region denotes the cortical areas.

Long range connections such as the one coming from the LGN to the V1 help us understand the visual information pathways. Inside a single cortical area, short range connections exists, the *lateral connections* (or *horizontal connections*) used by neurons in a single area to process information. Despite their shorter travelling distance, lateral connections are slower than connections between areas since they are not myelinated (0.1–0.2 m/s *versus* 2–6 m/s in the macaque[93]).

The information emitted by the retina is not directly transmitted to all the cortical areas in the visual cortex, and a hierarchical structure is constructed, where *higher* areas are less directly connected to the retina since information passes through a lot of *lower* areas. For instance, most of the visual input concerning motion will go from the retina to the LGN from where its projected to V1, then passed through MT, and later to MST and other higher cortical areas (see Fig. 3.2).

In addition to the forward stream carrying information from the retina to the higher level areas, a backward stream transmits

back information to areas closer to the retina. Those *feedbacks* transmit information across large regions of the visual field[8] and are proposed to play a major role in contextual information processing.

3.2 PROPERTIES OF NEURONS

3.2.1 *Retinotopy*

The first areas of the visual cortex preserve neighbourhood properties: two neighbouring neurons in the retina are connected to two neighbouring neurons in the cortical area V1. This preservation of topology is called *retinotopy* and is partially guided by chemical markers during the development phase, as first hypothesised by Sperry.[215,214,110] Those chemical markers are assumed to encode the position of neurons in the receptive field.

Yet such a chemical mechanism, also called *chemospecificity*, is likely to be complemented by other processes, in particular due to the numerous cells in the visual system areas. Among the various principles proposed to guide retinotopy, Hebbian approaches are the most popular and biologically plausible. Among them, the LISSOM model by Miikkulainen, Bednar, Choe and Sirosh[156] provides the most extensive implementation and explains the emergence of *receptive fields*.

3.2.2 *Receptive fields*

Because the connections from the retina to the visual cortex tend to maintain a neighbourhood, neurons in V1 receive input from a small region of the retina, corresponding to a small portion of the visual field. The region in the visual field that a neuron sees is called its *receptive field* and is assumed to be at the origin of the aperture problem described in Section 1.1. Indeed only a few neurons in V1 see the 2D cues such as the one at the extremities of a translating bar, and much more neurons are activated by the 1D cues.

The higher areas of the visual cortex, the larger the receptive fields are. Indeed each neuron in a cortical area receives input from multiple neurons of a lower cortical area, and its receptive field is thus a combination of the lower ones. Typical

values for the receptive fields are 1° for V1 cells or 6–10° for MT cells. Those values are however variable inside a single cortical area as they depend on the position of the receptive field of the cell. The distance of the receptive field of a cell from the centre of the visual field is called the *eccentricity*. The further the eccentricity, the larger the receptive field is, following a log-polar scheme.

3.2.3 *Motion tuning*

In each cortical area, neurons are tuned for specific features. Yet not all the neurons in an area are tuned for the same feature. For instance only 30% of the neurons in area V1 are selective to the direction of motion, but this percentage raises to 92% in area MT.[212] Indeed area V1 is not specific not motion, since it is the entry point of the visual cortex (see Section 3.1.1) and includes all kind of visual information.

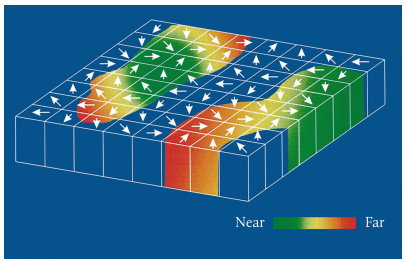


Fig. 3.3 Disparity columns in MT. From DeAngelis and Newsome.[65]

Neurons in V1 and MT are also selective to speed, but with different tuning. V1 cells prefers small speeds (around 2°/s) but MT cells are tuned to higher speeds (10-20°/s) despite taking their input from V1.[150]

Several cells in the visual system are tuned for multiple features and may thus provide the substrate for cross-modality interactions. In MT, DeAngelis and Newsome[65] found patches of direction selective cells which are also part of binocular disparity columns (see Fig. 3.3). Based on recordings from Basole and colleagues[21] in the ferret, Mante and Carandini[148] postulate that the tuning of V1 cells for motion direction is multiplexed with orientation preference, viewed a blob in the spatio-temporal frequency domain.

3.2.4 *Receptive field structure*

The receptive field is often structured, in that different regions have different contributions to the neuron response. Moreover some regions of the receptive field can be inhibitory, meaning that neuron response will be lower if this part of the receptive field is also stimulated, rather than the excitatory part only. Furthermore receptive fields can be divided into two parts: its centre, also called *classical receptive field* (CRF), and the *surround*.

An activation of the centre is required to elicit a response of the neuron, but action in the surround can modulate this response. Activation in the surround alone, even if it is excitatory, cannot elicit a response. In Fig. 3.4 we show the three types of inhibitive surround found by Xiao and colleagues in MT.[258]

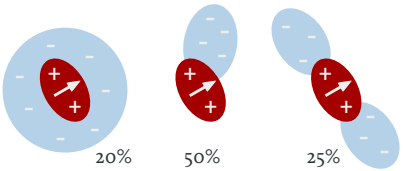


Fig. 3.4 Three main types of surrounds found by Xiao and colleagues[258] in the cortical area MT.

In addition to spatial receptive fields, neurons in the visual cortex are tuned for certain properties, like orientation or spatio-temporal frequencies. Those properties are sometimes combined together in the more general concept of spatio-temporal receptive fields often defined in the frequency domain. Indeed it is not sufficient to consider the tuning of neurons separately of its other spatial properties. For instance the preferred orientation tuning of V1 cells changes with direction or speed,[21] a change easily understandable if we consider the spatio-temporal tuning of the neurons.[148]

3.2.5 *Receptive field dynamics*

Most models explaining the visual processing at work during the early stages of the brain assume that the tuning of the neurons, i.e. their receptive fields, does not change during the experiments. However various experiments in the literature show that the receptive fields are reshaped according to various factors: the time of the experiment, the contrast in the stimulus, the locus of attentions, etc. Receptive fields reshaping is also a mechanism found at all the levels of the visual cortex, e.g. in the retina,[112] in the LGN,[155] in V1,[141] in V4,[159] in MT,[128] or in LIP.[28] In this section we focus on motion processing cortical areas, namely V1 and MT.

Spatial changes

Pack and colleagues[174] recorded changes in the receptive field size varying the contrast. For high contrast data, the recorded cells show a suppression of their activity as the stimuli size increase. On the contrary low contrast data, the recorded cell show only an increased activity while the stimuli size increase. These results are consistent with segmentation at high contrast, and diffusion at low contrast, as observed in the literature.

Experimentalists proved that the classical centre-surround

receptive fields of MT neurons is not fixed spatially but can be modified by attention.[257,10] Having the attention fixed at a certain point moves the spatial receptive field structure towards the attended point. Moreover the spatial structure is changed depending on whether the fixated point is inside or outside the classical receptive field. If the attended point is inside, the receptive field is shrunk, while it is expanded if the point is outside.

Tuning changes

Probing the response of MT monkeys cells with plaids, Pack, Berezovskii and Born[169] shown that their preferred direction changes over time. As shown in Fig. 3.5 the direction tuning of an MT neuron evolves over time. Initially neurons respond to the two 1D gratings motion, and later converge towards the average 2D plaid motion. The convergence does not appear in anaesthetised monkeys.

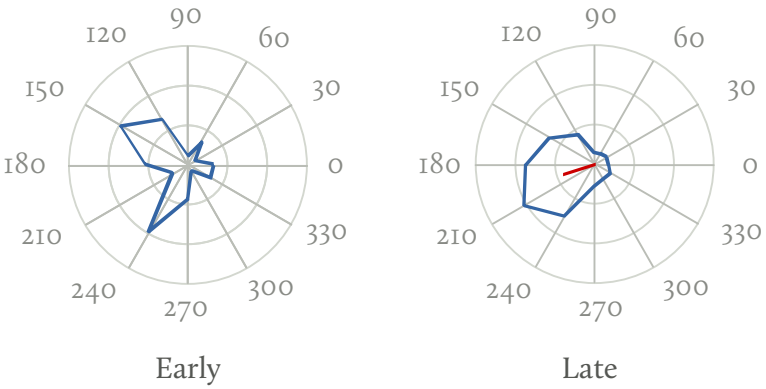


Fig. 3.5 MT neuron response to a plaid stimulus (from Pack, Berezovskii and Born[169]). *Left:* Direction tuning for the first 20 ms. *Right:* Direction tuning averaged over the last 1500 ms. Outer circle correspond to 90 spikes/s.

Similar dynamics can be found by probing the firing-rate of MT neurons with barber pole stimuli:[172] Although during the early stage, MT neurons have their preferred direction similar to when gratings are used, after a time the preferred direction moves towards the orientation of the barber pole.

As in the aperture problem, the shift of the preferred direction can be explained thanks to delayed inhibitive mechanisms, as shown by Escobar and colleagues in V1.[74]

Attention can also modulate the gain in neuron responses,

in particular it can increase or decrease the responses in MT neurons for a specific direction.[179]

Pattern versus *component cells*

Pack and colleagues shown that the response of an MT cells to plaids can vary over time to shift from the two component motions, to the pattern motion. Movshon and colleagues qualified neurons having this property *pattern direction selective* (PDS).[161] But other MT neurons, the *component direction selective* (CDS), are shown to enhance their component motion preference.

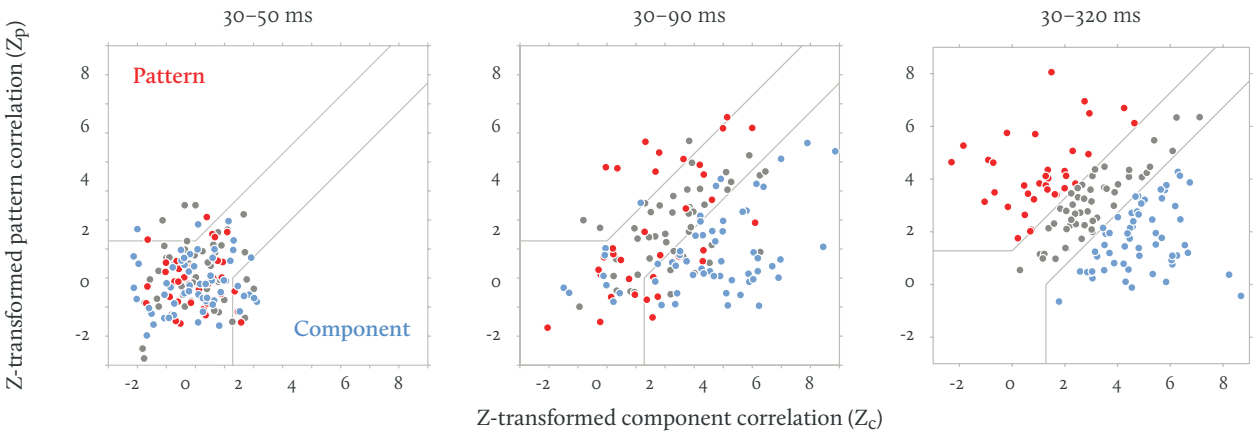


Fig. 3.6 Evolution of cell classification as pattern or component in MT (from Smith, Majaj and Movshon[211]).

Using pattern and component correlations, Smith, Majaj and Movshon shown that the pattern versus component cell classification evolves over time.[211] During the first milliseconds, all neurons are unclassified, and their correlations changes during the first 300 ms (see Fig. 3.6). Note that component cells are classified sooner than pattern cells.

3.3 NEURAL COMPUTATION

3.3.1 *Pooling of V1 responses*

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.[34] In macaque area MT, neurons solving the aperture problem have been found by various studies, using different 2D motion stimuli.[162,171,172,211]

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,[162] albeit some neurons seems to act as local features detectors such as end-stopped cells.[116,176] Thus, there seems to be a good intuition that 2D motion computation is a two-stage mechanism with local extraction feeding global integration.

3.3.2 *Aperture problem*

Interestingly, when presented with a set of small oriented bars, direction selectivity of MT neurons exhibit dynamics similar to the one observed at the tracking (see Section 2.2) and perception (see Section 1.1) levels: their optimal direction slowly rotating from the component orthogonal to the bars orientation to the 2D motion direction over a 150 ms response period.[171]

Chapter 4 Existing motion models

This chapter reviews three models of motion integration among the numerous propositions in the literature. The selected models share common mechanisms that we believe to be fundamental to understand the links between cortical activity and the dynamics of motion perception and smooth pursuit. The model of Simoncelli and Heeger presented in Section 4.1 combines half-squared divisive inhibition and the doughnut mechanism to account for responses in individual MT cells. The detailed models of Grossberg, Mingolla and colleagues, presented in Section 4.2, explain a wide range of perceptual results based on neurons interactions. The model of Bayerl and Neumann presented in Section 4.3 investigates the role of simple biological mechanisms and apply them to computer vision problems.

4.1 SIMONCELLI AND HEEGER (1998)

Heeger and Simoncelli proposed a linear/non-linear model of motion integration mapping the cortical areas V1 and MT.[208] Their model successfully reproduces responses of V1 and MT neurons for several experiments with gratings, plaids and random dots.

4.1.1 *Model description*

Computations in the model are mapped to three populations of the visual cortex.

The model considers a greyscale video $I(t, x) : \mathbb{R}^+ \times \Omega \in \mathbb{R}^+$ as input, where t denotes time, and $\Omega \in \mathbb{R}^2$ is the considered spatial domain. Response of *V1 simple cells* is denoted by $L(t, n) : \mathbb{R}^+ \times \mathcal{O}$, where $\mathcal{O}$ is the set of orientations along which the input image is filtered. 28 different orientations are sampled to filter the input with third-derivative Gaussian filters. Then, the *V1 complex cells* response is computed by averaging simple cells responses having similar orientation and phase. Finally,

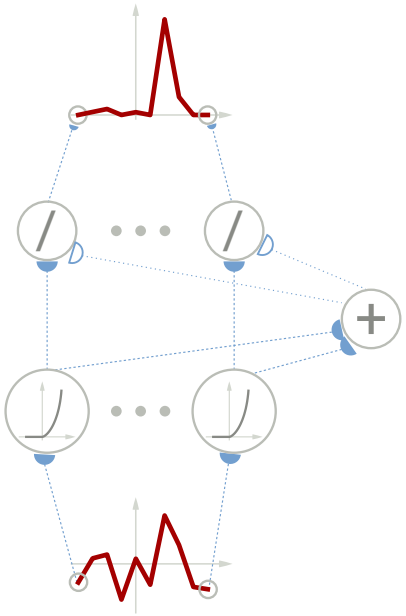


Fig. 4.1 Inhibition in the model of Simoncelli and Heeger.[208] First the input signal is half-squared, then divided by the sum of the half-rectified signals, thus implementing a selection mechanism.

the computation of MT cells responses combines the complex cells output using a divisive inhibition.

The divisive inhibition is applied at two stage in the model: at the output of the filters, and at the output of MT. It modulates a single cell activity $p(t,i)$ tuned for any given feature i by the population average (see Fig. 4.1):

$$q(t,i) = \lambda_1 \frac{|p(t,i)|_+^2}{\sum_j |p(t,j)|_+^2 + \lambda_2}, \tag{4.1}$$

where $|x|_+ = \max(0,x)$ denotes a positive rectification to account for the firing rate activity, and $\lambda_{1,2}$ are constants.

4.1.2 *Discussion*

The authors propose a *doughnut mechanism* to combine the responses of the cortical area V1 which act more like filters, into direction selective cells in MT. Indeed, the receptive field of V1 cells can be characterised by blobs in the spatio-temporal frequency domain.[124] Moreover, the frequency response of various stimuli, moving with the same velocity but with different frequencies, all lies on a plane. Hence, Simoncelli and Heeger propose that MT cells tuned for a given velocity pool the responses of V1 cells lying on the plane corresponding to their tuning.

One should note that the filters used in this model are not causal.

4.2 Grossberg and Mingolla (1997, 2001, 2007)

Grossberg, Mingolla and colleagues proposed several incarnations of their model incorporating form and motion processing to reproduce perception on a wide range of stimuli. The authors map their model to various cell populations in cortical areas V1, MT and MST.

4.2.1 *Model description*

Chey, Grossberg and Mingolla model the initial stages of motion detection in details using transient and directional cells to compute motion direction.[56] The three other cell populations

model lateral interactions (spatial integration and centre/surround mechanisms).

Each population of the various incarnations of the models defines the activity p of its cells as a membrane equation :

$$C_m \frac{\partial p}{\partial t} = -(p - E_e)g_e - (p - E_i)g_i - (p - E_l)g_l \qquad (4.2)$$

where g_e and g_i represent the total inputs from excitatory and inhibitory neurons synapsing on the cell, i.e. their receptive fields, g_l is a leak conductance, and the $E_.$ are reversal potentials.

One of the interesting features of the model is the detection of motion. Whereas other approaches take spatio-temporal filters or correlation detectors responses as input, the author model the mechanisms leading for detecting motion in their first cell populations. Moreover they investigate the role of ON and OFF streams as found in the LGN and V1.

Motion detection is handled by the following populations:

- Transient cells detect changes in the input and are divided into ON and OFF channels.
- Directional interneurons integrate transient cells response by receiving excitatory input from activity at the same position and inhibitory input from directional interneurons sensitive to opposite direction.
- Directional transient cells received their input from transient cells and directional interneurons.
- Short-range motion filters accumulate motion in each direction at two different spatial scales.

Cells are classified as either belonging to the ON or the OFF channels (for the first populations), or one of the two spatial scales (for highest populations). Height possible motion directions are considered in the model and speed (motion magnitude) is not handled.

In a subsequent version of the model,[96] the authors included recurrent connectivity between higher and lower cortical areas, as well as form modulation using their FORMOTION model.[94] In

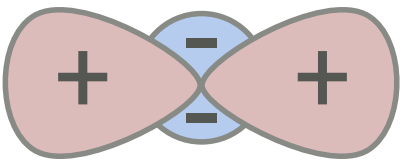


Fig. 4.2 Bipole cells of Grossberg and Mingolla. Due to their peculiar centre/surround, the inhibits information from T-junctions.

particular the peculiar centre-surround organisation of *bipole cells* (see Fig. 4.2) allows the detection of some extrinsic junctions, in particular in the chopstick illusion.

4.2.2 *Discussion*

In comparison to the model of Heeger and Simoncelli, the authors proposed recurrent model incorporating spatial interactions. They extensively studied the stimuli proposed by the psychophysics literature and successfully reproduced the perceptual results. In addition, Grossberg and Mingolla implement a *dynamical model* in which the output of a cell is not fixed but varies in time. This dynamics allows to see the resolution of the aperture problem in time and can also serve as input to other mechanism such as smooth pursuit.[173]

4.3 BAYERL AND NEUMANN (2004)

The model of Bayerl and Neumann provides an interesting small framework to study motion perception. They propose a recurrently connected system mapping V1 and MT, and apply it to a several simple synthetic stimuli to understand resolution of the aperture problem. An important contribution of their model is that it is also able to process large video sequences such as the one used in computer vision and provide good results. The authors also investigated several extensions of this model.

4.3.1 *Local motion estimation*

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

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

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 ∂^2_α denotes the second order directional derivative in the direction $\alpha \in \mathcal{O}$.

From these filtered outputs, the Bayerl and Neumann 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.

Due to its simple two-frame correlation, this local motion estimation is limited in its frequency support, and is not equivalent to the *elaborated Reichardt detectors*[236] and thus not to the energy models.

4.3.2 *Model description*

The two cortical areas implemented by Bayerl and Neumann can be described by their activity:

$$p_i : (t,x,v) \in R^+ \times \Omega \times \mathcal{V} \to p_i(t,x,v) \in [0,1], \quad i \in \{1,2\}, \tag{4.3}$$

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.

Processing in each cortical areas follows the following stages:

$$
\begin{aligned}
m_1(t,x,v) &= k_1(t,x,v)\,(1+100\,p_2(t,x,v)),\\
n_1(t,x,v) &= m_i^2(t,x,v) \overset{v}{*} G_\sigma,\\
p_1(t,x,v) &= \frac{n_1(t,x,v) - (1/2|\mathcal{V}|)\ \sum_{w\in\mathcal{V}} n_1(t,x,w)}{0.01 + \sum_{w\in\mathcal{V}} n_1(t,x,w)},
\end{aligned}
$$

$$
\begin{aligned}
n_2(t,x,v) &= p_1^2(t,x,v) \overset{x,v}{*} G_\sigma,\\
p_2(t,x,v) &= \frac{n_2(t,x,v) - (1/2|\mathcal{V}|)\ \sum_{w\in\mathcal{V}} n_2(t,x,w)}{0.01 + \sum_{w\in\mathcal{V}} n_2(t,x,w)},
\end{aligned}
$$

where m_i and n_i are intermediate stages to compute p_i, k_1 is the local motion input described in Section 4.3.1, $\overset{v}{*}$ denotes convolution with respect to the spatial domain with a Gaussian whose standard deviation is 0.75 and $\overset{x,v}{*}$ denotes convolution with respect to the spatial and velocity domains with a Gaussian whose standard deviations are 0.75 (velocity) and 7 (spatial).

4.3.3 *Discussion*

The model of Bayerl and Neumann[24] was successfully applied to moving squares with fixed velocities. In addition it proves that a simple model can achieve good performances with computer vision videos. In this respect it can be linked to the model of Castellanos Sánchez, Giraud and Alexandre[49] in which computer vision sequences are used as input, and later implemented on FPGA.[233] Bayerl and Neumann also extended it to support transparent motion and incorporate form cues. Those extensions allow the model to report correct results on random dots stimuli and overlapping squares.

The model uses a divisive inhibition similar to the one proposed by Simoncelli and Heeger[208] (see Section 4.1) but extend it with spatial interactions.

Compared to Grossberg and Mingolla, there are no dynamics and the modelling is less detailled, allowing comparison with perceptual recordings to a lesser extent. As with the models of Grossberg and Mingolla, the model incorporates spatial interactions and recurrent connectivity. The main difference is the

small set of equations linking in a fashion similar to Heeger and Simoncelli, and the focus on computer vision stimuli, at the expense of more perceptual experiments.

Part II

Models for motion integration

Part 2 Models for motion integration

In Chapter 5 we describe the common architecture of our models and discuss its biological plausibility. Our models consider the dynamical activity in two cortical areas involved with motion integration, V1 and MT. We also describe the input of the models and give a readout to compare the results to motion dynamics such as smooth pursuit.

The first incarnation of our model is given in Chapter 6. We show that a small set of key characteristics is sufficient to explain a wide variety of motion percepts. Moreover the dynamics of our model allow us to compare its output to perceptual and behavioural recordings and link them to the computation of motion in the visual cortex.

Yet some stimuli require more complex mechanism to solve motion ambiguities. In Chapter 7 we propose an extension of our initial model in which motion cues are gated by luminance information from a cortical area in processing form. Such a luminance modulated diffusion is an alternative to more complex models relying on features detectors and proposes biological foundations to the motion segmentation problem.

Finally, we transpose our initial model into the neural field formalism in Chapter 8. We show that this change of framework does not impact the results obtained using our initial model, and allows us to prove several properties of our model. We prove the existence and uniqueness of the solution and experiment properties of the selection mechanism.

Chapter 5 Models architecture

In this chapter we describe the common architecture of our models and briefly discuss its plausibility. In Section 5.1 we introduce the rationale of our models focusing on its spatial and dynamical properties. The general description of our models is given in Section 5.2. We define an input to our models and a readout to compare to smooth pursuit dynamics in Section 5.3. Section 5.4 gives some implementation details.

MAIN CONTRIBUTIONS

- A dynamical cortical architecture to study motion integration.
- A model of *readout* to compare output to smooth pursuit.

5.1 INTRODUCTION

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.[152]

There are two aspects that have been largely ignored by most of the two-stage feedforward models.[162,255,208,199]

First, motion integration is intrinsically a spatial process. Since most of the natural objects are rigid, propagating non-ambiguous motion information is an essential aspect of motion integration.[106,164,98,248] The role of such diffusion process has only been investigated in a small number of biologically inspired models. Grossberg and colleagues[94,32] investigated how local form and motion cues can be integrated through recurrent diffusion (see Section 4.2). A similar solution was also de-

veloped by Bayerl and Neumann,[22,23] albeit with a more simple and realistic motion computation algorithm (see Section 4.3).

Second, 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 (see Section 1.1). And similar dynamics have been found with smooth pursuit eye movements (see Section 2.2). Such temporal courses can reflect the dynamical neural solution to the aperture problem. Indeed, over a time course of several tens of milliseconds, neurons in the area MT solve the aperture problem, so that late but not early preferred direction corresponds to pattern motion direction.[171,172,211]

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. Moreover, our models suggests that the dynamics of spatial integration and the time course of motion perception can be intrinsically linked.

5.2 DESCRIPTION OF THE MODELS

5.2.1 *Global structure*

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. Motion information is extracted and processed at different spatial scales within layers that are recurrently interconnected. As illustrated in Fig. 5.1, the common basis of our models is described by three layers of motion processing.

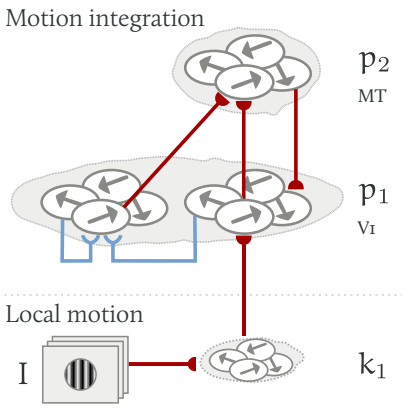


Fig. 5.1 Schematic view of the model. Recurrently connected areas p_1 and p_2 implement motion integration and take local motion cues k_1 as input.

The first layer extracts local motion energy through spatio-temporal filtering, corresponding to simple and complex cells of the primary visual cortex.[208]

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.[190]

The third layer implements MT neurons, which integrate motion over larger portions of the image through the conver-

gence of cortical layer 2 cells. Our MT-like layer 3neurons have larger receptive fields and are tuned for lower spatial frequencies and higher speed than striate-like layer 2 cells. This fact is consistent with the view that V1 and MT operate at different scales.[34] Feed-forward models of motion integration are heavily rooted on such evidence[208,199,255,147] and we will compare our results to one of them. However, V1 and MT areas are recurrently interconnected.[207] Existing models have shown that such recurrent connectivity can play a role in solving the aperture problem in synthetic and natural sequences[56,24,22] as well as implementing contextual effects observed in V1 and MT neurons.[8]

Dynamics are the major innovation of our model. Indeed, we do not consider

5.2.2 *Models overview*

Our model implements interactions between several layers processing motion information. The model estimates dynamically the velocity information given an input grey level image sequence denoted by:

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

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

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.

$$p_i : (t, x, v) \in \mathbb{R}^+ \times \Omega \times \mathcal{V} \to p_i(t, x, v) \in [0, 1], \quad i \in \{1, 2\}, \tag{5.1}$$

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 k_1 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 imple-

mentation of detection and integration stages that correspond to cortical areas V1 and MT.

The coupling between layers defines the connectivity rules using a set of coupled differential equations. With that respect, our model follows some previous contributions.[56,24,32] Forward connections transmit information from layers closer to the eye to layers deeper in the system while backward connections project back to the areas closer to the eye as discussed in Section 3.1.3. 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.

5.3 INPUT AND OUTPUT

5.3.1 *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.[193,236,245,1] 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.[102,208,199] 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.[85,208,68] In addition, mechanisms to combine the output of such filters have been largely studied. For instance, the *doughnut mechanism* is described and studied by Simoncelli and Heeger.[208]

* See for example Section 4.3.1 where we describe an alternative local motion detection based on correlation of two frames.

Our local motion input is based on an energy model computed from the filters of Derparnis and Gryn,[68] namely the second derivative of a Gaussian and its quadrature pair—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 proposed by Alexiadis and Sergiadis.[4] This choice is mo-

tivated 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:

$$f^r(t,x,v) = \sum_{n=0}^{N} \left(\sum_{m=1}^{M} t_m^r(s_n^r(v))(y_m^r \overset{x}{*} I)(t,x,v) \right)^2, \quad (5.2)$$

where $r \in \{o,e\}$, f^o and f^e are the odd and even responses of the filters, N is the order of the chosen filters, $M = (N+1)(N+2)/2$, 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 $\overset{x}{*}$ denotes convolution with respect to the spatial domain.

Then, based on the expression (5.2), we defined the activity (energy) of our first layer k_1 by:

$$k_1(t,x,v) = f^o(t,x,v) + f^e(t,x,v). \quad (5.3)$$

5.3.2 *Defining readouts*

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 provide at each position x the instantaneous activity of a neuron tuned for the velocity v, as shown in Fig. 5.2.

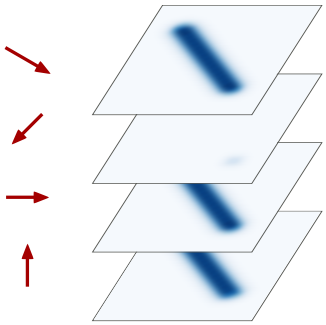


Fig. 5.2 Distributed activity response for a translating bar.

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:

$$m_i(t,x) = \frac{\sum_{v \in \mathcal{V}} p_i(t,x,v)\, v}{\sum_{v \in \mathcal{V}} p_i(t,x,v)}, \quad i \in \{1,2\}. \quad (5.4)$$

This velocity field can be represented either by arrows or by a colour coded image indicating speed and direction (see Fig. 5.3). Here, we use the Middlebury colour code[15] as illustrated in Fig. 5.4. This colour code emerged as the *de facto* standard in the optical flow computer vision community and it is motivated by colour

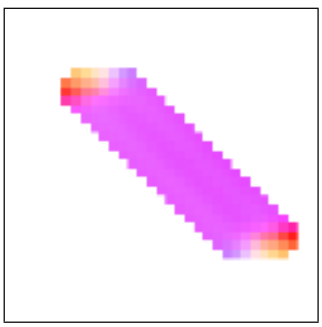


Fig. 5.3 Optical flow representation by sampled velocities or colour code.

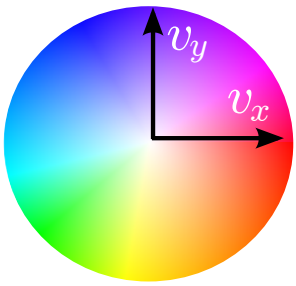


Fig. 5.4 Middlebury colour code. For each velocity, direction is represented by the hue, and speed by the saturation.

perception experiments. It associates a single colour to each velocity. 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 readout such as the eye movement 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 readout 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:

$$\frac{dw}{dt}(t) = \lambda \left(\sum_{x \in \Omega} m_2(t, x) - w(t) \right), \tag{5.5}$$

where m_2 is defined by (5.4). Thanks to the definition of this readout, we will define in Section 6.2.1 an estimated direction errors, so that direct comparisons with biological data will be possible.

5.4 IMPLEMENTATION DETAILS

In this thesis we document the performance of our models 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.

As far 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 ms. The set of possible velocities $\mathcal{V}$ also needs to be sampled. We chose $\mathcal{V} = [-3, 3]^2 \in \mathbb{Z}^2$ so that the velocities are sampled in a 7×7 pixels grid.

Simulations were written in the Python programming language and used the SciPy scientific library.[120] Integration was performed using a fourth-order Runge-Kutta (RK4) method with at most 10 iterations between two successive input frames. Luminance-gating described in Chapter 7 was implemented on GPU using the CUDA library.

A notable exception is the work for the SEARISE European project which mostly uses C++ source code, GPGPU technologies, and relies on frame-by-frame approaches (see Chapter A).

Chapter 6 Dynamics of motion integration

In this chapter we propose a dynamical model of motion integration accounting for perception and smooth pursuit eye movements and based on our common architecture. The model is described in Section 6.1 by the activity in two cortical areas, V1 and MT. In Section 6.2 we show that a variety of motion percept and dynamics are taken into account. Results are given for line drawings and gratings stimuli. We conclude in Section 6.3 by discussing the dynamics of the model, its relation with existing approaches, and its limitations.

MAIN CONTRIBUTIONS

- A model explaining a wide range of motion percept
- And accounting for smooth pursuit dynamics

6.1 MODEL

6.1.1 *General connectivity*

Given the activity k_1, the core of our model is defined by the interactions between the two layers p_1 and p_2, which are modelled by two coupled differential equations:

$$\frac{\partial p_1}{\partial t} = -\lambda_1 p_1 + (1 - p_1) \left| \lambda_1^f k_1 + \lambda^b k_1 p_2 - \lambda_1^l G_{\sigma_1^l} \overset{x}{*} \int_{\mathcal{V}} p_1(t, x, w)\, dw \right|_+, \tag{6.1}$$

$$\frac{\partial p_2}{\partial t} = -\lambda_2 p_2 + (1 - p_2) \left| \lambda_2^f G_{\sigma_2^f} \overset{x}{*} p_1 - \lambda_2^l G_{\sigma_2^l} \overset{x}{*} \int_{\mathcal{V}} p_2(t, x, w)\, dw \right|_+, \tag{6.2}$$

where G_σ is a Gaussian function of variance σ, λ's and σ's are constants, and $|x|_+ = \max(0, x)$ denotes a positive rectifica-

tion.

The two main characteristics of our model (6.1)–(6.2) are summarised as follows:

- FEEDBACK, from p_2 to p_1, which is modulated by $\lambda_b k_1$ in a multiplicative way.[24] Therefore we used a modulating rather than driving feedback, similar to that found in studies of the motion processing system in primates.[207]
- LATERAL INHIBITION, 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 selection or winner-take-all mechanism.[64,262] Instead of the divisive inhibition as found in some models,[167,24] we implemented a subtractive inhibition.

6.1.2 *Parametrisation*

The model is fully specified by a set of 12 parameters. These parameters, whose values are given in Table 6.1, were found by matching the time scale dynamics of psychophysical experiments. The simple line drawing stimulus was used to fit the parameters that were then kept constant for all other motion stimuli.

In addition to the time scale matching procedure, we also investigated the role of the parameters. For instance, the λ_1^l and λ_2^l parameters representing the weight of the inhibition are necessary to achieve a selection like mechanism.[262,64] We evaluated the acceptable range for those inhibition parameters to be between 1.8 and 8.0.

Table 6.1 Chosen parameters setting for our dynamical system.

(6.1)	$\lambda_1 = 2.0$	$\lambda_1^f = 1.0$	$\lambda_1^b = 24.0$	$\lambda_1^l = 4.0$	$\sigma_1 = 2.0$
(6.2)	$\lambda_2 = 2.0$	$\lambda_2^f = 16.0$	$\sigma_2^f = 8.0$	$\lambda_2^l = 4.0$	$\sigma_2 = 2.0$

6.2 EXPERIMENTAL RESULTS

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. Second we continue our exploration by using different plaid patterns. Finally we describe the effect of the aperture shape on 2D information and its consequence for motion perception.

6.2.1 *Dynamics on 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: the translating bar stimulus detailed. For short durations, its perceived direction is biased towards the direction orthogonal to its orientation (see Section 1.1). Consistently, it has been demonstrated that initial tracking direction exhibits the same bias (see Section 2.2).

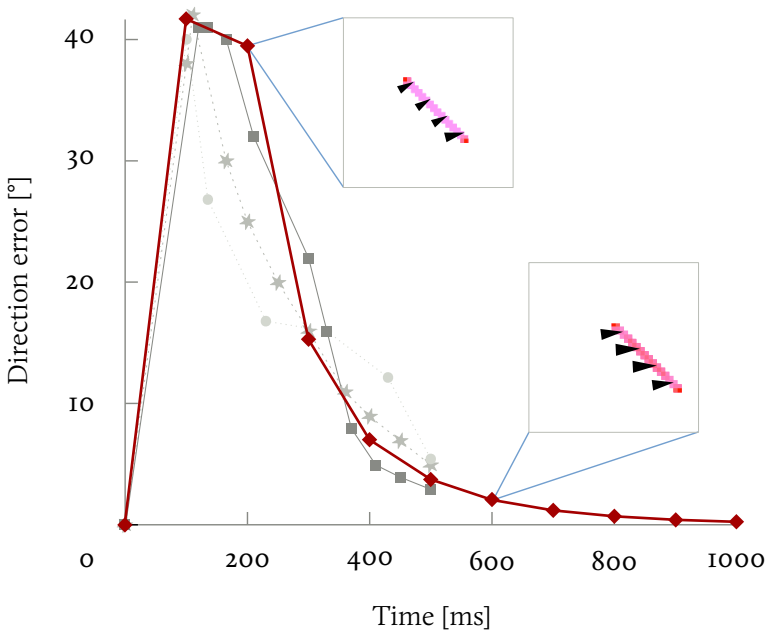


Fig. 6.1 Temporal evolution of our model estimated direction error $w(t)$ on a 45° tilted bar, moving rightward (*diamonds*). For comparison, we replotted human and macaque tracking error, as well as perception error from Fig. 2.2 (*squares, stars, disks*). The optical flow $m_2(t, x)$ is shown at 200 and 600 ms.

We define the *observed direction error* 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,[34,171,172] the perceived direction[144,51,205] or the tracking direction of smooth pursuit

eye movements[24,1,35,158] (see Fig. 2.2).

It should be noted that since the observed direction error is an *angular* error computed from motion, it is highly imprecise during the first hundred milliseconds. At that period of time, responses are slow, noisy and rapidly varying so that computation of the effective angles becomes unstable.

In order to compare our results with experimental data, we define the *estimated direction error* as the difference between the angles of the true translation direction and our global readout $w(t)$. As illustrated in Fig. 6.1, 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. We found a smooth 2D motion diffusion inside the bar as shown by the gradual evolution of the velocity fields. Thus our model can solve the aperture problem at both local and global scales. After a short period of time where the direction error stays constant at about 40°, the estimate of the global motion converged to the true direction—a null direction error—with an exponential decay. It should be noted that the dynamics observed at output stage of our model closely mimics the experimental data measured for both pursuit and perception.

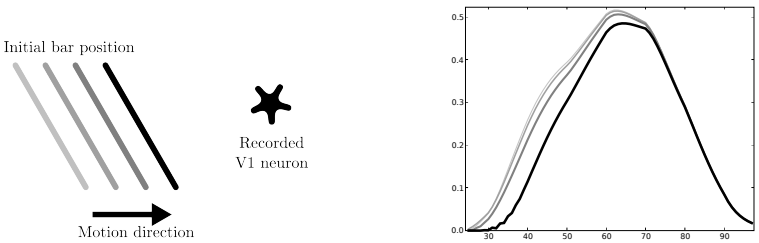
Motion anticipation

In the motion anticipation experiment, we record the response of a V1 neuron in our model to the translating bar stimulus with different initial spatial position (see Fig. 6.2). The further the initial position is from the V1 receptive field, the sooner the V1 response starts. Thus, a motion anticipation mechanism seems to enlarge the receptive field of the neuron in direction of the initial bar position. Yet such a mechanism is highly limited by our purely multiplicative feedback.

Variations of the translating bar

Next, we introduce two variations to test the model behaviour.

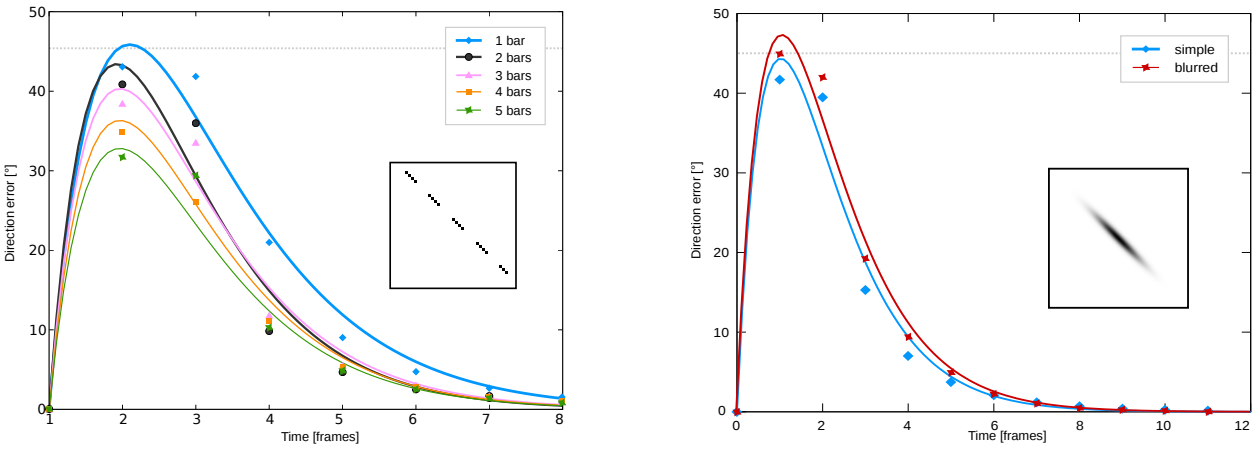
First, we show in Fig. 6.3 (a) the changes in direction error when the bar is cut into an increasing number of line segments.



(a) (b)

Fig. 6.2 Anticipation in our V1 model cells. (a) Translating bars are presented from different distance to the recorded cell. (b) The farther the distance, the sooner the V1 response starts.

Introducing more line-endings both reduced the initial bias in the global motion estimation (from 44 to 32°) and produced a faster exponential decay of the direction error, a results found in the psychophysical literature.[144] Similar changes were reported when filling a moving diamond with 2D texture elements.[241] On the contrary, smoothing the luminance profile by applying a Gaussian filter along the bar orientation reduced the contrast of line-endings (see Fig. 6.3 (b)) and thus resulted in a larger initial bias, reaching the asymptotic error of 45° and a somewhat longer time constant for error reduction. Similar results were reported with smooth pursuit eye movement in humans.[240]



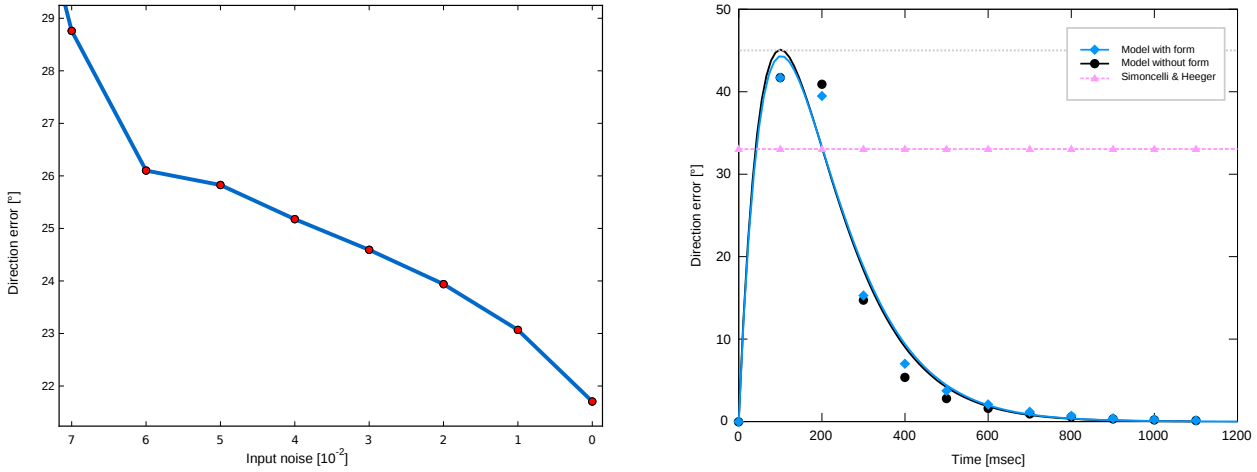
(a) (b)

Fig. 6.3 Varying strength of 2D motion cues. (a) A tilted line is cut into small segments, introducing more line-endings. Direction error is plotted against time for different numbers of segments in the 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 – larger bias – and a slower time course for computing the exact translation of the bar.

Second, we show in Fig. 6.4 (a) how the early direction error depends on the level of noise added to the input stimulus. We considered additive Gaussian noise with different variances. The early direction error was estimated at a fixed time $t = 250$ ms, around the steepest decrease.

Similar to the effects of contrast which have been observed in both psychophysical studies[51] and behavioural studies,[241] 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.



(a) (b)

Fig. 6.4 (a) Early direction error as a function of the variance of the Gaussian noise added to the input. Similar to the effects of contrast which have been observed in both psychophysical studies[51] and behavioural studies[241], higher levels of noise resulted in larger initial biases. (b) Direction error dynamics on the translating bar: Comparisons between Simoncelli and Heeger's model[208] and ours.

Still our model performs much better than the static model of Simoncelli and Heeger [208]

that we use for comparison. Output of their model is plotted as continuous dotted line in Fig. 6.4 (b). 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 and Heeger.

Rotating ellipses

To conclude this line-drawing section, we briefly mention that similar psychophysical observations with other types of line-drawing objects.[143,204,154] Our output was always consistent with experimental data, for both initial bias estimate and time course. One interesting example is given by rotating ellipses.[244] Weiss and Adelson[248] 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. With *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 Fig. 6.5, 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.

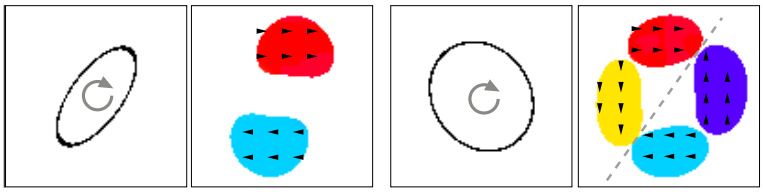


Fig. 6.5 Model response to the gelatinous ellipses. (a) We first process a thin ellipse of ratio 9:20 and the resulting motion (b) is compatible with rotation. (c) We then process a thick ellipse of ration 3:4 and the resulting motion (d) is a deformation incompatible with rotation *left downward yellow* patch and *violet upward right* patch should be inverted, and not pointing towards the *diagonal line*.

6.2.2 *Dynamics of pattern motion using plaids*

Plaid patterns have been largely studied to elucidate 2D motion integration both at psychophysical level[2,79,92,260] and physiological level.[162,194] One interesting aspect of plaid motion is that, depending on the relative direction of the two components, different perceived directions can be predicted from the different computational solutions proposed so far: vector averaging (VA), intersection-of-constraints (IOC) or 2D feature tracking. 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.[171,211] Such neuronal dynamics could explain why perceived direction[260] as well as eye tracking direction[151] 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[79] as well as unikinetic plaid patterns.[92] Fig. 6.6 (a) illustrates the model output in response to a type II plaid such as used in [40]. 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 estim-

ate 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. Fig. 6.6 (b) illustrates the model performance for another type of plaid. With unikinetic plaids, the IOC solution cannot be applied since only one component is drifting. The VA solution collapses 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 2D features tracking 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 δ in Fig. 6.6 (b)), similar to that observed in humans[151] and monkeys.[18]

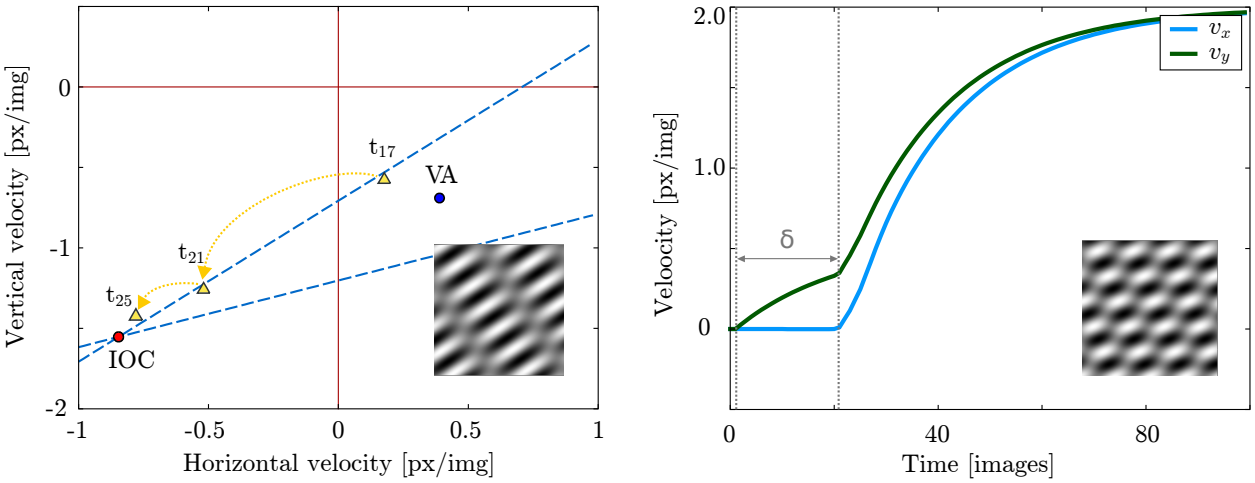


Fig. 6.6 Model responses to plaid pattern motion. (a) Model output obtained with a type II plaids where the two component directions are separated by 25°.[40] The temporal dynamics is illustrated by the instantaneous output direction at three different points in time (triangles). In the same plot, the predictions made by the vector average (VA) and intersection-of-constraints (IOC) models are illustrated. Note that IOC response is similar to 2D features tracking response in this case. (b) Response to an unikinetic plaid.[151] The initial response following the moving plaid switches with time. Note that we observe a delay δ between the vertical and horizontal responses as described for eye movements.[151]

6.2.3 *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 rectan-

gular 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.[243] The bias depend of the aspect ratio, defined by ratio between the long and short axes of the aperture, and increases with it. Moreover, human ocular tracking[153] 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).[172]

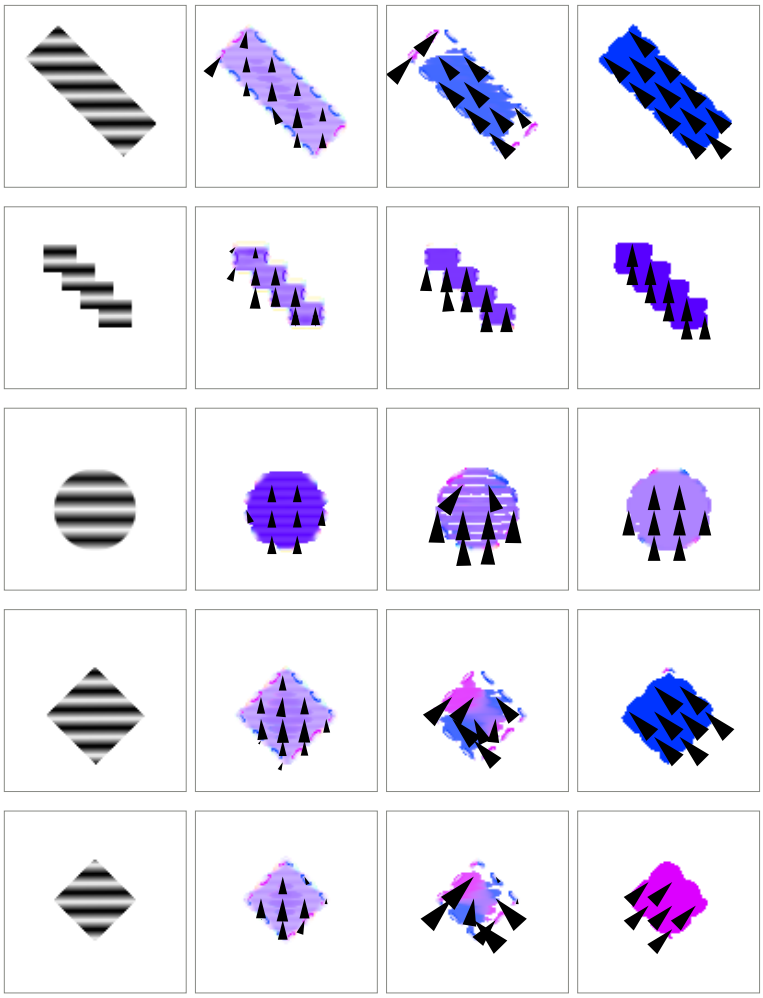


Fig. 6.7 Model response to upward moving gratings presented behind various aperture shapes. For a given row each column shows respectively one image of the stimulus, and three representation of the motion field for the initial response, intermediate response, and steady state. We tested the following 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.

Our model reproduces these different aspects of motion integration for barber poles (see Fig. 6.7). 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.[188,125,153,143] Perceived direction changes towards the grating motion direction as the size of the indentation increases. Our model simulated such behaviour. As illustrated in Fig. 6.7 (b), 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 Fig. 6.7 (c)).

Barber pole motion stimuli with an aspect ratio of 1:1—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.[50,153] 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 and colleagues[50] reported 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 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 k_1. 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 Fig. 6.7 (d–e)). Introducing a small additive Gaussian noise (av-

erage: 0.5, variance: 0.02) into k_1 resulted in similar switches. Thus, small changes in stimulus characteristics can lead to totally different estimates of global motion in our dynamical model.

6.3 DISCUSSION

In the present study, we proposed a motion integration model to solve 2D motion integration and segmentation. We implemented and applied our two 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.

6.3.1 *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.[171,211,172] 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.[153,151,19] 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. This temporal dynamics might be related to the timing of the underlying centre-surround interactions.[14] However, the relative contribution of both lateral and feedback recurrent connectivity to the temporal dynamics of centre-surround interactions is still unclear.[8]

Next, our model can reproduce the time course of 2D motion integration as evidenced by a large number of studies at

psychophysical,[260,51,144] behavioural[154,241,35] and neurophysiological[171,211,172] levels. In brief, the estimate of global motion, as computed by our simple readout 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.

6.3.2 *Relations to other approaches*

Several other models have been designed to simulate the temporal dynamics of 2D motion integration. A first attempt was made by Wilson and colleagues to explain the transition of perceived direction between vector average and IOC solutions for type II plaids in human observers.[255] This model was further expanded to account for barber pole and line motion perception.[147] 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.

Pack, Grossberg and Mingolla[173] modelled various populations inside the cortical areas MT and MST to account for the second phase of smooth pursuit: the maintenance of ongoing pursuit movements while the image is fixed on the retina. In

our model we focus on the first stage of the pursuit focusing on motion direction and velocity estimation.

6.3.3 *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.[24,23] 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 and colleagues[250] 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.[154,241,35] However, a full model should incorporate contrast-dependant local motion filters (such as the one described by Escobar[73]) as well as contrast-dependent spatial integration mechanisms as found in areas V1[201] and MT.[175]

A second limitation of the model is its disregard for the role of other segmentation cues such as luminance, texture or disparity. In the next chapter we investigate the role of luminance smoothness as a segmentation cue for motion integration.

Chapter 7 Luminance-gated diffusion

In this chapter we propose that motion integration is gated by luminance information. In Section 7.1 we present previous proposals in cross-modal interactions which are based on junction detectors. The model is described in Section 7.2 as an extension of the approach presented in the previous chapter. The implementation details follow in Section 7.3. Section 7.4 presents the new results obtained with this model. We conclude in Section 7.5 by linking our approach to computer vision algorithms as well as other biological models, and end by discussing its biological interpretation.

MAIN CONTRIBUTION

- A luminance-gated mechanism segregating objects

7.1 INTRODUCTION

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.[106,164,98,248]

Considering the form information for motion integration is also a necessary condition to explain some experimental results, as we discussed in Section 1.3.

The role of motion diffusion process has only been investigated in a small number of models such as the one of Grossberg and colleagues where local form and motion cues are integrated through recurrent diffusion.[94,32] The various versions of their model succeed to solve the aperture problem in many different instances of motion stimuli investigated psychophysically.[143,51] However, they heavily rely on many different sub-types of local feature detectors and a huge number of cortical areas.

A similar solution was developed by Bayerl and Neumann,[22,23] albeit with a more simple and realistic motion computation al-

gorithm. 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 (the extrinsic junctions, see Section 1.3.1). Yet, such computational rules have not yet been demonstrated in the cortical processing of 2D moving patterns.

Here, we propose an extension of our dynamical model providing a simple solution for 2D motion integration combining 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 fuses both contour and surface representations, which have been found in cortical areas V1 and V2.[196,122,221] It also offers a simple solution for the edge versus surface (or global) smoothness constraints used by different models of motion integration.[248] 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.

7.2 MODEL

We extend the dynamical model developed in Chapter 6 by postulating 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.[165,80,186,121]

There are evidence for involving both lateral connections[36,216] and recurrent input[132,131] 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.

Our model is represented in Fig. 7.1. The motion integration between p_1 and p_2 which was a simple Gaussian in Chapter 6 is now gated by luminance information from a new cortical layer processing form information, ϕ. Its activity is defined by a function

$$\phi(t, x, \theta) \in \mathbb{R}^+ \times \Omega \times [0, 2\pi) \to \mathbb{R}^+. \tag{7.1}$$

and represents the local orientation of the luminance profile from position x in the direction θ. Note that function ϕ 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[138] but also can encode surface brightness.[178] In future development of the model, such function can also be extended to form information extracted from other cues such as colour or texture.

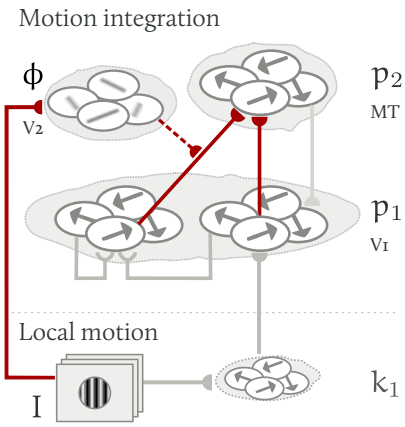


Fig. 7.1 Luminance-gated motion integration. A form layer ϕ gates the integration of information between p_1 and p_2, thus implementing anisotropic diffusion.

The integration from p_1 to p_2 is now defined by:

$$\frac{\partial p_1}{\partial t} = -\lambda_1 p_1 + (1 - p_1) \left| \lambda_1^f k_1 + \lambda^b k_1 p_2 - \lambda_1^l G_{\sigma_1^l} \overset{x}{*} \int_{\mathcal{V}} p_1(t, x, w)\, dw \right|_+, \tag{7.2}$$

$$\frac{\partial p_2}{\partial t} = -\lambda_2 p_2 + (1 - p_2) \left| \lambda_2^f \int_{\Omega} K(t, x, y)\, p_1(t, y, v)\, dy - \lambda_2^l G_{\sigma_2^l} \overset{x}{*} \int_{\mathcal{V}} p_2(t, x, w)\, dw \right|_+, \tag{7.3}$$

where K is defined by:

$$K(t, x, y) = G_{\sigma_2^f}(x - y)\, \phi(t, x, \widehat{xy}), \tag{7.4}$$

where $\widehat{xy}$ denotes the angle between the vector $\overrightarrow{xy}$ and the horizontal axis.

This luminance-gated diffusion, is the main novelty of our model. Rather than diffusing motion information isotropically from p_1 to p_2, we defined an anisotropic diffusion where p_1 is

integrated in a spatial neighbourhood using the weight $K(t, x, y)$, defined in (7.4). This weight is composed of two terms.

The first term, $G_{\sigma_2^f}(x - y)$, 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. In this chapter, we propose that the integration depends on the form so that the layer ϕ is defined by:

$$\phi(t, x, \theta) = \int_\Omega G_{\sigma_x}(x-z)\, G_{\sigma_\theta}(\theta-\widehat{xz})\, G_{\sigma_s}(I(t,x)-I(t,z))\, dz. \quad (7.5)$$

The layer ϕ describes the luminance smoothness at position x and along the direction θ. In (7.5), the term $G_{\sigma_x}(x-z)\, G_{\sigma_\theta}(\theta-\widehat{xz})$ defines an oriented spatial neighbourhood around x (see Fig. 7.2). The last term, namely $G_{\sigma_s}(I(t,x) - I(t,z))$, corresponds to a brightness similarity measure describing form information using luminance as a criterion.

A representation of the layer ϕ for all the directions and for a given set of sampled positions is shown in Fig. 7.3. The main property of ϕ is to facilitate integration inside similar spatial structures of the image, a property shared by neurons as observed in both psychophysics[204,142] and cell recordings in macaque area MT.[113]

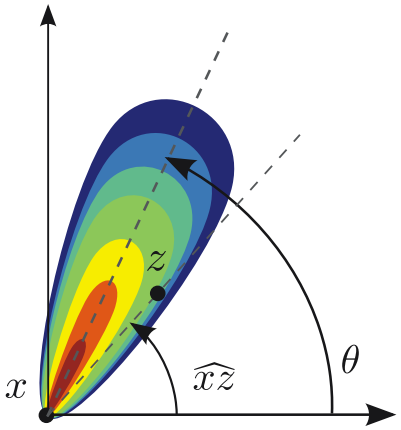


Fig. 7.2 Oriented spatial neighbourhood around x in the direction θ used to compute ϕ. Luminance in this oriented neighbourhood is compared with the luminance at the origin x.

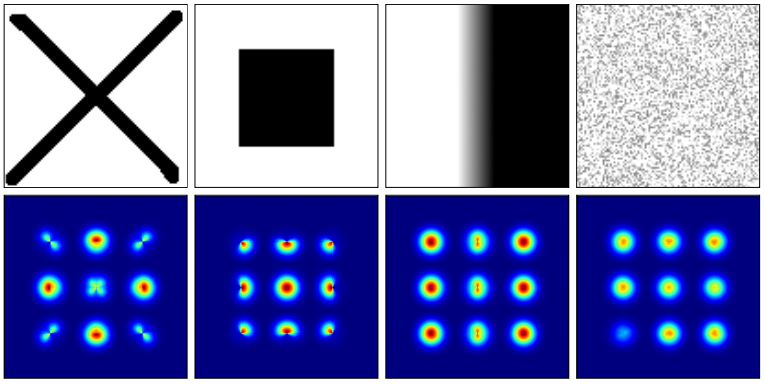


Fig. 7.3 Diffusion of information for different spatial structures. Upper row gives a set of input images with different luminance distribution. Lower row shows a representation of K indicating for a given set of sampled position, the weight by which their neighbourhood is integrated.

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 ob-

served for instance in macaque area MT.[175] 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 Fig. 7.3), but also propagates inside iso luminance regions.

7.3 IMPLEMENTATION

(7.2)	$\lambda_1 = 2.0$	$\lambda_1^f = 1.0$	$\lambda_1^b = 24.0$	$\lambda_1^l = 4.0$	$\sigma_1 = 2.0$
(7.3)	$\lambda_2 = 2.0$	$\lambda_2^f = 16.0$		$\lambda_2^l = 4.0$	$\sigma_2 = 2.0$
(7.4)	$\sigma_2^f = 8.0$				
(7.5)	$\sigma_x = 12.0$	$\sigma_\theta = \pi/8$	$\sigma_s = 0.4$		

Table 7.1 Chosen parameters setting.

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 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 sometimes limited to 60×60 binary images.[32]

7.4 RESULTS

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 chapter, 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.[113,205] 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).[204,143,206] 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.

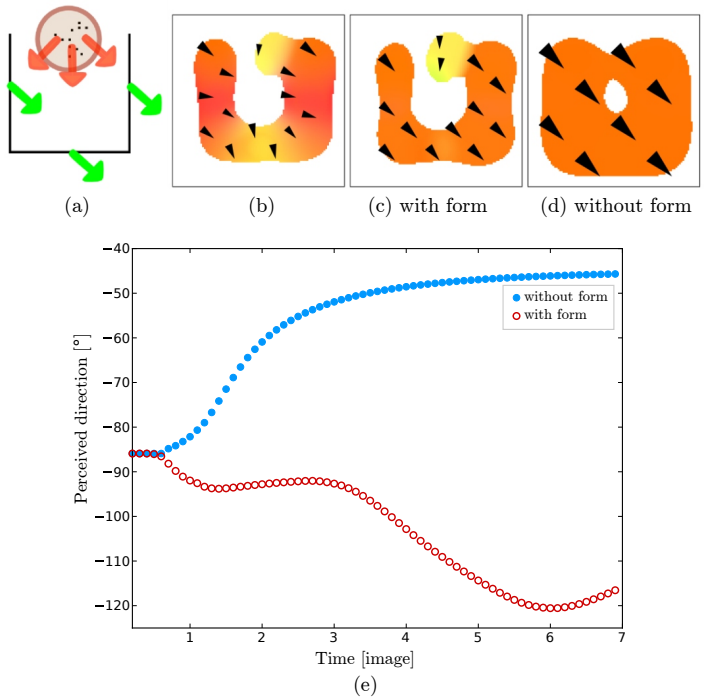


Fig. 7.4 Model response to the motion stimulus proposed byin area MT. (a) A square moving diagonally downward and to the right is presented together with a patch of moving dots instead of the upper segment of the square (see text for more details). (b) Initial model response illustrated by velocity field, 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.

7.4.1 *Preventing capture: the dotted square stimulus*

In Fig. 7.4, we considered the stimulus proposed by Huang and colleagues[113] 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 Fig. 7.4 (b)). As illustrated in Fig. 7.4 (c), 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 Fig. 7.4 (c) and Fig. 7.4 (d)).

7.4.2 *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.[9] The first stimulus consists in two horizontally translating bars, as shown in Fig. 7.5 (a). 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 Fig. 7.5 (a), we illustrate the velocity field $\mathfrak{m}_1$ 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.[9] We also show that velocity flow fields are coherent at the two different spatial scales $\mathfrak{m}_1$ and $\mathfrak{m}_2$ showing that feedback allows the model to compute coherent motion representation at different stages along the motion pathway. Removing the ϕ layer, results 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 Fig. 7.5 (b)). 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. Fig. 7.5 (b) 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 $\mathfrak{m}_2$ flow fields) but are not

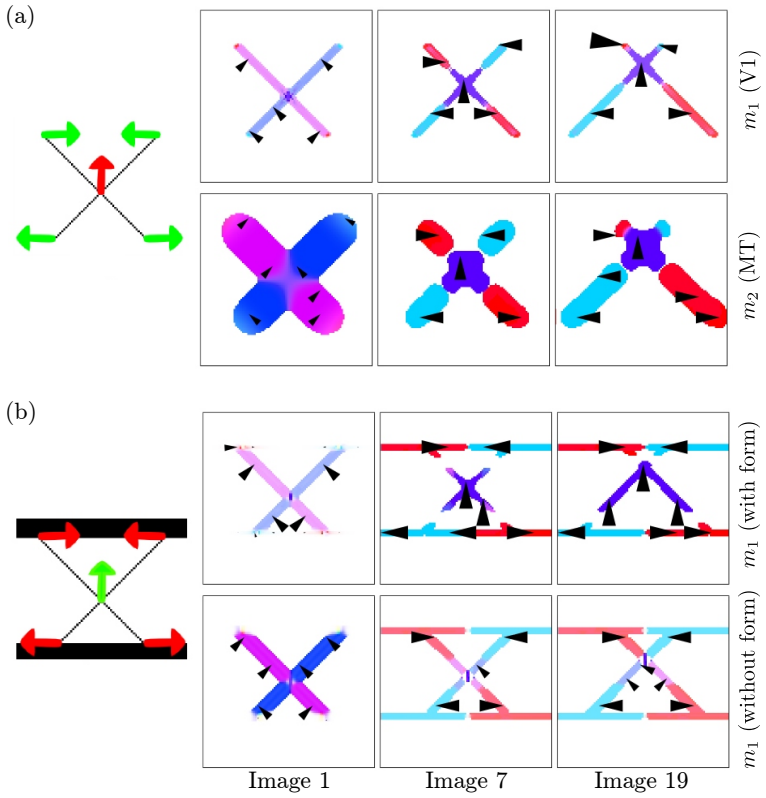


Fig. 7.5 Model response to chopsticks motion. (a) Two tilted and crossing bars are translating in opposite motion direction resulting in two horizontal perceived motions.[9] We display the velocity fields m_1 and m_2 to illustrate the time course of motion computation at two different spatial scales. (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.

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 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 occluding edges. This result demonstrates the role of the layer ϕ 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.

7.4.3 *Geometry controlled diffusion: diamonds*

Another challenging set of experiments was provided by the study of Lorenceau and Alais[142] as illustrated in Fig. 7.6. 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 counter-clockwise. The percentage of correct responses have been replotted in Fig. 7.6 (*light grey bars*) for the 10 different shapes used in this study. The authors 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 the original paper,[142] we defined a rotation coherence readout as follows. First, we decomposed local motion as given by the activity measurements of our model into a radial-rotational space. The biological plausibility of such a decomposition, as well as its links to human motion percept, have already been described by Barraza and Grzywacz.[16] 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. Fig. 7.6 plots this ratio for the

same 10 shapes (*dark 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 Fig. 7.6), most probably because of the distance between the corresponding line-endings.

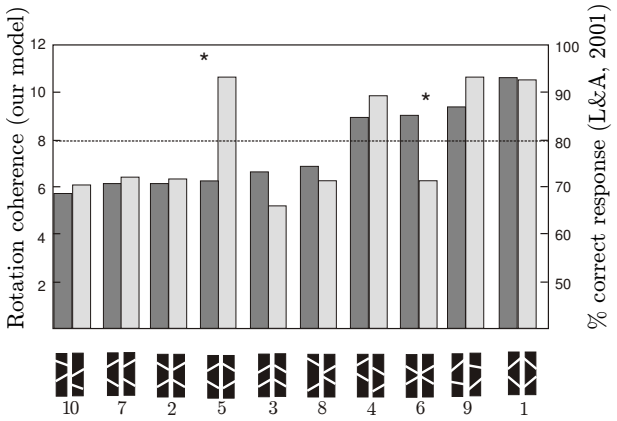


Fig. 7.6 Results on the whole set of stimuli presented by Lorenceau and Alais.[142] All the stimuli are made of four edges varying a diamond shape and are numbered according to the original paper. The *light grey* bars represents the correct rotation response from the psychophysical experiments (see rightward). The *dark grey* bars corresponds to the responses of our model (see leftward axis).

7.5 DISCUSSION

7.5.1 *Links to computer vision approaches*

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.[217,168,173] 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, on 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 and Schunck[111] 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 chapter 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.[6,259,12,246]

But there is yet another set of approaches using also form/luminance modulation for the diffusion process. For example, Hildreth[107] presented a model that calculates the velocity field of least variation along a contour in the scene, corresponding a contour smoothness constraint. Similarly, Nagel and Enkelmann[163] 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.

7.5.2 *Form modulation in biological models*

Several biologically-inspired models were designed to investigate the role of motion diffusion in the context of motion integration.[56,32,23] These models are able to capture several aspects of motion integration such as the propagation of feature tracking estimates.[56,96] Some of these models implement isotropic motion diffusion by using Gaussian distributions of activity both within layers and between layers through recurrent connectivity.[24,23] They can simulate the final results 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 mo-

tion integration that can be biased either by local form information as well as by depth cues.[32] A similar approach using depth cues was proposed recently by.[25] 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.[204,142] or the chopsticks.[9] 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 feed-forward and feedback interactions are implemented at different levels[32] and the model postulates the existence of several types of highly specific form and motion detectors. In contrast, in this chapter 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 to constrain motion diffusion. We believe that such a powerful model can then be extended to understand how cortical architectures implement more complex operations.

7.5.3 *Luminance smoothness: a simple rule for gating motion information*

We implemented a simple mechanism of form modulation 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 found in the literature.[32] Here, the layer ϕ indicates directions in the image along which luminance is nearly uniform. Such an 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.[116] Abrupt changes in the luminance profile along the contour can be signalled by another type of neurons found in area V1, *end-stopped cells*.[115,172] 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.[32] The layer ϕ used in the present study only signals in isoluminance 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 ϕ 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.[248]

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.[196,183,87] 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.[122,221] Such representation of uniform surfaces based on luminance distribution has been related to brightness perception.[196] 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.[170] On the contrary, these cells remained unresponsive to a circle of same diameter. Our model can reproduce this dynamics, thanks to the layer ϕ.

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.[86] Considerable attention has been paid to the statistics of contrast distribution and its relationships with the properties of elementary local features detectors.[209] Recent studies have shown the importance of luminance distribution as well[149,84] and pinpoint its role in the neural dynamics of local information processing[149,87] but also in surface segmentation.[82] 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[142] present the advantages to have identical motion energy. The main difference between the ten stimuli illustrated in Fig. 7.6 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 us-

ing 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 interconnected 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 collinear and aligned within visual space.[80,104,105] The collinear facilitatory effects seen for contrast detection,[186,187] static and dynamical contours detection[80] and apparent motion perception[88,203] are often through to be mediated by intra-cortical short-range lateral connections.[36,216]

In our model, we did not implemented specific features detectors, neither their particular temporal dynamics. Also, we did 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.[130] Nevertheless, our simulations 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 investigate how the precise timing of 2D motion integration can be simulated by implementing the timing architecture of the early visual pathways.[130,47]

Chapter 8 Neural fields model

Models developed in the previous chapters are expressed in a formalism sharing common notations with the literature, yet slightly different. In order to facilitate the mathematical analysis we chose the neural fields *framework described in Section 8.1 for which theoretical tools already exists. In Section 8.2 we show how a new integral term behaves asymptotically as a Laplacian operator. Section Section 8.3 and Section 8.4 demonstrate the existence, uniqueness and boundedness of the solution given by our neural field model. We conclude in Section 8.5 by analysing the selection mechanism in simplified version of the model.*

MAIN CONTRIBUTIONS

- Proofs of existence, uniqueness and boundedness of the solution of our model.
- A numerical analysis of the mechanisms involved in our models.

8.1 THE NEURAL FIELDS FORMALISM

One major difficulty in observing or modelling the brain is that its study involves a multi-scale and thus is a multi-disciplinary analysis. There is a large variety of scales but there is no integrative model yet covering all of them at once.[60] Thus, one has to choose a given scale and define a suitable mathematical framework for that scale.

In this work we consider a *neural mass* 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[254,253] and Amari,[7]

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 is still a challenging field of research.[72,61,76,238] Applications in vision of neural fields models include the work on *spatial attention* by Fix, Rougier and Alexandre,[83] on *orientation tuning* by Bressloff and colleagues[43] or Veltz and Faugeras,[238] or on *saccades generation* by Taouali, Rougier and Alexandre.[222]

Our goal is to use this formalism for the problem of motion estimation. Following the general structure of the previous models, we describe the activity of a continuum of neuron populations are denoted by:

$$p_i : (t, x, v) \in \mathbb{R}^+ \times \Omega \times \mathcal{V} \to p_i(t, x, v) \in \mathbb{R},$$

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). Thus, $p_i(t, x, v)$ is the average activity—more precisely, the *synaptic drive*[184]—of the population corresponding to position x and tuned to velocity v.

The general neural equation for an activity based model is:

$$\frac{\partial p}{\partial t}(t, r) = -\Lambda \cdot p(t, r) + S\left(\int_{\Omega \times \mathcal{V}} W(t, r, s) p(t, s)\, ds + K(t, r)\right), \tag{8.1}$$

where $p = (p_1, p_2)^T$, $r = (x, v)$ characterises the population (position and velocity tuning), $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2)$ is a matrix describing the temporal dynamics of the membrane potential, and $S(x) = (S_1(x_1), S_2(x_2))^T$ is a matrix of sigmoid functions defined by:

$$S_i(s) = 1/\left(1 + e^{-s}\right).$$

K is an external current that models external sources of excitations (in our case, $K = (\lambda_1^f k_1, 0)^T$ since there is no external input to map p_2). More importantly, $W(t, r, s)$ describes how the population s (at position y and tuned to the velocity w) influences the population r at time t.

In the right-hand side of equation (8.1), the first term de-

notes 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.

In order to fit into the neural field framework, we rewrote our dynamical model defined in Chapter 6 as follows:

$$\begin{aligned}\frac{\partial p_1}{\partial t}(t,x,v) = &-\lambda_1 p_1(t,x,v)\\ &+ S_1\Big(k_1(t,x,v)(\lambda_1^f + \lambda^b p_2(t,x,v))\\ &\quad - \lambda_1^l G_{\sigma_1^l} \overset{x}{*} \int_{\mathcal{V}} p_1(t,x,w)\,dw\\ &\quad + \lambda_1^d (G_{\sigma_1^d} \overset{x,v}{*} p_1(t,x,v) - p_1(t,x,v))\Big),\end{aligned} \tag{8.2}$$

$$\begin{aligned}\frac{\partial p_2}{\partial t}(t,x,v) = &-\lambda_2 p_2(t,x,v)\\ &+ S_2\Big(\lambda_2^f G_{\sigma_2} \overset{x}{*} p_1(t,x,v)\\ &\quad - \lambda_2^l G_{\sigma_2^l} \overset{x}{*} \int_{\mathcal{V}} p_2(t,x,w)\,dw\\ &\quad + \lambda_2^d (G_{\sigma_2^d} \overset{x,v}{*} p_2(t,x,v) - p_2(t,x,v))\Big),\end{aligned} \tag{8.3}$$

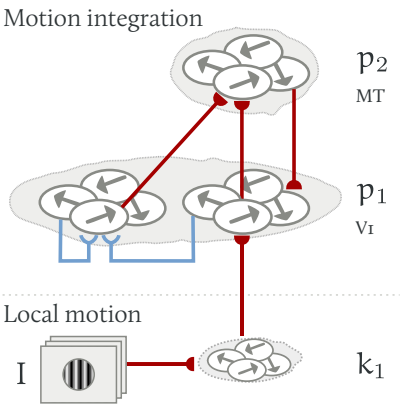


Fig. 8.1 Schematic view of the neural fields model. It follows an architecture similar to the one defined in Chapter 6.

The major change is the use of a sigmoid instead of the modulated positive rectification in previous models (6.1)–(6.2) and (7.2)–(7.3). In addition, we also added a lateral diffusion term in both equations, weighted by $\lambda_{1,2}^d$. Such a diffusion term ensures a smoother solution and its effect is studied in Chapter 10.

Note that the system defined in (8.2)–(8.3) does not completely fit into the classical neural field framework as defined by (8.1). Indeed, our coefficient matrix W depends on p since we model a multiplicative feedback – $k_1(t,x,v)\,p_2(t,x,v)$.

8.2 LATERAL DIFFUSION

Several models implemented gap junctions as lateral diffusion through a Laplacian operator. In this section, we prove that

such a Laplacian operator can be approximated by the diffusion terms introduced in (8.2)–(8.3).

LEMMA 8.1 Given a Gaussian kernel G_σ as defined by

$$A e^{-\left(\frac{(x-x_0)^2}{2\sigma_x^2}+\frac{(y-y_0)^2}{2\sigma_y^2}\right)},$$

with $\sigma = (\rho, \eta)$, let us denote

$$A = G_\sigma \overset{x,v}{*} p(x,v) - p(x,v), \tag{8.4}$$

then we have

$$A = \frac{\rho^2}{2}\sqrt{\pi}D_x^2 p(x,v) + \frac{\eta^2}{2}\sqrt{\pi}D_v^2 p(x,v) + o(\rho^2, \eta^2, \rho\eta). \tag{8.5}$$

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 (8.4) 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

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

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

thanks to the moment conditions,

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

and we finally obtain (8.5). □

8.3 EXISTENCE AND UNIQUENESS OF THE SOLUTION

In order to study the well-posedness of our model (8.2)–(8.3), let us consider the results presented in,[75] for neural field equations. 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 (8.2 and 8.3). 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 8.2)

Let $\mathcal{F}$ be the set $L^2(\Omega \times \mathcal{V})$ of square integrable functions defined on the product set $\Omega \times \mathcal{V}$ and taking their values in $\mathbb{R}$, and $\mathcal{F}^2 = \mathcal{F} \times \mathcal{F}$. The basic idea is to rewrite (8.2)–(8.3) as a differential equation defined on the set $\mathcal{F}^2$. With a slight abuse of notation we can write $p_i(t)(x,v) = p_i(t,x,v)$ and note $p : \mathbb{R} \to \mathcal{F}^2$ the function defined by the following Cauchy problem:

$$p(0) = p_0 \quad \in \mathcal{F}^2, \tag{8.6}$$
$$p' = -\Lambda p + S\left(W(t) \cdot p + K(t)\right), \tag{8.7}$$

with $p = (p_1, p_2)^T$, $K = (\lambda_1^f k_1, 0)^T$, and $S(x_1, x_2) = (S_1(x_1), S_2(x_2))$. The operator W is the 2×2 connectivity matrix function defined

by the four linear mappings from $\mathcal{F}$ to $\mathcal{F}$:

$$
\begin{aligned}
W_{11} \cdot p &= -\lambda_1^l G_{\sigma_1^l} \overset{x,v}{*} p + \lambda_1^d (G_{\sigma_1^d} \overset{x,v}{*} p + p), \\
W_{12} \cdot p &= \lambda^b k_1 p, \\
W_{21} \cdot p &= \lambda_2^f G_{\sigma_2} \delta_v \overset{x}{*} p, \\
W_{22} \cdot p &= -\lambda_2^l G_{\sigma_2^l} \overset{x,v}{*} p + \lambda_2^d (G_{\sigma_2^d} \overset{x,v}{*} p + p).
\end{aligned}
$$

Functionally W_{11} and W_{22} correspond to lateral interactions in maps VI and MT, W_{12} denotes the backward connection from MT to VI, and W_{21} denotes the forward integration from VI to MT. In the following we note f the mapping defined by the right-hand side of (8.7):

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

Note that the time dependence in the definition of f arises solely from the function k_1 that occurs in W_{12} and in K. We prove the existence and uniqueness of a solution to (8.7) 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 (8.7), and that its interval of definition is an open interval $(-\alpha, \alpha)$ containing 0.

LEMMA 8.2 If $k_1(t)$ is measurable for all $t \in I$, f maps $I \times \mathcal{F}^2$ to $\mathcal{F}^2$.

Proof. Let $p = (p_1, p_2) \in \mathcal{F}^2$. If $k_1(t)$ is measurable for all $t \in I$, so is $W_{12} \cdot p_2$. All the other elements of $W \cdot p$ are simple or weighted sums (convolutions) of a measurable function p and thus $W \cdot p$ is measurable. This implies that $S(W(t) \cdot p(t) + K(t))$ is in $\mathcal{F}^2$ for all $t \in I$. □

LEMMA 8.3 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

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

where S' is the maximum value taken by the derivatives of the sigmoids S_1 and S_2. $\|W(t)\cdot(p-q)\|$ 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:

$$\|G\overset{x,\nu}{*}p\|_{\mathcal{F}}\leqslant k\|p\|_{\mathcal{F}}\quad\forall p\in\mathcal{F},$$

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 $\|k_1(t)p_2\|\leqslant\bar{k}_1\|p_2\|_{\mathcal{F}}$. This completes the proof that f is Lipschitz continuous with respect to the second variable. □

THEOREM 8.1 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 (8.6–8.7) has a unique solution, hence is in $\mathcal{C}^1(J,\mathcal{F}^2)$.

Proof. Thanks to Lemmas 8.2 and 8.3 the conditions of the Cauchy-Lipschitz Theorem are satisfied. □

8.4 BOUNDEDNESS OF THE SOLUTION

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

PROPOSITION 8.1 The solution described in Theorem 8.1 is bounded for all $t\in J$.

Proof. The variation of constant formula yields:

$$p(t)=e^{-\Lambda t}p_0(t)+\int_0^t e^{-\Lambda(t-s)}\,S(W(s)\cdot p(s)+K(s))\,ds,$$

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

$$\begin{aligned}\|p(t)\| &\leqslant \|e^{-\Lambda t}\|\|p_0\| + \\ &\qquad \left\| \int_0^t e^{-\Lambda(t-s)}\, S(W(s)\cdot p(s) + K(s))\, ds \right\| \\ &\leqslant e^{\max(\lambda_1,\lambda_2)\alpha}\|p_0\| + \max\left(\frac{S_{1m}}{\lambda_1}, \frac{S_{2m}}{\lambda_2}\right)\left(e^{\max(\lambda_1,\lambda_2)\alpha} - 1\right) \\ &\leqslant e^{\max(\lambda_1,\lambda_2)\alpha}\left(\|p_0\| + \max\left(\frac{S_{1m}}{\lambda_1}, \frac{S_{2m}}{\lambda_2}\right)\right),\end{aligned}$$

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

As Faugeras and colleagues,[75] we can extend this local result from $(-\alpha, +\alpha)$ to $(-\alpha, +\infty)$, assuming that the hypotheses on p_0 in Theorem 8.1 are satisfied for $t \in (-\alpha, +\infty)$. Indeed, either $+\alpha = +\infty$ and the result is proved or there exists $0 < \beta < \alpha$ such that p is not bounded for all $\beta \leqslant t < \alpha$, thereby obtaining a contradiction.

We summarise these results in the following theorem:

THEOREM 8.2 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 (8.6–8.7) has a unique bounded solution, hence in $\mathcal{C}^1((-\alpha, +\infty), \mathcal{F}^2)$.

8.5 SELECTION MECHANISM ANALYSIS

8.5.1 *Local model with direction*

To study the effect of the lateral inhibition on the input, we can restrict our model to a single population model, not considering space, and with a constant input. Thus the input is given by $I(\nu) : \mathcal{V} \to \mathbb{R}$, and the model is defined as:

$$\frac{\partial p}{\partial t}(t,\nu) = -\frac{1}{\tau}p(t,\nu) + S\left(\int_{\mathcal{V}} -k\, p(t,y)\, dy + I(\nu)\right), \quad (8.8)$$

where S is a sigmoid and τ and k are positive constants.

LEMMA 8.4 If $\sup S'|\mathcal{V}|\,|k|\,\tau < 1$, then the solution p_e of the model described in (8.8) is given by:

$$p_e(\nu) = \tau S\left(I(\nu) - k\tau\bar{\alpha}\right).$$

where $\bar{\alpha}$ is constant across $\mathcal{V}$.

Proof. At the equilibrium, we can set the left part of (8.8) to zero. Then the steady-state p_e is defined by:

$$\frac{1}{\tau} p_e(v) = S\left(\int_{\mathcal{V}} -k p_e(y)\, dy + I(v)\right).$$

Switching from the activity p to the voltage u formulation, by $p = \tau S(u)$, the steady-state voltage u_e is defined by:

$$\begin{aligned} u(v) &= \int_{\mathcal{V}} -k\tau S(u_e(y))\, dy + I(v) \\ &= -k\tau \int_{\mathcal{V}} S(u_e(y))\, dy + I(v) \\ &= -k\tau\alpha + I(v). \end{aligned} \tag{8.9}$$

where we α is defined as:

$$\alpha = \int_{\mathcal{V}} S(u_e(y))\, dy. \tag{8.10}$$

Then, using (8.9), we have:

$$\int_{\mathcal{V}} S(u_e(v))\, dv = \int_{\mathcal{V}} S(-k\tau\alpha + I(v))\, dv,$$

that is, with the definition (8.10):

$$\alpha = \int_{\mathcal{V}} S(-k\tau\alpha + I(v)\, dv. \tag{8.11}$$

Denoting by $F(I)$ the right hand side of (8.11), this function is a contraction mapping if $\sup S'|\mathcal{V}|\,|k|\,\tau < 1$. Then, under this assumption and thanks to the fixed point theorem, there exists a unique $\bar{\alpha}$ verifying (8.11) that depends on model parameters and I.

Switching back to an activity model we have:

$$\begin{aligned} p_e(v) &= \tau S(u_e(v)) \\ &= \tau S\left(-k\tau\bar{\alpha} + I(v)\right) \end{aligned}$$

which concludes the proof in the activity domain. □

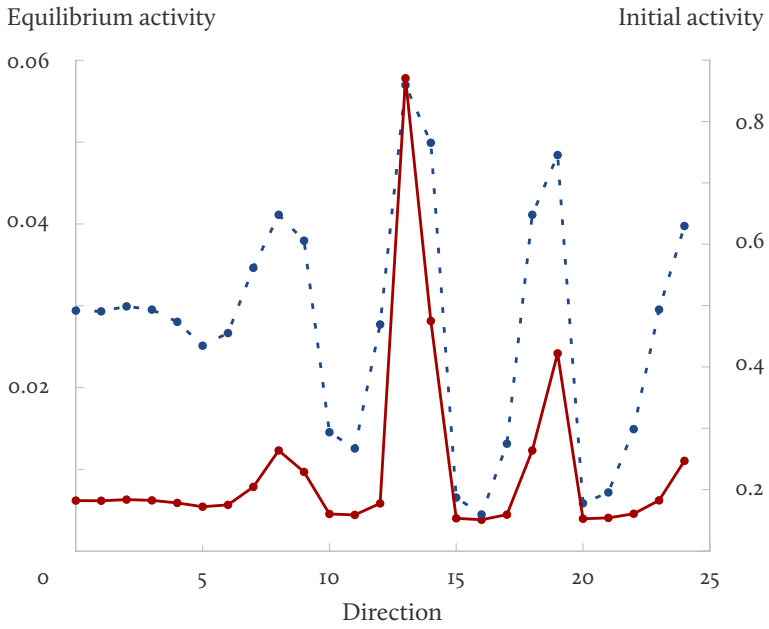


Fig. 8.2 Selection mechanism on direction model. *Dotted* line represent the initial input $I(v)$. *Solid* line show the equilibrium state in which multiple velocities are still active.

Contrarily to divisive inhibition implemented for instance by Yuille and Grzywacz,[262] our lateral inhibition is not sufficient to implement a *winner-take-all* mechanism (see Fig. 8.2). Indeed, the role of the lateral inhibition here is only to subtract a constant value at each direction, since $\bar{\alpha}$ does not depend on v (Lemma 8.4).

Yet numerical results in our complete neural-field model suggest a strong selection mechanism influenced by lateral inhibition. We hypothesise that selection is thus the combination of lateral inhibition and spatial integration. Indeed, in the next section we add a 1D spatial domain to the model to obtain a direction selection.

8.5.2 *1D spatial model with direction*

To account for the selection mechanism obtained in our complete model, we add a 1D spatial dimension to the previous purely directional model. The 1D spatial model has its activity defined by:

$$\frac{\partial p}{\partial t}(t,x,v) = -\frac{1}{\tau}p(t,x,v) + S\left(G_\sigma \overset{x}{*} p(t,x,v) - \int_v k\, p(t,x,y)\,dy + I(x,v)\right), \tag{8.12}$$

where $p : \mathbb{R}^+ \times \Omega \times \mathcal{V} \to \mathbb{R}$, and Ω is a 1D open bounded set of $\mathbb{R}$.

The two sets of connectivities defined in this model are enough to reproduce various properties of the full neural field model. At the same time the model is small enough to apply *bifurcation* analysis tools on it. The analysis was conducted using a C++ implementation of the model and the Trilinos library.[103]

As an illustration, we studied the multi-stability phenomenon described in Section 6.2.3 and redrawn in Fig. 8.3. Thus we defined an input constant in time mimicking the aperture problem found in a 1D cut of the square gratings stimulus (*red line* in Fig. 8.4).

We analysed the bifurcation by varying the steepness of the sigmoid (see Fig. 8.5). For sigmoids steep enough we observe several branches stable and unstable of solutions. The two solutions shown in Fig. 8.3 are observed and correspond to the two stable branches of the right figure (they are overlapping on the left figure). In addition, an unstable solution consisting of an average of the two stable solutions is found by the analysis.

Fig. 8.3 Response of model (6.1)–(6.2) on almost similar stimuli. The output direction changes dramatically by varying the input conditions. The *red line* denotes the cut used as input for the model presented here.

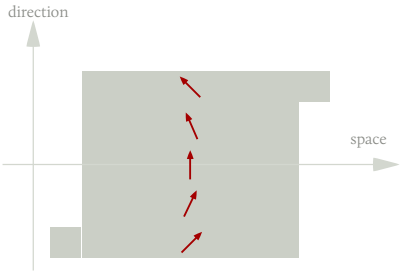


Fig. 8.4 Input stimulus for the direction+space model. It reproduces the aperture problem in an horizontal 1D slice of the square barber pole.

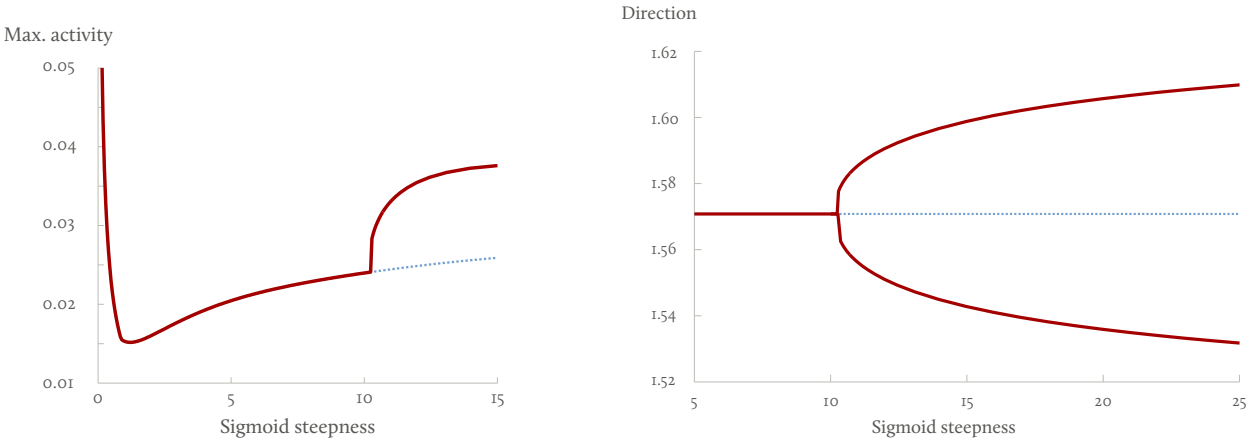


Fig. 8.5 Bifurcation analysis on the 1D spatial model with direction (8.12). (a) The maximum activity is shown with respect to the sigmoid steepness. Since they have the same maximum value, two stable solutions are overlapping after the pitch fork. (b) The average direction is shown with respect to the sigmoid steepness. *Dotted* lines denotes unstable solutions.

Part III

Performance analysis

Part 3 Performance analysis

Generalising the evaluation of motion integration models discussed in the previous results, we advance a new evaluation methodology of motion estimation approaches. Our benchmark described in Chapter 9 is divided in two. A static evaluation allows comparison of the final percept and can be applied to existing computer vision approaches. A dynamical evaluation considers the time evolution of the perception or of the eye position. We provide baseline results for state of of the art computer vision and modelling approaches.

To conclude we investigate the performance of our neural field model in a computer vision fashion. In Chapter 10 we start by inputting videos from classical and state of the art computer vision benchmarks into our model in its neural field formulation. The results demonstrate that the proposed approach can successfully be applied to computer vision sequences and that biological inspiration can be an effective source of innovation.

Chapter 9 Comparison to human performance

Evaluating models of motion integration is not easy and no evaluation methodology currently exists. We propose to apply computer vision ideas related to benchmarking as the basis of an evaluation based on performance of the visual system (Section 9.1). We start by describing the main difficulties of designing a benchmark in Section 9.2. The stimuli database is made of stimuli used in behavioural and perceptual experiments described in the previous chapters. We specify it in Section 9.3. We illustrate the scoring procedure on an example and show baseline results obtained on both biologically inspired artificial vision models and computer vision models in Section 9.4. Section 9.5 concludes and mentions possible extensions of this work.

MAIN CONTRIBUTIONS

- An evaluation methodology for models of motion integration.
- Baseline results on classical and state of the art approaches in computer vision and modelling.

9.1 INTRODUCTION

Offering proper evaluation methodology is essential to continue progress in modelling the neural mechanisms involved in vision information processing.

In this chapter we define a motion evaluation methodology where the visual system performance acts as ground truth. Since models aim at elucidating both the computational principles and the computing architectures involved in motion processing, comparing their outputs to biological responses is therefore a strong requirement. Such an evaluation methodology is

very different from classical computer vision benchmarks where flow fields are compared together. In our 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. Thus defining global readouts is necessary in order to compare output from models with observable quantities measured in neuroscience experiments. Moreover behavioural and perceptual experiments provide numerous types of data such as perceived motion direction and speed or smooth pursuit eye movements.

▸ Stimuli, scoring procedure and baseline results are available online at: http://www-sop.inria.fr/neuromathcomp/motionpsychobench

9.2 COMPARISON DIFFICULTIES

9.2.1 *Stimulus parametrisation*

Comparing models performance to biological data requires the definition of an homogeneous stimuli set. For example the stimuli are characterised by their physical size, the distance to the observer, and their visual field size. In a benchmark stimuli set we need to ensure a constant mapping between the physical and the numerical dimensions, otherwise incoherences will occur in the simulated results. Other physical quantities such as duration and luminance also require precise mappings.

9.2.2 *Discretisation*

Stimuli discretisation implies the necessity to define properly a scale factor for converting real-world values characteristics into computer parameters. The main problem with discretisation procedures is the *aliasing problem*. One has to make sure that frequency of the input does not cross the Nyquist frequency.

The spatio-temporal discretisation maps the time and the visual field as a succession of discrete images sampled uniformly. Such a representation is geometrically different from the log-polar retinotopic disposition of the visual cortex. Moreover the precision of the input has to be sufficient to avoid aliasing problems with the stimuli sizes used by the experimentalists. The finer the precision the larger the data and a good compromise between those two quantities has to be chosen.

Luminance is usually encoded by an eight bits values at each

pixel. Such a coarse quantisation is a severe restriction since contrast has a profound impact of the temporal dynamics of most visual percepts and is responsible for many dynamical non-linearities.

9.2.3 *Inhomogeneity of the motion representations*

All motion models do not have the same motion representation. Their output can be described by global velocity likelihoods,[157] velocity distributions at every position,[95,24] filter responses,[1] time-correlated spike trains,[53] or 2D flow fields.[17] A typical biological model of motion integration might include V1 layers with filter-like responses, MT layers corresponding to local pattern translational motions and MST layers giving indications of global rotation or expansion motions. Because of the variety of motion representations it is necessary to define common observable quantities which are comparable to experimental measurements. These common observable quantities are called *readouts* in both simulations and experiments presented herein.

9.2.4 *Lack of ground truth*

In computer vision the ground truth is the true velocity field, which is easily defined for synthetic videos, and which can also be estimated for real scene videos. For example in a recent benchmark Baker and colleagues[15] proposed videos of real scenes with the true velocity field. Algorithms can be evaluated based on local comparisons of the 2D flow fields against the estimated flow resulting from different algorithms.

In psychophysical studies the notion of ground truth is less obvious and it is impossible to define it 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. Moreover many stimuli are bi-stable or multi-stable, and an additional difficulty in defining a ground truth.

Among the set of experimental stimuli studied in neuroscience some provided results at different levels. For instance, the coherence level necessary for perceiving global motion in a

random dot patterns has been measured in human subjects but also in single neurons in areas MT, MST, and LIP. In the 2D motion integration example being considered here, a consistent set of global direction estimates have been collected at these different levels as well as for human perception, and monkey and human smooth pursuit.[152] When available these datasets collected for different responses with a single set of motion stimuli should be used to benchmark models.

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 restrict our study to a set of fundamental questions described in the following section.

9.3 DATABASE DESIGN

In this chapter we focus on two fundamental aspects of motion integration. Namely, we want to evaluate models performance with stimuli showing the respective influence between 1D versus 2D cues, and the dynamics of motion integration. We chose four stimuli fitting into two classes: line-drawings objects and gratings. For the purpose of our evaluation we selected stimuli for which smooth pursuit eye movements and motion perception data were available.

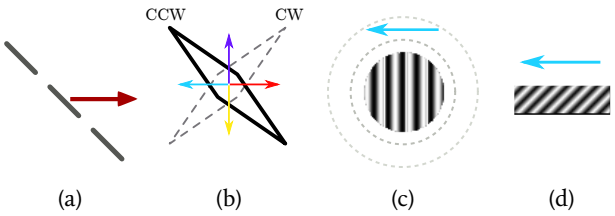


Fig. 9.1 Database design. The proposed stimuli fit into two classes: line drawings and gratings. (a) Translating bar. (b) Translating diamond. (c) Grating size. (d) Barber pole.

In the translating bar experiment (Fig. 9.1 (a)) a bar is cut in an increasing number of segments, thus increasing the number of 2D cues.[144] This stimulus is described in Section 1.2.2.

In the translating diamond experiment (Fig. 9.1 (b)) a diamond is presented moving either vertically or horizontally and with two possible orientations.[154] This stimulus is described in Section 2.2.2.

In the grating size experiment (Fig. 9.1 (c)) a translating grating is viewed through different aperture sizes.[20] This stimulus is described in Section 2.3.

In the barber pole experiment (Fig. 9.1 (d)) a translating grating is viewed through a rectangular aperture.[243] This stimulus is described in Section 1.2.2.

9.4 RESULTS

9.4.1 *Readouts definition*

Common output is necessary in order to compare models together and abstract all implementation dependent issues coming from the large variety of motion models. For example, if eye-movement like output can be defined, then models can be compared together in term of dynamics. As illustrated by this example, our goal here is to propose output formats corresponding to classical readouts as defined in psychophysics. Our goal is to describe for each readout, what they are supposed to measure, and how they are measured in psychophysics.

Since the notion of local measurement has no clear interpretation in term of neural architecture or activity, the readouts defined herein correspond to global motion estimations. For example, it is known that the preferred motion of neurons in both V1 and MT changes depending on the stimulus,[175] or that the perceived motion and the neural activity can differ.[114] Considering not only the local estimates but also more distant features makes an important difference with classical computer vision methodology for optical flow where only precision of local estimates matters.

Our goal is to define qualitatively which common outputs are needed from models (i.e. readout inspired from real neuroscience experiments). However no general formula for readouts can be given for three main reasons. The first reason is that the cortical mechanisms leading to a readout from neural activity are usually not clearly established and a fortiori it is hard to model them rigorously. The second reason is that the variety of motion representations in models makes it impossible to write a general formula that would be valid for any kind of

representation. The last reason is that readouts defined here are sometimes inherent to models. Indeed some models already provide an eye-movement output[157] whereas others considers neural activity in cortical areas.[23]

In this chapter two kinds of evaluation are considered: The *static* evaluation considers only the result at convergence whereas the *dynamic* evaluation focuses on the dynamics of motion integration.

- For the *dynamic evaluation*, the readout is expressed as a time independent value, such as a perceived motion direction. Given a stimulus, some experiments require subjects to give their perception concerning the motion they perceive. In general, this perceived motion readout ignores time evolution. From a modelling point of view, the perceived motion readout can be a global velocity corresponding to the steady state. For the 2D motion integration tasks, we can assume that perceived motion corresponds to the final output from eye movement readout.

- For the *static evaluation*, the readout is expressed as a time dependant value, such as smooth pursuit eye movements. Voluntary eye movement to track motion are directly related to our interpretation of the scene in term of motion content. Primates use two types of voluntary eye movement to track objects of interest: smooth pursuit and saccades. Pursuit eye movements are driven by visual motion and rely on both low-level and high-level motion processing. Pursuit initiation is critically dependent upon visual motion processing in cortical areas MT and MST(see Chapter 2. It presents the interest of being a simple motor responses that requires an accurate estimate of the global direction and speed of a single object, despite its properties such as shape or colour. It is therefore a good probe of object motion processing and in particular it reflects many of the dynamical properties of low level motion computation. From a modelling point of view, smooth pursuit eye movement is a single time-dependent vector, and we only consider the eye direction since speed is generally ignored in experiments.

9.4.2 *Scoring procedure*

The full scoring procedure for each class of stimuli is available online. For each stimulus, instructions are detailed (see for example Fig. 9.2 for translating diamonds stimuli and Fig. 9.3 for the grating sizes). In order to show the general idea, let us explain what is the scoring procedure for static evaluation of the translating diamonds stimuli. Our reference paper for this case is the one of Masson and Stone[154] as it presents ocular following measurements that we can use in our evaluation.

For a given an approach, our evaluation procedure starts from the estimated global motion direction at every frame and for each of the stimuli in this class, i.e. for translating diamonds translating in one out of the four possible directions (right, up, left, down) and oriented either clockwise (CW) or counter clockwise (CCW) . Let us denote by $e_S(t)$ the estimated global direction dynamics for a stimulus S, with $t \in [0, 450\,\text{ms}]$ and $S \in \{\text{right,up,left,down}\}\times\{\text{CW,CCW}\}$. From this global estimated direction, the instantaneous direction error $\varepsilon_S(t)$ is defined by

$$\varepsilon_S(t) = e_S(t) - \hat{e}_S(t),$$

where $\hat{e}_S(t)$ is the true object motion. Some results are shown in Fig. 9.4 and Fig. 9.5 for the biologically inspired artificial vision model proposed in Chapter 6. Here the estimated global motion direction was obtained from the MT layer activity.

Translating diamonds: static

Masson, G.S. and Stone, L.S.
J Neurophysiol, 2001.

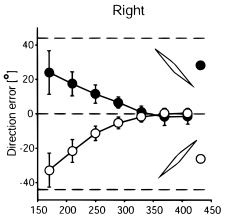
1. For each stimulus, compute global direction

a. The static global motion direction is computed for each stimulus

$$d_s \in [0, 2\pi[$$

$$\text{with } t \in [0, 450],\ s \in \{\text{up}, \text{down}, \text{left}, \text{right}, \} \times \{\text{CW}, \text{CCW}\}$$

2. For each stimulus, we compute a direction error



a. We know the true object motion direction e_s for each stimulus.

b. We can compute the difference between the two

$$e_s = \hat{d}_s - d_s$$

3. We score the direction errors

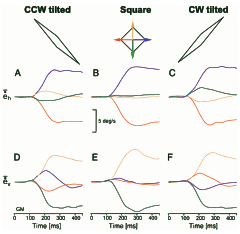
a. Assuming a normal law with data from the literature, where $\hat{e}_s$ is the average direction error, and σ_s its deviation:

$$k_s = G_{\sigma_s}(e_s - \hat{e}_s)$$

Translating diamonds: dynamic

Masson, G.S. and Stone, L.S.
J Neurophysiol, 2001.

1. For each stimulus, extract eye direction error



a. The eye direction dynamics is denoted for each stimulus s by:

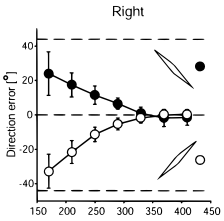
$$e_s(t) \in [0, 2\pi)$$

$$\text{with } t \in [0, 450],\ s \in \{\text{up}, \text{down}, \text{left}, \text{right}, \} \times \{\text{CW}, \text{CCW}\}$$

b. From the veridical object motion, $\hat{e}_s$, we compute a direction error ϵ_s:

$$\epsilon_s(t) = \hat{e}_s(t) - e_s(t)$$

2. For each of the motion direction, we fit the error



a. For a given motion direction, we average absolute errors for the clockwise and counter clockwise orientations.

$$h_d(t) = 0.5 \,\text{abs}\,(e_{d,\text{CW}}(t) - e_{d,\text{CCW}}(t))$$

b. We fit the result with the following function

$$f_{A,B,\tau}(t) = A \exp\left(-\frac{t}{\tau}\right) + B$$

3. We score the fitted parameters

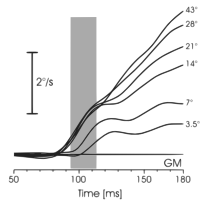
Assuming a normal law with data from the literature.

Fig. 9.2 Slides describing the static and dynamics scoring procedure for the translating diamonds (available on the website).

Grating size: dynamic

Barthélemy, F., Vanzetta, I. Masson, G.S., J Neurophysiol, 2006.

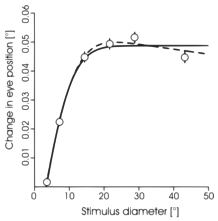
1. For each grating size, extract eye velocity



The eye velocity dynamics is denoted for each size s by:

$$v_s(t) \text{ with } s \in \{3.5, 7, 14, 21, 28, 43\}, t \in [0, 180]$$

2. For each time window, fit change change wrt size



a. We consider 20 ms time windows centered on:

$$t_i \in \{105, 125, 145, 165\}$$

b. We average v(t) on the time windows to get changes in eye position.

c. For each time window i, we fit the change in eye position w.r.t. grating size by:

$$f(s) = A_e \text{erf}\left(\frac{s}{\omega}\right) + R_0$$

3. We score the values for ω

Assuming a normal law with data from the literature.

Fig. 9.3 Slides describing the dynamics scoring procedure for the grating sizes (available on the website).

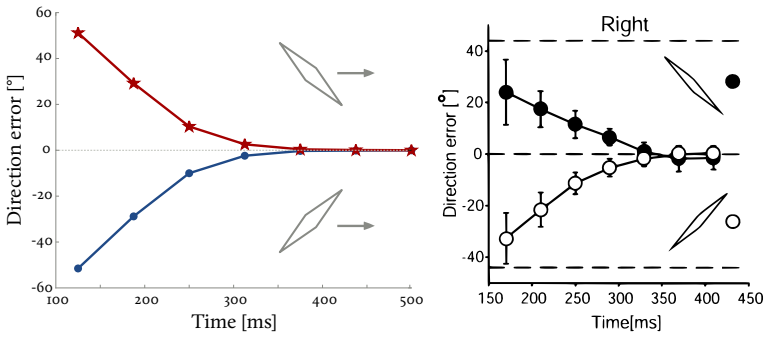


Fig. 9.4 (a) Direction error for rightward moving diamonds estimated by the model of Chapter 10. (b) Oculomotor dynamics recorded by Masson and Stone.[154]

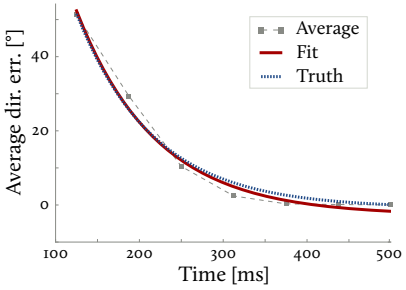


Fig. 9.5 Average directional error for all moving direction and its fit to the associated function compared to the observed fitted function.

Our goal is to compare this estimated direction error $\varepsilon_S(t)$ to the direction error observed with human subjects (see Fig. 9.4 (b) from Masson and Stone[154]). We followed the same procedure as the one defined by the authors. Estimated direction error is fitted with the function:

$$f_{\alpha,\beta,\tau}(t) = \alpha \exp(-t/\tau) + \beta,$$

where α, β, and τ are the fitting parameters to be adjusted. These parameters can then be compared to what is obtained with human observers. The original authors estimated the average values and standard deviations for these parameters – denoted respectively by $\mu(\hat{\eta})$ and $\sigma(\hat{\eta})$ for a parameter $\eta \in \{\alpha, \beta, \tau\}$). Assuming a Gaussian distribution for parameters coming from human subjects and given a stimulus S, we defined a score $s \in [0, 1]$ for each parameter by:

$$s_\eta = \exp(-(\eta - \mu(\hat{\eta}))^2/\sigma(\hat{\eta})^2), \qquad \text{with } \eta \in \{\alpha, \beta, \tau\}.$$

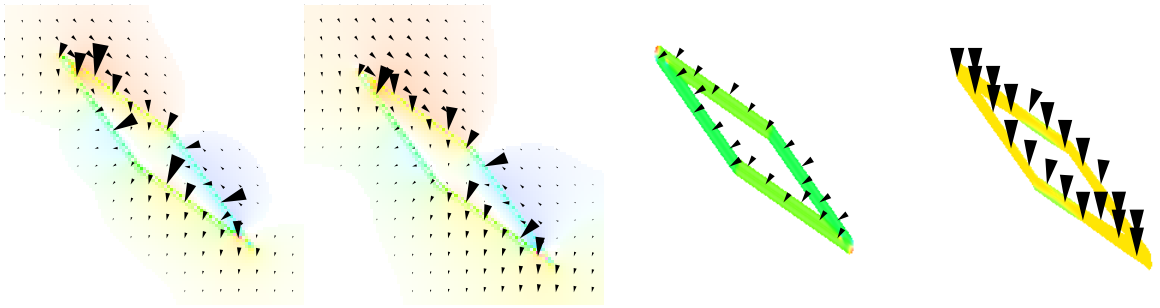
Finally, a global score can be obtained by averaging scores over all stimuli and parameters.

9.4.3 *Baseline results*

We applied our evaluation methodology to both biologically inspired artificial vision models[23,227] and computer vision models[111,146,220,45] by running either the original implementation from the authors or the code that was available in the Opencv library.[42] A single set of parameters were experimentally tuned in order to achieve the overall best score across all experiments. As defined in Section 9.4.1, we discuss below the results obtained in the static and dynamic evaluations. For each scenario, results are presented into tables by scores between zero (low performance) and one (high performance. Algorithms are ranked according to their average score across all experiments. Complete details for each stimuli and evaluation procedure can be found on the associated website.

Static evaluation

Results are presented in Table 9.1. As a general comment, it is interesting to remark that models performance somewhat follows research evolution. For example the classical approaches for optical flow proposed by Horn and Schunck[111] or Lucas and Kanade[146] show quite a poor performance on most stimuli (see Fig. 9.6).



(a) (b) (c) (d)

Fig. 9.6 (a) and (b) Output flow from Horn and Schunck[111] for two different frames. (c) and (d) Same output frames from the model described in Chapter 10.

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, with hardly no texture, the input to differential algorithms is not very informative and leads to an aperture problem that is hard to solve numerically.

Approach	Avg.				
TMK·10	1.00	1.00	1.00	1.00	1.00
SRDB·10[220]	0.86	1.00	1.00	0.65	0.78
BM·10[45]	0.74	1.00	1.00	0.00	0.98
BN·04[24]	0.68	1.00	1.00	0.36	0.38
LK·81[146]	0.45	0.81	0.00	0.99	0.00
HS·81[111]	0.39	0.52	0.00	1.00	0.03
BMOCV [42]	0.19	0.00	0.32	0.44	0.00

Table 9.1 Static evaluation results. For each approach and each experiment a score between 0 (worse) and 1 (best) is given depending on the final motion direction error. BMOCV denotes the block matching algorithm found in the OpenCV library, and TMK·10 is the model of Chapter 10.

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 Sun and colleagues[220], or Brox and Malik,[45] which are now among the best computer vision models (see the latest results online from

Baker and colleagues [15]). Interestingly, those models also perform well for most of our experiments.

Biologically inspired artificial vision models (Bayerl and Neumann[24] or model of Chapter 10) show high performance. In particular the model proposed in Chapter 10 obtains the maximum score. The major strength of this 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 result of this evaluation methodology because it allows to show that taking biology into account can lead to extra performance.

Dynamic evaluation

Results are presented in Table 9.2. Studying the dynamical properties of motion integration is quite a new topic and very few biological data (psychophysical, oculo-motor, or neural dynamics) were available when most models were proposed. For this reason, most models were often static, i.e. they were interested in estimating the final percept or optical flow, ignoring how the solution evolves in time to the final percept.

Table 9.2 Dynamic evaluation results. For each approach and each experiment a score between 0 (worse) and 1 (best) is given depending on the fitting procedure described in the main text.

Approach	Avg.			
TMK·10	0.68	1.00	0.96	0.08
LK·81[146]	0.37	0.75	0.36	0.00
SRDB·10[220]	0.37	0.75	0.36	0.00
BM·10[45]	0.35	0.50	0.36	0.18
BN·04[24]	0.32	0.50	0.39	0.07
BMOVC [42]	0.31	0.50	0.36	0.05
HS·81[111]	0.26	0.75	0.03	0.00

Studying the dynamics has mainly been considered in Chapter 6. This model shows the best performance on translating bars and diamonds, considering the recent neuronal, psychophysical, and behavioural findings.[171,172,154,241] For all other models, since there is no true dynamics, scores are not very informative. Yet, dynamics remain an open issue in the modelling community. None of models tested here performed correctly on the gratings stimulus, suggesting that the underlying mechanisms remain to be

found.

9.5 DISCUSSION

In this chapter we set the basis for a new evaluation methodology for motion models which is based on human performance. This work generalises in a rigorous way the evaluation procedures done for most motion models proposed in computational neuroscience. By carefully defining a unified database and proper scoring procedures, it is now possible to perform non-biased comparisons between models, since stimuli are not optimised for a given approach. Our database is freely available on the web together with scoring instructions and results. We provided baseline results for both biologically inspired artificial vision models and state-of-the-art computer vision models. Our results also contribute to show what is the interest to consider biologically inspired models in the computer vision community. By considering stimuli from the psychophysics community, one can further challenge motion models, in addition to existing optical benchmarks.

The proposed evaluation methodology can of course be extended. In this study several properties affecting the motion integration mechanisms were ignored. For instance, disparity used in binocular experiments is missing, and thus it could be probably possible to evaluate other kinds of models where motion and depth are combined.[25,252,200,219] Another property ignored herein is the contrast. In a wide range of psychophysical and neurobiological stimuli, contrast has a considerable effect on motion integration. For instance the receptive field size in areas V1 and MT changes with contrast.[201,175] Contrast also influences behavioural results.[154,241,35] Those kinds of stimuli variations should also be considered in a more comprehensive data set. One could also think about other stimuli. For instance, it is well known 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.[117] Incorporating this multi-stability in models is still only at the sketch level in models,[89,238,227] and mostly ignored in motion benchmarks.

The general idea of benchmarking has been very well understood and applied in computer vision where challenging benchmarks are now available for several key problems allowing models to be compared and further improved. For example, motion estimation performance in computer vision increased significantly thanks to several classical benchmarks, which pointed out strength and weaknesses of state of the art approaches.[17,15]

For instance the benchmark for optical flow introduced by Baker and colleagues[15] defines a set of challenging image sequences with associated ground truth. The choice of sequences was guided by the needs to evaluate models performance on key difficulties encountered by modellers (motion at objects boundaries, occlusions, non-rigid motions, large displacements). The proposed evaluation methodology consisted of several quantified criteria based on local comparisons between ground truth and output from computer vision models.

In the next chapter we investigate how our model behaves on natural scenes used in computer vision, even if it was not proposed in this scope.

Chapter 10 Comparison to computer vision

Following the approaches of Castellanos Sánchez et al.[49] and Bayerl and Neumann,[24] we investigate the suitability of a biological model to analyse realistic videos from classical and recent computer vision benchmarks. Section 10.1 discuss the links between our neural fields model and computer vision approaches. In Section 10.2 we describe the input motion estimation fitted to computer vision requirements in. In particular we consider the effect of temporal and spatial borders. In Section 10.3 we consider two ways of improving the precision of the estimated flow: lateral diffusion and subpixel velocity spaces.

MAIN CONTRIBUTION

- A biologically inspired algorithm for motion estimation.

10.1 RELATIONS TO COMPUTER VISION

One essential aspect of the neural fields 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 several authors[67,70,62] and further extended by Viéville and colleagues.[239] 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 and colleagues[46] or Aubert and Kornprobst.[13]. Thus, one can see a direct relation between the neural fields framework and PDE-based approaches.

As such, introducing the neural fields framework for motion estimation can be related to the series of papers proposing

PDE-based approaches for optical flow estimation, starting from the work of Horn and Schunck.[111] In computer vision, this seminal work has been further improved by many authors.[71,33,247,166] 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 fields 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 fields 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.

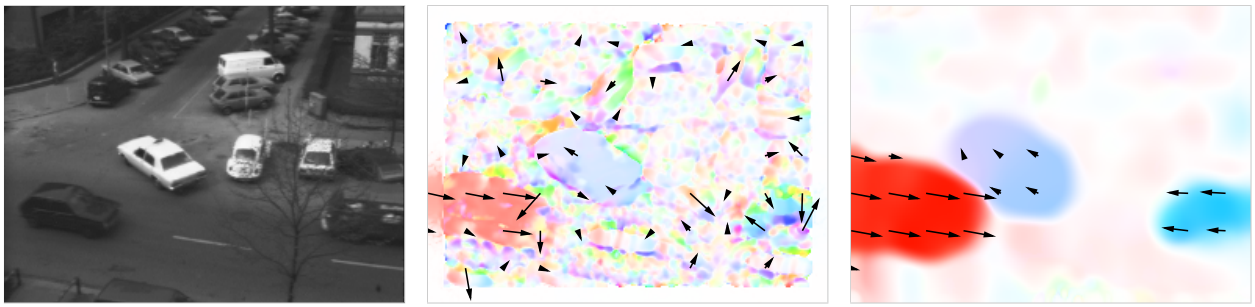
10.2 LOCAL MOTION ESTIMATION

10.2.1 *Frames correlation*

In this chapter, as an alternative to the motion filters described in Section 5.3.1, we use correlation between two frames as input to our model. This choice makes more sense in a computer vision approach since it is a simpler mechanism—two frames are compared—often used in computer vision algorithms. Indeed the filtering approach described in Section 5.3.1 does not consider the initial and final frames due to its large temporal support. Moreover this change proves that our model is relatively independent of its input. We use the correlation motion detector described in Section 4.3.1 and proposed by Bayerl and Neumann[24] but enhanced to support subpixel velocities estimates.

10.2.2 *Flow inpainting*

In order to avoid high errors at the borders in the input, since its computed from displaced frames, we set them to a small value for all the velocities—0.02 in this case. Displaying the optical flow for p_1 and p_2, we can see that those regions get filled when a velocity dominates close to the borders. For instance in the Hamburg taxi sequence (see Fig. 10.1), we observe a filling-in of the borders in regions close to the cars.



(a) (b) (c)

Fig. 10.1 Filling-in the borders. (b) Input Hamburg taxi sequence. (b) Optical flow in p_1. (c) Optical flow in p_2.

10.3 OPTICAL FLOW PRECISION

10.3.1 *Lateral diffusion*

One can wonder if the model defined by (8.2)–(8.3) in Chapter 8 is usable in robotics applications. The precision of the optical flow is particularly important in this case and benchmarking databases were designed to compare the algorithms described in computer vision.[15]

The classical synthetic Yosemite sequence (see Fig. 10.2) provides a good example of a dense velocity field, where a motion is associated to each position.[17] When given as input to the neural fields model with no lateral diffusion—$\lambda^{d}_{1,2} = 0$—the output of the upper layer p_2 is made of patches, as described in Fig. 10.3 (a). A similar effect appears with any of the other models since they do not incorporate lateral diffusion.

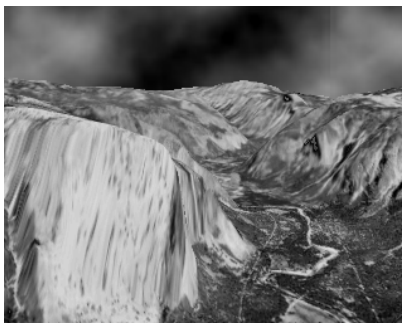
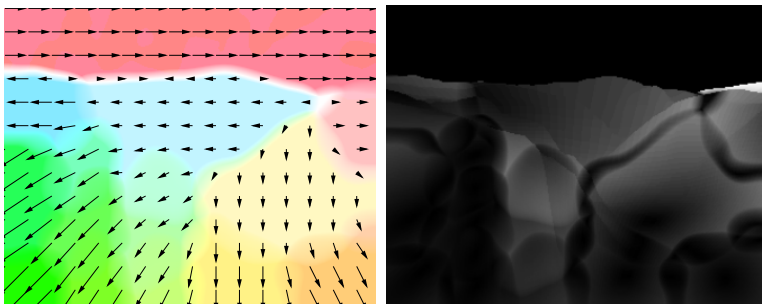


Fig. 10.2 Synthetic Yosemite sequence. This video is classically used to evaluate optical flow approaches. Since it is computer generated, its ground truth is available.

To compare the output of the model to the ground truth, we compute the angular error of the optical flow m_2 (see Section 5.3.2) by:



(a) (b)

Fig. 10.3 Model output on the Yosemite sequence. (a) Optical flow from p_2. The optical flow is made of patches inside which a single velocity dominates. (b) Angular error of the optical flow. The borders between patches have a lower error (*darker*) than the inside of the patches since multiple velocities coexist.

$$\arccos\left(\frac{u_1v_1 + u_2v_2 + 1}{\sqrt{u_1^2 + u_2^2 + 1}\sqrt{v_1^2 + v_2^2 + 1}}\right)$$

where $u = (u_1, u_2)$ is the correct velocity and $v = (v_1, v_2)$ is the estimated velocity.

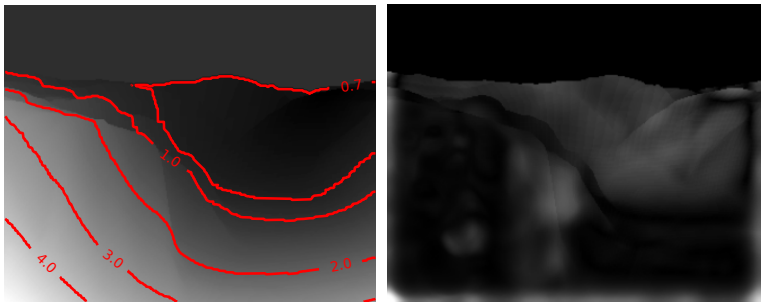
It is due to the selection mechanism, implemented in the model by lateral inhibition, that patches appear in the output. Indeed, a single velocity is preferred at a single position and because of the lateral interactions, a single velocity is preferred in a local neighbourhood. Noticeably, multiple velocities appear at the intersection of patches and lead to an average velocity closer to the ground truth. This averaging can be seen by the lower error made at the patches intersections in Fig. 10.3 (b).

In order to smooth the velocity field, we investigated the addition of a smoothing mechanism implemented as a short range lateral diffusion, as defined in the previous chapter. By increasing the values of λ_1 and λ_2 to 6 and 8, we smooth the output and decrease the average angular error to 6.48°.

10.3.2 *Subpixel velocities*

Plotting the norm of the ground truth and the spatial distribution of the angular error, one can remark that the highest errors are spatially localised in the subpixel velocities region (see Fig. 10.4).

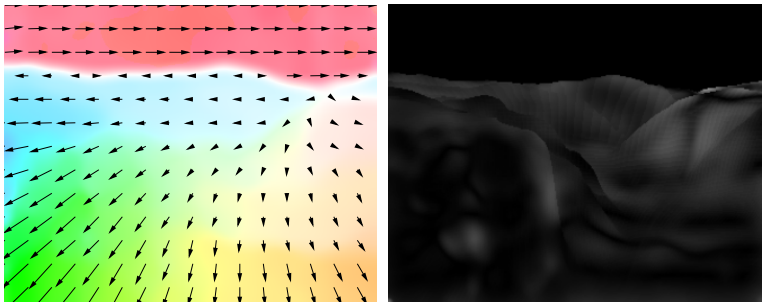
Indeed, due to the selection mechanism Thus, instead of using integer values for the coordinates in the velocity space $\mathcal{V}$, we can sample it by $\mathcal{V} = \{-5, -4.5, ..., 4.5, 5\}^2$. As shown in Fig. 10.5, this choice decreases the error inside the subpixel



(a) (b)

Fig. 10.4 Effect of a coarse velocity space. (a) Norm of the ground truth. (b) Angular error localisation.

velocities regions, and lower the average angular error to 3.97°, a value inside the range of baseline results in computer vision.[15]



(a) (b)

Fig. 10.5 Effect of a subpixel velocity space. (a) Optical flow output from the model (b) Angular error localisation.

10.3.3 *Dynamics*

In Fig. 10.6 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.

10.3.4 *Other results*

Results for the rubber–whale sequence[15] are shown in Fig. 10.7. In this sequence we obtain an average angular error of 10.40°(median 4.00°). The high- est 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.

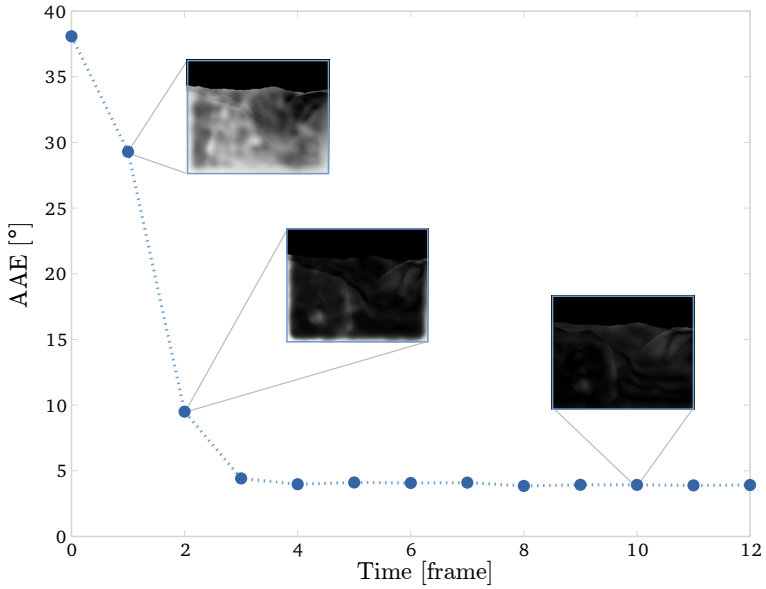
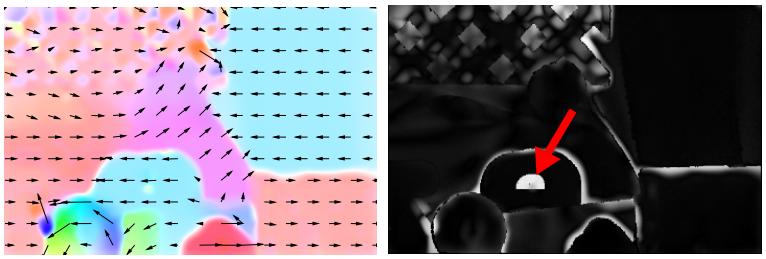


Fig. 10.6 Dynamical evolution of the average angular error (AAE) on the Yosemite sequence.

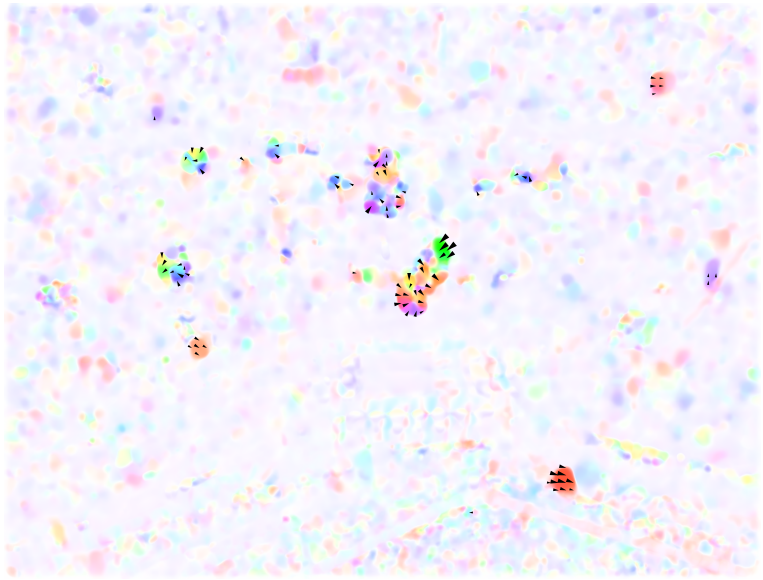


(a) (b)

Fig. 10.7 Rubber whale sequence. (a) Activity response in our p_2 area. (b) Angular error in p_2 (average is 10.40°).

In the context of the European project SEARISE, described in Chapter A, we tested our model on different sequences taken from video-surveillance cameras. Results are show in Fig. 10.8 for the stadium configuration and in Fig. 10.9 for the Duisburg entrance configuration and the parking configuration.

(a)

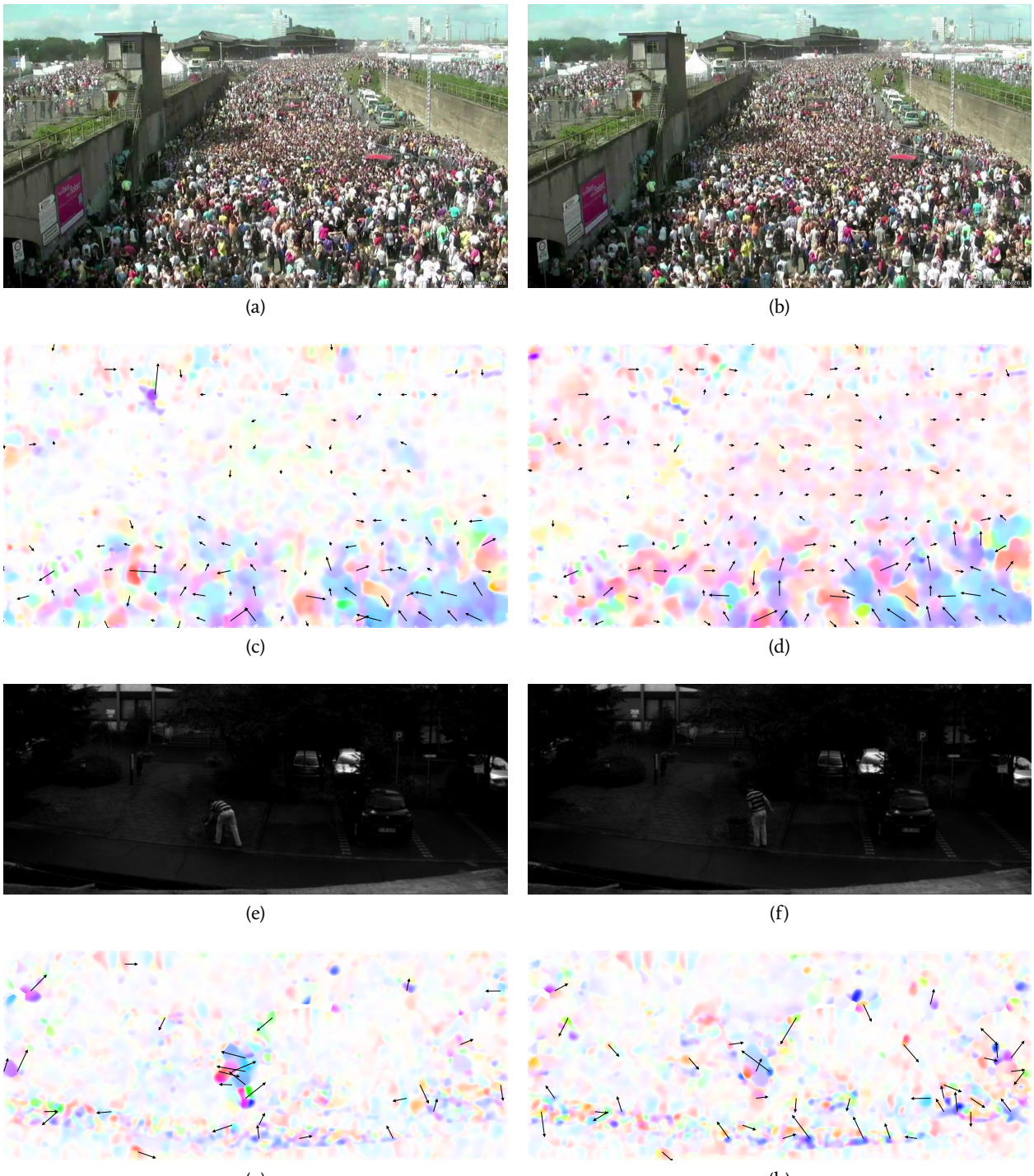


(b)

Fig. 10.8 Results on the SEARISE video-surveillance videos inside a stadium.

Yet applying our neural model on such large video sequences (1616×1228 pixels for the stadium) with large velocity spaces ($|\mathcal{V}| = 441$) has a high memory (several gigabytes per iteration) and

computational (several minutes per frame) cost. Reducing the velocity space, as done by Bayerl and Neumann with the Census transform,[23] or using a log-polar configuration, is one of the possible options to handle high velocities without dramatically increasing the memory usage. Due to the highly parallel nature of the model where all operations are local, the usage of massively parallel processors (GPU, FPGA, memristors) is a solution to decrease the computational cost.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 10.9 Results on the SEARISE video-surveillance videos at an entrance (*first two rows*) and on a parking (*last two rows*).

Conclusion

MODELS OF MOTION INTEGRATION

Perception

We designed and implemented models of motion integration able to reproduce behavioral performance on a wide range of moving stimuli. The tested set of stimuli included various configurations of line drawings, rotating ellipses and drifting gratings. To account for these experiments we designed models of cortical areas implementing small sets of computational mechanisms, without requiring specific motion or junction detectors. Compared to some previous models of motion perception our approach considers the neural implementation of the proposed mechanisms in V1 and MT.* Also, our approach explains a wider range of motion percept.†

* For example, descriptive models such as Bayesian models, generally focus on reproducing the observed percepts,[249,157] whereas we aim at explaining the underlying neuronal interactions.

† Mechanistic models are based on neuronal interactions but are often more restricted in the stimuli they explain, for instance they do not consider contextual influences[208] or non-binary gratings.[24]

FUTURE WORK

- Most experimental stimuli are multi-stable (see Section 1.4.1). an extension of our models should be able to account for change in perception over long durations (>1s). One possibility to achieve changes in time is to include *adaptation*. Such an adaptation can be implemented with second slower variable which inhibits the activity in order to select the second most active velocity. We expect an oscillatory solution to be found in this case.

  Yet other mechanisms such as stochasticity can explain changes of perception, and they those possibilities would have to be compared to experimental data.

- For simplicity and interchangeability, we designed models were V1 cells are speed- and direction-tuned. Yet the receptive fields of V1 cells can be considered as blobs in the frequency space[148] Finer details in the representation of motion in V1, for instance by implementing a feedback dir-

ectly of the filter cells, would be helpful to study the effect of changes in receptive fields (see Section 3.2.5).

We could implement a model where the feedback directly influences the cells corresponding to the filter blobs, without passing through our V1 cells. In doing so we expect the response of filters to evolve over time, allowing comparisons with experiments on tuning changes.[141]

Dynamics

The proposed dynamical model is able to account for motion integration dynamics, and we linked it to perception dynamics and smooth pursuit eye movements. Our model is able to reproduce the directional dynamics on various configuration of translating bars and we made some predictions concerning gratings and line drawings which are yet to be validated.

Compared to previous models our approach proposed a novel modelling link between neural interactions in V1 and MT and integration dynamics. Previous similar models focused on motion perception (except the model of Pack, Grossberg and Mingolla[173] which focus on the maintenance of smooth pursuit). Yet perception and dynamics are two intertwined aspects of motion integration, and we are convinced that modelling of both behavioural and perceptual experiments would be fruitful.

FUTURE WORK

- Predictions of the model concerning dynamics remain to be tested by psychophysics. For instance the dynamical shift from the vector average to the intersection of constraints shown in Fig. 6.6 (a) is well known in the literature. Yet, in our experiment, it appears to be following one velocity line. Comparing smooth pursuit or precise perception dynamics on this example might validate this model behaviour.
- To analyse the effect of the contrast we varied the amplitude of the motion input, or added noise. Yet a real mechanism to account for contrast remain to be added in the

model, and in particular in the motion detectors which are still contrast invariant. A solution would be to use the motion detector used by Weiss, Simoncelli and Adelson.[251]

- *Motion anticipation* described in Section 6.2.1 is currently being experimentally tested in macaques. In our model anticipation is still spatially limited by the purely multiplicative feedback. A subthreshold preparation feedback might help the anticipation of motion over larger spatial domain. Together with contrast handling in the model, anticipation might produce faster adaptation for higher contrasts. With the addition of different delays (see neural fields section below), we would be able to verify if anticipation comes from lateral of feedback interactions.
- Most comparison where made with the dynamics of direction error, since it is the data most often found in the literature. Yet our model is able to output *speed dynamics*, the magnitude of the motion, as another kind of prediction. We expect future smooth pursuit recordings to provide more speed data and thus further validate the proposed model.

Contextual influences

We proposed a novel form-motion interaction to account for motion segmentation. We suggest that luminance information, from cortical area V2, gates motion integration in V1–MT. Such a simple model explain the motion perception in stimuli with different moving objects such as the simulated edge of Huang, Albright and Stoner[113] or the chopstick illusions.

Compared to previous models our approach proposes a novel mechanism for motion integration. It also proposes an alternative to more complex junction detectors and depth layers[23,32] in some stimuli. Yet depth ordering is not considered in our model.

FUTURE WORK

- We proposed that luminance information gates motion integration, but the mechanisms underlying such a contex-

tual modulation remains to be found. One possible explanation for the binding of cross-modality informations—luminance and motion here—is *spike correlation*.

Several spiking models have been proposed to segment motion or luminance, as well as implement cross-modal interactions through synchronisation.[53,54,265,261] If such a synchronisation is observed in the visual cortex, it might serve as a binding mechanism between cortical areas.

- More generally, finding a mediator of cross-modality interactions might help incorporate other features (for instance texture) to help motion integration. In addition to synchronisation, one other possibility for binding is multiplexing—cells tuned for several features. For instance DeAngelis and Newsome[65] found disparity columns in cortical area MT, for patches of neurons also selective to specific motion direction. Mante and Carandini[148] propose thaht certain cells of V1 exhibit a tuning combining orientation and motion direction, based on physiological recordings.[21]
- We proposed the luminance-gating mechanism as a form cue to segregate between objects. Other models often use *depth ordering* to support for multiple motions.[32,23] Yet these approaches are limited to the number of layers implemented in the model. To account for arbitrary multiple motions—also named *transparent motion*—we suggest two alternatives.

  The first is to localise the inhibition in the velocity space. Currently the selection mechanism is inhibiting equally all the velocities at a given position. By sharpening its connectivity to a Gaussian function we can allow several solution to coexist.

  The second is to consider a synchronisation model either as a full spiking models as discussed previously, or by adding a *phase* domain. Then different motion can coexists at a given position, if the connectivity is set accordingly.[55]

Neural fields

We investigated the mechanisms involved in our model both analytically and numerically. To facilitate the analysis as well as the linking to other models, we rewrote our model with neural fields equations. Such a common framework allows to compare the mechanisms in our model to others and unravel the neural computations carried across the cortex. We also numerically investigated the role of lateral inhibition and searched for bifurcations in a simplified version of the model.

Almost none of the previous models use neural fields equations for motion analysis. One notable exception is Giese[89]. Indeed studies of neural fields equations are mainly theoretical[254,7,76] or focusing on saccades,[222] attention,[83] or orientation in V1.[237]

FUTURE WORK

- Following neural field research[238], further bifurcation analysis could be applied to models and explain various multi-stable percepts. In particular our model is able to give multiple solutions, with small changes to the initial conditions, on the multi-stable stimuli presented in Section 1.4.1.

  Because of the numerous equations to be solved in our model, in particular due to the spatial and velocity discretisation, applying continuation tools directly of the full model is extremely computationally demanding. We suggest two non-exclusive alternatives. First, the model can be approximated by Pincherle-Goursat kernels,[234] as suggested by Veltz and Faugeras.[238] Second, we can simplify the model to keep only selection and spatial diffusion in a 1D spatial domain and 1D directional domain, as we did in Section 8.5.2.

- The luminance diffusion mechanism remains to be incorporated into the neural fields formalism. We already show how to include Laplacian operators and multiplicative feedback into neural fields.

- Adding *delays*[76] in the equation should constrain the connectivity of the model. First, although distant connections are low, the model is currently not causal, as are almost

all models in the literature. Then, Deco and colleagues[66] showed that delays are useful to model forward/backward interactions. As described above, delays taken from the physiological literature can constrain models enough to be able to differentiate between lateral and feedback mechanisms, for instance in motion anticipation. Physiologically this work has been started by Angelucci and Bullier[8] who shown that long-range lateral connections in V1 are too slow to explain the orientation-selective surround field. Finally, delays can model the speed difference between the parvo- and magno-cellular pathways.

PERFORMANCE ANALYSIS

Comparison to human performance

We proposed a motion estimation evaluation methodology for motion integration approaches. Our goal was to evaluate the results of a given model by scoring how much it fit to the experimental data. In this view, we established mappings between experimental time or space configuration and benchmark sequences to maintain the relations across stimuli. Our benchmark is divided in two evaluations: The static evaluation, which allows comparison with computer vision approaches; The dynamical evaluation to compare with motion integration dynamics such as smooth pursuit.

FUTURE WORK

- The set of stimuli is still relatively small and should be extended. Random dot kinematograms are still to be incorporated as a third class of stimuli. Contextual influences such as the various barber poles and chopsticks stimuli are also missing.
- Luminance—and thus contrast—is not yet considered by the database, although being a strong contextual modulator of motion integration. Its discretisation will require the usage of another file format since computer images are often limited to 8 bits.

- Our perceptual performance is based on motion direction only. Yet several other measurements are used in psychophysics, including coherent motion, discrimination, and forced choice.

Comparison to computer vision

We compared the results of our model to computer vision approaches of motion estimation. A set of video sequences from the computer vision show that our model is able to give good results even for large input and complex stimuli. Several characteristics allow a V1–MT model of motion perception to perform well for computer vision. First, the hierarchical architecture of the visual system enables large scale diffusion. Second, the model is able to *fill-in* regions with invalid input, such as borders. Finally, our dynamical model remembers the past, and does not process the input frame by frame.

Previous biologically inspired algorithms have been proposed to handle computer vision problems. For most of them, biology is a source of inspiration and they do not consider motion percepts.[49] The model of Bayerl and Neumann[24] is a successful model bridging the gap between modelling and computer vision and served as inspiration to our work.

FUTURE WORK

- Large velocities remain a problem in the current implementation due to peculiar nature of our distributed representation. A sparse or log-polar motion tuning will certainly improve the computational cost and the results, without having to change the model. An improved connectivity and readout mechanism might then compensate the lack of precision in the velocity field by interpolating the population response in MT. Such as mechanism will also increase the biological plausibility since target speed can be obtained by taking a weighted average of the population, and not of single cells.[59,189]
- In terms of validation, a careful evaluation of optical flow

performance remains to be done. In particular recent computer vision benchmarks[15] could be used to fully quantify the results of our models.

Part IV

Appendices

Appendix A European project SEARISE

This PhD thesis was partially funded by the European project SEARISE.

The SEARISE project developed a trinocular active cognitive visual system, Smart-Eyes*, for detection, tracking and categorisation of salient events and behaviours. The system has capability to learn from and self-adjust to the ever changing visual input; fixate at salient events and follow their motion; performs visual categorisation of salient events based on environmental context and a set of policy rules. The "heart" of the visual system is made of two active stereo cameras, the binocular cameras, which automatically fixate at the salient object, follow its motion, and then switch the attention to another salient location. The system performs multi-scale analysis by zooming individual parts of attended events, which might either uncover object's identity or display its salient actions in details. The "brain" of the visual system is a cognitive model of visual processing replicating computational strategies supported by neurophysiological studies of the mammalian visual cortex. The Smart-Eyes system combines an engineering paradigm for coordinated eye-like movements of the binocular cameras with an innovative computational theory of visual cortex. The system responds to events happening in its field of view by continuously switching its attention to those objects or object parts exhibiting most salient actions.*

A.1 MAIN CONTRIBUTION

The motion models described in thesis provided the substrate for a strong scientific collaboration with Heiko Neumann and Jan D. Bouecke from the Ulm University. They allowed a common motion processing model to be implemented in the SEARISE system, with better parameters set up, and faster implementation.

The collaboration has been extended to the development of the evaluation methodology for motion estimation models based on human performance presented in Chapter 9.

A.2 OTHER CONTRIBUTIONS

A.2.1 *Hardware implementation*

Back at the beginning of the European project, the SEARISE team was supposed to work on conventional CPU hardware. Thanks to the previous knowledge developed by partners in the former Odyssée team (now Neuromathcomp), we were able to propose several alternative technologies which eventually appeared to be far superior. Indeed the members of the Odyssée team, had a strong experience on advance high performance computation.

The persons which were involved at the time include Horacio Rostro-González for FPGA, and Émilien Tlapale and Jérôme Piovano for GPGPU. Note that the GPGPU technology pushed on the scene by the INRIA appears to be the dominant technology in all computationally expensive algorithms in the scientific fields.

Thus, our experience has been able to be disseminated into the other teams of the SEARISE project, and is now used intensively by, for instance, Heiko Neumann and Silvio Sabatini's project.

Moreover our previous experience has allowed to debug problems appeared long after the beginning, that were intricate to the GPGPU.

A second benefit of proposing this alternative has been to be able to upgrade, by a graphic card upgrade, the expensive system initially bought by SEARISE, without having to replace the full system.

A.2.2 *Engineering collaboration*

We proposed and defined a *common format* for storing input and results of simulations, allowing modules to be independently tested by the team members. This proposal involved the HDF5 format, in which we already had a lot of experience with, which is becoming a *de facto* standard for scientific format interchange. Common formats allows to test different modules, both developed by the project, but may also be used by other external teams.

A.2.3 *Core software*

Among the various areas where we led the way was the GPUData library. Based on our previous experience in the team with *CImg*, *SciPy*, and several other multi-dimensional scientific libraries development, as well as our GPGPU experience, we were able to propose a high-performance GPGPU library for CUDA, probably the first existing at that time. Eventually most of the code produced by the SEARISE project is built on top of our initial contribution.

Appendix B Aiding low vision patients

In this appendix we propose a reading aid software for low vision patients. This work started as a postgraduate fellowship in a collaboration with Éric Castet and Jean-Baptiste Bernard from the CNRS. It was extended during the PhD thesis, although not being our main objective. Section B.1 introduces the problem of low vision in and its consequence to reading. In Section B.2 we describe the existing reading aid as well as their capabilities. In Section B.3 we propose a reading aid software, Navisio*, and describe its functionalities for complex documents reading. Helped by orthoptists at the hospital* La Timone *in Marseille, we conducted a comparison experiment on 26 low vision patients and present the results in Section B.4. We conclude in Section B.5, suggesting future extensions of our reading aid.*

MAIN CONTRIBUTIONS

- A reading aid software for low-vision patients
- Experimental validation with patients and experts

B.1 LOW VISION

Low vision is defined as a permanent form of visual impairment which cannot be corrected with an optical correction and which leads to the inability to read the newspaper from a conventional viewing distance of 40 cm.

Leading causes of low vision are age-related eye diseases like macular degeneration (≥ 50%), glaucoma, diabetic retinopathy and cataract. They cause a decrease of the visual acuity and dramatically affect everyday life. Most of these diseases induce central field loss: Patients are blind in their central visual field (this is defined as a scotoma) as they cannot use the fovea, the highest resolution part of the retina to explore visual scenes (see Fig. B.1).

(a) (b) (c)

Fig. B.1 Qualitative representation of the perceived image by a patient in front of a zoomed text displayed on a computer screen. (a) Image of a text displayed on a standard video monitor, with (b) the simulated acuity depending on the gaze and eccentricity and (c) with a black area corresponding to an absolute scotoma area. The legible letters correspond to the average visual span.[192]

In 2004, the World Health Organization estimated that 124 millions people suffered from low vision in the world, using the visual acuity (ability to recognise small letters) to characterise the low vision impairment (see Table B.1).

In most cases, medical treatments can slow down the development of the pathology to blindness. Unfortunately, there is no treatment to recover the damaged part of the eyes in low vision. Visual prosthesis and ophthalmic devices recently made substantial progress, but we are still far from allowing patients to perceive fine details.[232] Consequently, patients have to adapt their everyday life activities, and especially reading, which is often declared as the most dramatic loss.

Table B.1 Ranges of visual acuity loss (*Source*: World Health Organization)

Visual Acuity (Snellen scale)	Visual Impairment
≥ 20/30	Normal Vision
20/30 to 20/70	Mild Visual Impairment
20/70 to 20/200	Moderate Low Vision
20/200 to 20/500	Severe Low Vision
20/500 to 20/1000	Profound Low Vision
≤ 20/1000	Blindness

Reading speed is currently the most chosen indicator to measure reading performance for low vision patients. Different calibrated tools measuring reading speed allow to characterise precisely the reading deficiency of the patient (see the MNREAD test[135]). In normal vision, reading speed increases with character print size until a maximum reading speed is reached.[134] When this maximum reading speed is reached, the corresponding character size is called the critical print size (CPS). The same observation is true for low vision patients,[137] but the CPS is higher, and the maximum reading speed is less. Thus raising the print

size only for low vision readers is not sufficient to recover the "normal vision" maximum reading speed. This is explained by other limiting factors besides visual acuity in low vision reading such as the restriction of the visual field (holes and distortions in the visual field), lateral masking (also called crowding[181,182]), fixation instability[63] or oculomotor inabilities.[48]

The complexity of the document (as shown in Fig. B.2) is also an important limiting factor in low vision reading. The patient has to extract different paragraphs and pictures and to read them in a defined order. Reading speed for these documents is naturally much slower.

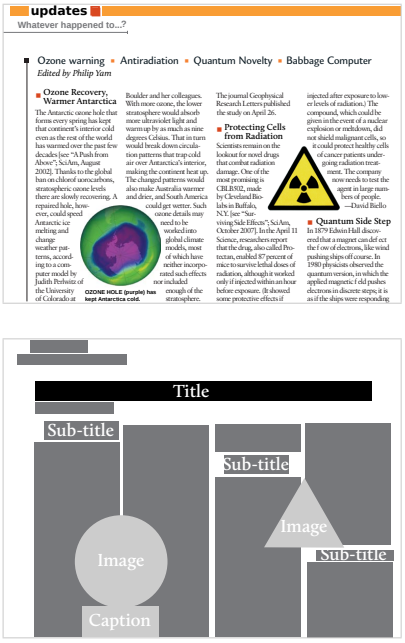


Fig. B.2 Illustration of a complex document layout composed as a set of titles, paragraphs and images.

B.2 READING AIDS IN LOW VISION

B.2.1 *Comparison criteria*

Optimal character display As mentioned in the introduction, texts must be displayed at a sufficient size to be read with a maximal reading speed. This size can be very high for severe pathologies (≥ 2° visual angle for a mean character high) whereas this CPS is about 0.2° for normal vision.[134]

Many psychophysical studies showed the interest to improve the legibility of a text to increase reading performance. People with low vision usually read better with white-on-black text[210] and a maximum contrast is necessary for an optimal reading.[197] A strong variability between patients exists for fonts[11] and for background/foreground colours[136] in reading speed.[133,198]

Some experiments tried to decrease crowding effect between letters in low vision reading. These studies tried to increase inter-letter spacing,[57] interline spacing[31] or to inverse polarity between each adjacent letter.[58] They showed little (for the interline variation) or no significant gain on the reading speed.

To summarise, an optimal character display is a high luminance contrast display with a magnified print size superior to the CPS. Other display parameters like colour or font have to be defined individually for each patient.

Optimal presentation of a text Magnifying a text on a screen implies moving it dynamically because of the *window-size effect*: a trade-off between the angular size of characters and the field

size.

Automatic scrolling was proved to be the fastest way to read a text on a screen for low vision patients, rather than manual methods using visual-motor combination to scroll text.[100] The automatic drift mode was also the preferred mode in normal and low vision, and no difference was found between horizontal and vertical scroll.[39]

With the Rapid Serial Visual Presentation (RSVP), each word is successively displayed on a screen to limit eye movements during reading. Exposure time can be constant, length-word dependant or the patient can decide when to display another word. RSVP reading speed is as fast as with scroll presentation, but users usually do not consider RSVP as pleasant.

The local/global navigation problem A complex page layout (see for example Fig. B.2) introduces a local/global navigation problem: It is necessary to enlarge one or different local regions of the document to infer the global layout.

B.2.2 *Existing reading aids*

In this section we give an overview of the main existing systems for accessibility. Table B.2 summarises some of the key features. Note that useful links are available in the electronic version of this document.

	Clearview+	myReader2	ZoomText	Jaws	Navisio
Printed documents	✓	✓	—	—	—
Electronic documents	—	—	✓	✓	✓
Dedicated hardware	✓	✓	—	—	—
Audio output	—	—	✓	✓	—
Text reformatting	—	✓	✓	—	✓
Automatic scroll	—	✓	✓	—	✓
Local/Global display	—	—	—	—	✓

Table B.2 Comparison of typical existing systems and Navisio. Note that text reformatting is the possibility to change the text display more than with a simple zoom.

Reading aid systems for printed documents These systems are frequently used by low vision patients who want to read or write. A region of the document is captured by a camera and magnified onto a monitor. One can distinguish two main systems:

- The closed-circuit television (CCTV) magnifiers (such as *Clearview+*): CCTV magnifiers display in real-time the captured image onto a monitor. It is usually possible to maximise the contrast and to select the background and foreground colours. Text can be zoomed up to 60x with a high-quality display. Portable hand-held magnifiers are also proposed. As far as reading is concerned, using a CCTV magnifier requires combining eye movements with manual movements to move the magnifier over the text. Thus, reading with a CCTV magnifier is more difficult than reading a static enhanced text. In particular, there is an important local/global navigation problem with CCTV magnifier because of the difficulty for the patient to know his relative position in a complex document. Jumping to the next line of a text also takes time: from 17% up to 50% of the total reading time is dedicated to the retrace time. More details on CCTV performance evaluation can be found in the literature.[26,38]

- The *myReader2* tool: *myReader2* is a reformatting system which scans the document, performs text analysis and recognition (which may take several seconds), and displays text in a new format (text parameters, scrolling type). This tool has been tested on low vision patients, and it was compared with CCTV: Its efficiency was proved in terms of reading speed and reading comfort.[101] A special control panel is also proposed to optimise ergonomics. However *myReader2* has some limitations to deal with complex documents.[180]

Reading aid systems for electronic documents Beside the built-in accessibility features proposed GUI, operating systems, navigation engines or edition software (zoom modes by Microsoft, PDF accessibility by Adobe, Firefox accessibility features), there is a wide variety of relevant software available. One can distinguish two main systems:

- The *computer-screen magnifier* aid: Some parts of the screen can be enhanced (see, *ZoomText* or *Lunar*). The navigation and the selection on the monitor are controlled by an

enhanced and slowed cursor moved thanks to keyboard or mouse. The text can be zoomed up to 20× and is displayed with the possibility to customise contrast, colours, and an anti-aliasing function in most software. Another interesting feature, for example proposed by *ZoomText* is that subjects can read a selected text in a special environment, so that text is reformatted for easier reading.

- The *screen reader* aid: Thanks to this processing, blind or low vision patients can listen to document files read aloud by synthetic speech (Jaws, Windows-Eyes or *Vocale Presse* for French press).

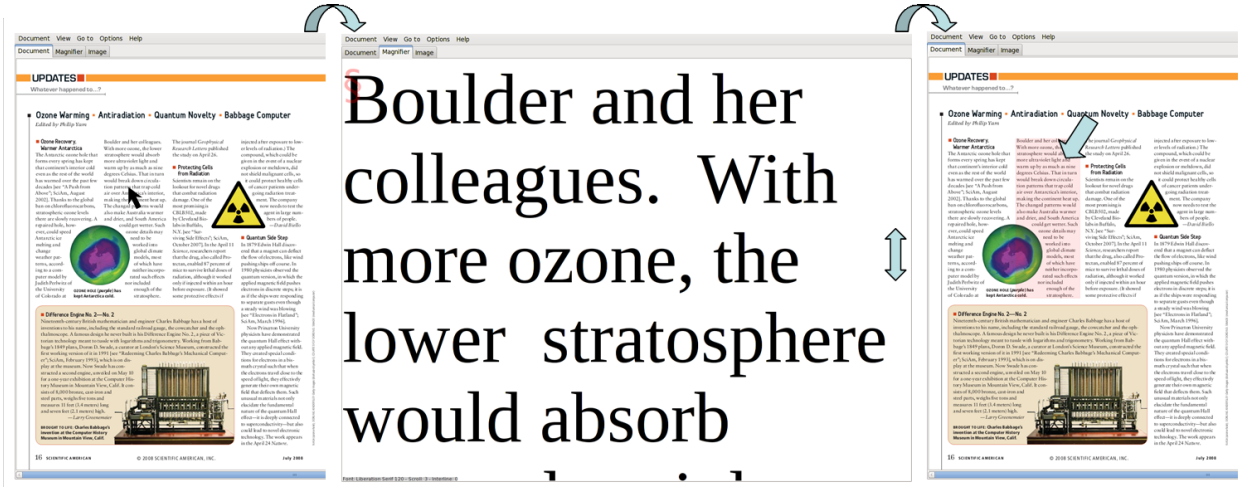
Usually, reading aid software propose the audio and enhanced functionalities and patients can use both modes.

B.3 ENHANCED READING WITH NAVISIO

Navisio is an electronic document magnification software. It was programmed in C++ and it runs on Windows or Linux systems. The goal of *Navisio* is to help low vision patients read simple and complex PDF documents with text and images. *Navisio* offers new navigation possibilities based on two view modes (see Fig. B.3): a document view and an enhanced view. The document view corresponds to the global standard view of the document, and the enhanced view corresponds to a reformatted view of a paragraph (or an image) which has been selected in the document view. The user can easily switch from one view to the other.

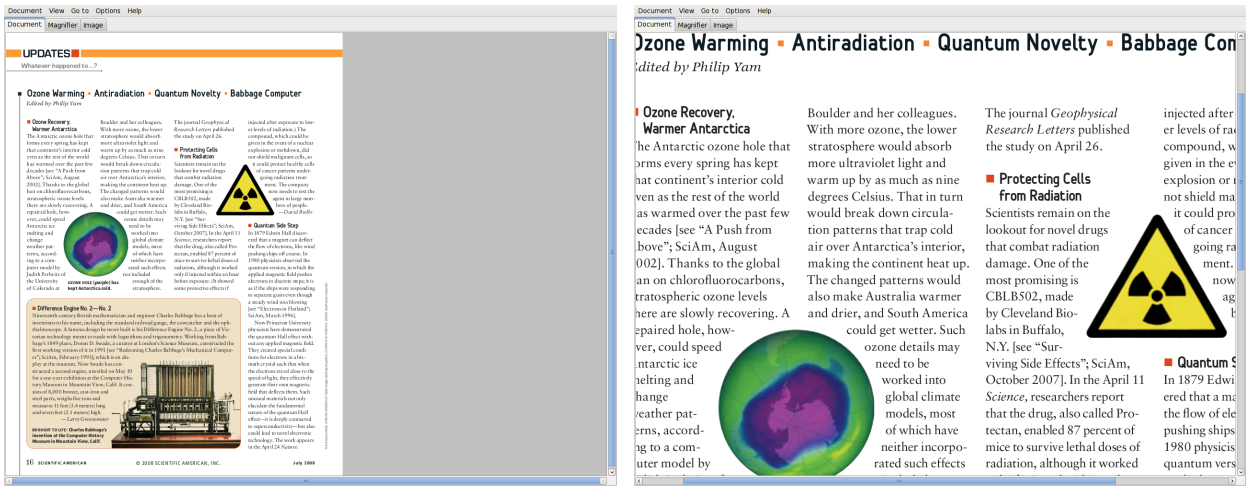
B.3.1 *The global document view*

In this mode, the raw document is presented with text paragraphs and images. The user navigates in the document with a cursor (shape and speed of the mouse are adjustable) controlled by the mouse. The user can zoom on the document before selecting a part with the cursor (see Fig. B.4). Once selected, the part of the document is displayed in the enhanced view (see Section B.3.3).



(a) (b) (c)

Fig. B.3 Navigation between the document view and the enhanced view. (a) The patient selects a paragraph in the document view. (b) This paragraph is displayed on the enhanced view according to patient (c) At any time, the patient can come back to the document view: the paragraphs just read are highlighted (*red*).



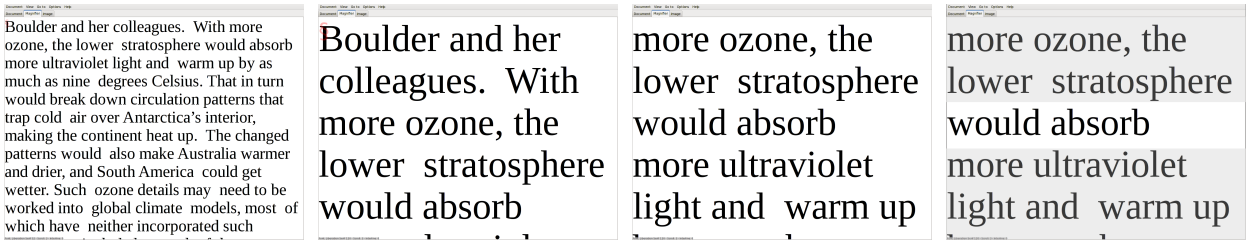
(a) (b)

Fig. B.4 Example of document view: (a) Once the document is opened, it is shown full page. (b) Then patient can zoom in and move to focus on specific parts.

Navisio proposes a colour code in the document view: Every paragraph which has been displayed in the enhanced view will be highlighted with a colour-code. The colour is different if the paragraph was just selected or selected previously (see examples in Fig. B.3 (c) and Fig. B.6 (a)). Thanks to this colour-code, the patient knows the exact position of the paragraph which was just read and also knows what remains to be read.

B.3.2 *The local enhanced view*

In the enhanced mode (see Fig. B.5), the user can magnify a selected paragraph with a large panel of reformatting parameters, most of which are related to psychophysics results (see Section B.2.1). With customisable shortcuts, the user can easily adjust the display (font, character size, interline size, background and fore- ground colours, and other enhancements) or select the scrolling type (horizontal or vertical). Note that all this personalised set of parameters can be saved in a configuration file so that several users can use *Navisio* and just need to reload their preferences.



(a) (b) (c) (d)

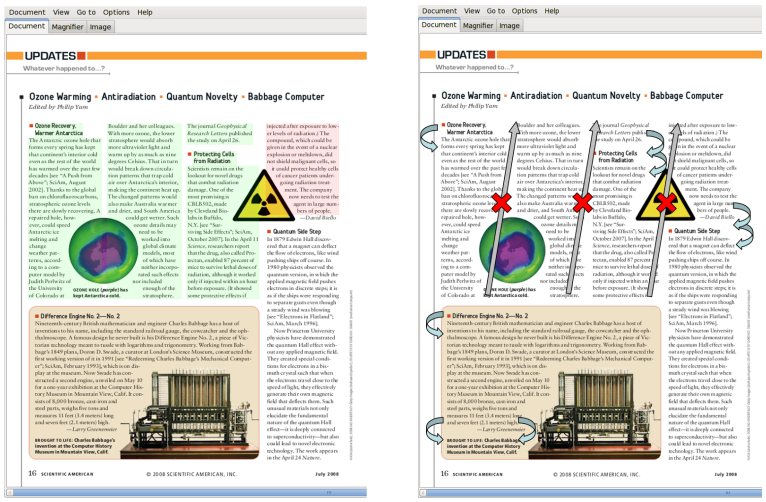
Fig. B.5 Some features of the enhanced view: (a) Example of an enhanced view. (b) Same paragraph with different display parameters. (c) The patient then reads the text scrolling with the mouse (the scrolling is manual; drift speed can be adjusted). One automatically reads all subsequent paragraphs if some have been found (see Section B.3.3). (d) Some enhancements are also possible such as putting more relative contrast to the central line.

If an image instead of a paragraph is selected, the display is automatically in the maximum resolution (the image can have been reduced in the raw document) and can also be zoomed.

B.3.3 *Document layout analysis with Navisio*

Navisio integrates a document layout analysis. Given a PDF document, each paragraph is segmented based on the PDF file structure, and some relations can be established between paragraphs using simple heuristics. In Fig. B.6, if one selects the paragraph

at the top of a column, then every paragraphs will be displayed sequentially (a symbol indicates the change between two paragraphs).



(a) (b)

Fig. B.6 Layout document analysis (a) As soon as a paragraph, defined by the layout analysis, is shown in the enhanced view, a colour code will indicate it in the document view. (b) In general, our document layout analysis allows the patient to read continuously different paragraphs related thanks to heuristics. Thus, for example, paragraphs in the same column are shown continuously in the enhanced view as the patient scrolls down. More complex transition such as changing columns are not handled in this version.

Of course simple heuristics are not sufficient to accurately deal with complex document. In order to improve this, one possibility is to better take into account PDF format structure, as specified by Adobe (see PDF accessibility by Adobe), as soon as the file was properly generated. Another possibility comes from the computer vision community concerning document layout analysis.[27,119] One advantage of the later solution is to allow one to consider any kind of document, including captured images of printed document (we refer to Section B.5 for more comments).

B.4 EXPERIMENTAL RESULTS

This section presents tests with 26 low vision patients. *Navisio* was evaluated according to two criteria: reading performance and reading comfort.

B.4.1 *Materials and methods*

Subjects 26 low vision patients (from 60 to 92 years old, average age is 78.2) participated in the tests. Their pathologies were mainly AMD (about 80% of the patients). Their maximum

reading speed, previously measured thanks to a French reading computer-based test ranged from 23 to 101 (average is 53.35) words/minute. Most patients had a weak knowledge of computer equipment but some often used electronic magnifiers.

Reading aid systems Subjects used two reading aid systems: an electronic CCTV magnifier (*Clearview+*) and a computer with *Navisio* installed. Note that both systems do not work with the same input, and that this study is valid for documents which exist in printed and electronic version, such as newspapers.

Texts and documents To estimate reading speed, one needs a database of calibrated texts since one cannot use twice the same text to compare two systems. In order to guarantee that different texts contained the same complexity and semantic information, we used the French standardised text database from Hanh and colleagues.[99] In this database, reading speed was shown to be equivalent for ten French texts for normal sighted people. These texts are left justified, and the selected font is *Times New Roman*, and they were matched in length (830±2 characters). In our experiments, we used them to compose two kinds of documents (simple and complex, see Fig. B.7). Electronic version is used with *Navisio* and a printed version is given for the magnifier. Proposed texts correspond to 0.5° of visual angle for a subject at 40 cm from the monitor and titles correspond to 1.5°.

(a) (b)

Fig. B.7 Sample documents
(a) Simple documents: One standardised text with a title is centred on the page (b) Complex document: Two standardised texts are split into 4 paragraphs and distributed on a page with titles.

Measuring reading performance Oral reading speed for a subject is measured for both simple and complex documents. If reading speed was considered as too slow by the orthoptist, only one standardised text was read in the complex document case. During the reading phase, each misread word was marked. Right after the reading phase, three questions were asked to estimate the understanding of the subjects for each text. When the test was finished, subject was asked to give a grade concerning the feeling of comfort for each tool.

Description of protocol Protocol is divided into two sessions. During the first session (about 30 minutes), the subjects learned how to use both tools and selected the optimal magnifying factor for the magnifier, as well as the best display parameters in the *Navisio* software. The distance to the screen was not fixed. During the second session (about 30 minutes), the subject randomly read simple and complex documents with the magnifier and with *Navisio*. For a simple document, the subject read the title and the text as soon as the page is given. For a complex document, the subject had to read two or four titles and texts (out of the four texts proposed) in a random given order to force the patient to navigate in the document. Mistakes in the reading order were marked by the orthoptist. Reading speed was measured during that session.

B.4.2 *Results*

Results about reading speed, reading understanding, words misreading proportion and reading comfort are shown in Fig. B.8.

Fig. B.8 (a) is about reading speed measurements for simple and complex documents in low-vision patients. It shows an overall increase of reading speed by 18.5% for *Navisio* with simple documents (close to *myReader2* benefit results[101]) and 16.5% for complex documents. Variation is significant with a T-test for dependent samples in simple document ($t=-5.09678$, $df=25$, $p=29\times10^{-5}$) and in complex document ($t=-4.79425$, $df=25$, $p=6.4\times10^{-5}$). Also, results show that reading a complex document takes 8% longer than reading a simple document with a CCTV magnifier and 11% longer with *Navisio*. It is likely that texts with more complex

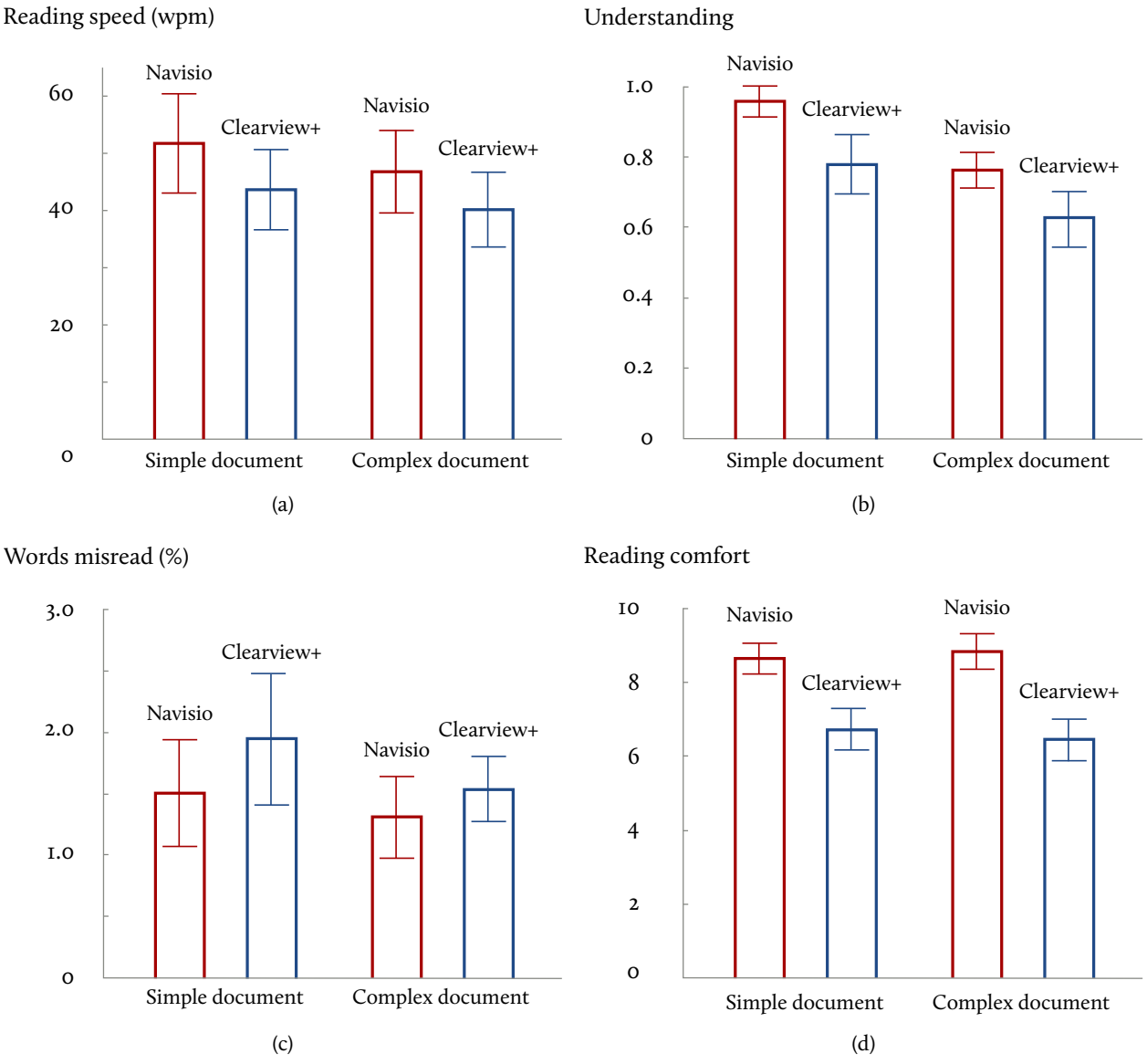


Fig. B.8 Experimental results with low vision patients (a) reading speed (words/minute), (b) understanding evaluation (ratio of correct answers to the questionnaire), (c) words misreading (percentage), (d) reading comfort (average grade between 0 and 10). Bars are 95% confidence interval.

structures would have shown a higher reading speed difference and a better profit of *Navisio*'s main purpose: the facilitation of the local/global navigation.

Fig. B.8 (b) and (c) show that *Navisio* also brings a meaningful advantage in terms of understanding (+22% for simple document and +23% for complex document) and reading mistakes (-23% for simple document and -15% for complex document). Thus for both kinds of documents, *Navisio* increases reading performances. For understanding, variation is significant with a T-test for dependent samples in simple document ($t=2.241040$, $df=25$, $p=0.034146$) and in complex document ($t=2.331032$, $df=25$, $p=0.028113$). For reading mistakes, variation is significant with a T-test for dependent samples in simple document ($t=-5.40062$, $df=25$, $p=1.3\times10^{-5}$) and in complex document ($t=-5.12989$, $df=25$, $p=2.7\times10^{-5}$).

Fig. B.8 (d) shows the most significant difference between *Navisio* and the CCTV magnifier: the reading comfort. *Navisio* was given an average grade of 8.65/10 for the simple document and 8.85/10 for the complex document, while for the CCTV magnifier it was 6.73/10 respectively 6.46/10. Variation is significant with a T-test for dependent samples in simple document ($t=-5.95238$, $df=25$, $p=3\times10^{-6}$) and in complex document ($t=-6.70751$, $df=25$, $p=10^{-6}$). The reading comfort proposed by *Navisio* is its most important interest, especially for complex document reading. Finally, when the patients were asked which tool they would prefer to use at home if they had the possibility to have the same printed and electronic document, 85% of the patients answered they would prefer *Navisio*. If we do not take into account patients who often use a CCTV magnifier, this percentage is about 96%. Interestingly, *Navisio* was a really appreciated reading aid system, even by patients not used to navigate on a monitor with a mouse. These results prove the interest for such a high parametrisable electronic document reader with local/global navigation.

B.5 CONCLUSION AND PERSPECTIVES

In this chapter, we described the new *Navisio* software which main goal is to help low vision patients read complex docu-

ments. The main feature is to enable a smooth navigation between a global document view and a local customizable view to read an electronic document (here a PDF file). We showed how the local view parametrisation is related to recent psychophysical results on reading performance. *Navisio* was evaluated with low vision patients and compared with a classical CCTV magnifier tool. Our tests revealed the efficiency of *Navisio* to read simple and complex electronic documents: Reading speed and comfort were significantly enhanced.

Of course, *Navisio* is intended to evolve and include more functionalities. Among forthcoming extensions that we consider, let us mention the addition of an audio support, further features to facilitate navigation and the possibility to select the RSVP in the enhanced mode.

Beside the realisation of this software, the goal was to integrate in a software existing and novel ideas to help low vision patients to read any kinds of electronic documents, and to have those ideas justified from the psychophysics point of view. The principles introduced here could be easily extended and applied to any kind of electronic documents, such as e-books for example, as soon as devices allow the human/machine interactions.

A step further, the important evolution which would highly benefit from recent advances in the computer vision community, is the possibility to handle any kind of text, such as hand-written texts or texts captured by a camera (similarly to *myReader2*). The crucial step is to have a robust document layout analysis,[27,119] in order to identify each component of a text, determine the logical structure of the paragraphs and perform optical character recognition (OCR). Much progress has been achieved but, to our knowledge, having a fully robust approach remains an open question. Such a pre-processing was tested in this context of low vision[223] but processing times and robustness remain an issue before integrating the OCR chain in a system (this is one weakness of *myReader2*). GPU-performance may help for processing times but new methods may also be required to improve robustness.

Finally, using an eye tracker system[223] with high frequency

performance is a challenging perspective. Such a system would allow to adapt the display in the enhanced view depending on the fixation. This could help to suppress the difficult mouse navigation task present in most of the reading aid tool.[129] Real time text deformations could also be useful to move the text on optimal health zones of the retina, but this is still a very difficult problem to validate such feature from a psychophysical point of view.

Acknowledgements

Advising

First, I would like to thank my advisers, Pierre Kornprobst and Guillaume S. Masson, for all the guidance they generously provided me before and during my PhD thesis, as well as for the preparation of my future. I greatly appreciated their permanent availability, contagious enthusiasm and open-mindedness.

Reviewing

I would like to express my gratitude to the examiners, Frédéric Alexandre, James A. Bednar and Ennio Mingolla, who kindly accepted to review this thesis, despite the important amount of work it requires. I also thank the other jury members, Olivier Faugeras, Jean Lorenceau and Gabriel Peyré, who accepted to be present during my defence.

I particularly thank Pascal Mamassian, Jean Lorenceau and Éric Castet for helpful discussions related to Chapter 6 and Chapter 7. Romain Veltz and Grégory Faye allowed me to understand the details of the neural fields framework and its possible applications in relation to our models. Pascal Mamassian also provided enriching thoughts related to the evaluation methodology of Chapter 9. Comments of the anonymous article reviewers were also of high interest.

Collaborations

The motion evaluation work presented in Chapter 9 arises from a collaboration with Jan D. Bouecke and Heiko Neumann. I am particularly thankful to Heiko Neumann for its scientific inspiration related to various aspects of my work and its help in various domains.

Numerical bifurcation analysis on the simplified neural fields model was done in collaboration with James Rankin who implemented the model using Trilinos (Chapter 8). Proofs of well posedness of the neural fields model were done in collaboration with Olivier Faugeras.

The low-vision research in Chapter B was conducted in collaboration with Éric Castet, Jean-Baptiste Bernard, Géraldine Faure and Anne-Catherine Scherlen. We

thank our colleagues from the Hospital La Timone in Marseille and specifically Fatiha Barouch as well as the patients who participated in the tests.

Family, friends & colleagues

Finally, I would express my gratitude towards my (growing!) family, my friends and my colleagues who supported (and distracted!) me during the last few years. I do love you all.

Funding

This research work has received funding from the Région PACA, the INRIA, the CNRS, the European Community (FACETS, Sixth Framework, No 025213, and SEARISE, Seventh Framework, No 215866), and the Agence Nationale de la Recherche (ANR, NATSTATS).

I finally want to thank Michele Bezzi and Fredrik Linaker, from *Accenture Technology Labs*, who accepted to act as referents of my work.

Publications

JOURNAL ARTICLES

E. Tlapale, G. S. Masson and P. Kornprobst. 'Modelling the dynamics of motion integration with a new luminance-gated diffusion mechanism'. In: *Vision Research* 50.17 (2010), pp. 1676–1692

J.D. Bouecke, E. Tlapale, P. Kornprobst and H. Neumann. 'Neural Mechanisms of Motion Detection, Integration, and Segregation: From Biology to Artificial Image Processing Systems'. In: *EURASIP, special issue on Biologically inspired signal processing: Analysis, algorithms, and applications* (2011). Article ID 781561

E. Tlapale, P. Kornprobst, G.S. Masson, J. Bouecke and H. Neumann. 'Bio-inspired motion estimation – From modelling to evaluation, can biology be a source of inspiration?' In: ***Submitted to*** *International Journal of Computer Vision* () (In review, see also INRIA research report 7447.)

PROCEEDINGS

E. Tlapale, P. Kornprobst, G.S. Masson and O. Faugeras. 'A neural field model for motion estimation'. In: ***To appear in*** *Proceedings of the Second conference "Mathematics and Image processing"*. 133–153

E. Tlapale, P. Kornprobst, J.D. Bouecke, H. Neumann and G.S. Masson. 'Evaluating motion estimation models from behavioural and psychophysical data'. In: *BIONETICS, special track on Bio-Inspired Machine Vision*. 2010

E. Tlapale, G. S. Masson and P. Kornprobst. 'Motion Integration Modulated by Form Information'. In: *Second plenary french conference on computational neuroscience*. 2008

J.-B. Bernard, E. Tlapale, G. Faure, E. Castet and P. Kornprobst. 'Navisio: Towards an integrated reading aid system for low vision patients'. In: *Proceedings of the Workshop on Computer Vision Applications for the Visually Impaired*. 2008

ABSTRACTS

E. Tlapale, G.S. Masson and P. Kornprobst. 'A dynamical neural model of motion integration'. In: *Vision Sciences Society*. 2010

J.-B. Bernard, E. Tlapale, A. Calabrese, E. Castet and P. Kornprobst. 'SolairePDF, un logiciel d'aide a la lecture de documents PDF pour les patients basse vision'. In: *Congres de la Societe Française d'Ophtalmologie*. 2008

E. Tlapale, G.S. Masson, T. Vieville and P. Kornprobst. 'Model of motion field diffusion controlled by form cues'. In: *European Conference on Visual Perception*. 2007

REPORTS

E. Tlapale, G.S. Masson, T. Vieville and P. Kornprobst. *Biological model of motion integration and segmentation based on form cues*. Tech. rep. RR-6293. INRIA, 2007

E. Tlapale, J.-B. Bernard, E. Castet and P. Kornprobst. *The SOLAIRE Project: A Gaze-Contingent System to Facilitate Reading for Patients with Scotomatas*. Tech. rep. RT-0326. INRIA, 2006

Bibliography

1 E.H. Adelson and J.R. Bergen. 'Spatiotemporal energy models for the perception of motion'. In: *Journal of the Optical Society of America A* 2 (1985), pp. 284–299.

2 E.H. Adelson and J.A. Movshon. 'Phenomenal coherence of moving visual patterns'. In: *Nature* 300.5892 (1982), pp. 523–525.

3 J.K. Aggarwal and N. Nandhakumar. 'On the gomputation of motion from sequences of images – A review'. In: *Proceedings of the IEEE* 76.8 (1988), pp. 917–935.

4 D.S. Alexiadis and G.D. Sergiadis. 'Narrow directional steerable filters in motion estimation'. In: *Computer Vision and Image Understanding* 110.2 (2008), pp. 192–211.

5 J.M. Allman and J.H. Kaas. 'A representation of the visual field in the caudal third of the middle temporal gyrus of the owl monkey (Aotus trivirgatus)'. In: *Brain Research* 31.1 (1971), pp. 85–105.

6 L. Alvarez, R. Deriche, T. Papadopoulo and J. Sanchez. 'Symmetrical dense optical flow estimation with occlusions detection'. In: *International Journal of Computer Vision (IJCV)* 75.3 (2007), pp. 371–385.

7 S.-I. Amari. 'Dynamics of pattern formation in lateral-inhibition type neural fields'. In: *Biological Cybernetics* 27.2 (1977), 77–87.

8 A. Angelucci and J. Bullier. 'Reaching beyond the classical receptive field of V1 neurons: horizontal or feedback axons?' In: *J Physiol Paris* 97.2–3 (2003), 141–154.

9 S. Anstis. 'Imperceptible intersections: The chopstick illusion'. In: *AI and the Eye*. 1990, p. 105.

10 K. Anton-Erxleben, V.M. Stephan and S. Treue. 'Attention reshapes center-surround receptive field structure in macaque cortical area MT'. In: *Cerebral Cortex* 19.10 (2009), p. 2466.

11 A. Arditi. 'Adjustable typography: an approach to enhancing low vision text accessibility'. In: *Ergonomics* 47.5 (2004), 469–482.

12 G. Aubert and P. Kornprobst. 'A mathematical study of the relaxed optical flow problem in the space BV'. In: *SIAM Journal on Mathematical Analysis* 30.6 (1999), pp. 1282–1308.

13 G. Aubert and P. Kornprobst. 'Can the nonlocal characterization of Sobolev spaces by Bourgain etal. be useful to solve variational problems?' In: *SIAM Journal on Numerical Analysis* 47.2 (2009), pp. 844–860.

14 W. Bair, J.R. Cavanaugh and A. Movshon. 'Time course and time–distance relationships for surround suppression in macaque V1 neurons'. In: *The Journal of Neuroscience* 23.20 (2003), pp. 7690–7701.

15 S. Baker, D. Scharstein, JP Lewis, S. Roth, M.J. Black and R. Szeliski. 'A database and evaluation methodology for optical flow'. In: *International Conference on Computer Vision, ICCV'07*. 2007, pp. 1–8.

16 J.F. Barraza and N.M. Grzywacz. 'Measurement of angular velocity in the perception of rotation'. In: *Vision research* 42.21 (2002), pp. 2457–2462.

17 J.L. Barron, D.J. Fleet and S.S. Beauchemin. 'Performance of Optical Flow Techniques'. In: *IJCV* 12.1 (1994), pp. 43–77.

18 F.V. Barthelemy, J. Fleuriet and G.S. Masson. 'Temporal dynamics of 2D motion integration for ocular following in macaque monkeys'. In: *Journal of Neurophysiology* 103.3 (2010), pp. 1275–1282.

19 F.V. Barthelemy, L.U. Perrinet, E. Castet and G.S. Masson. 'Dynamics of distributed 1D and 2D motion representations for short-latency ocular following'. In: *Vision Research* 48.4 (2008), pp. 501–522.

20 F.V. Barthelemy, I. Vanzetta and G.S. Masson. 'Behavioral Receptive Field for Ocular Following in Humans: Dynamics of Spatial Summation and Center-Surround Interactions'. In: *Journal of Neurophysiology* 95.6 (2006), 3712–3726.

21 A. Basole, L.E. White and D. Fitzpatrick. 'Mapping multiple features in the population response of visual cortex'. In: *Nature* 423.6943 (2003), pp. 986–990.

22 P. Bayerl and H. Neumann. 'Attention and figure-ground segregation in a model of motion perception'. In: *Journal of Vision* 5.8 (2005), pp. 659–659.

23 P. Bayerl and H. Neumann. 'Disambiguating Visual Motion by Form–Motion Interaction – a Computational Model'. In: *International Journal of Computer Vision* 72.1 (2007), pp. 27–45.

24 P. Bayerl and H. Neumann. 'Disambiguating Visual Motion Through Contextual Feedback Modulation'. In: *Neural Computation* 16.10 (2004), pp. 2041–2066.

25 C. Beck and H. Neumann. 'Interactions of motion and form in visual cortex – A neural model'. In: *Journal of Physiology – Paris* 104 (2010), 61–70.

26 P.J. Beckmann and G.E. Legge. 'Psychophysics of reading – XIV. The page navigation problem in using magnifiers'. In: *Vision research* 36.22 (1996), 3723–3733.

27 A. Belaïd. 'OCR: print'. In: *Survey of the state of the art in human language technology*. Cambridge University Press. 1997, 71–74.

28 S. Ben Hamed, J.R. Duhamel, F. Bremmer and W. Graf. 'Visual receptive field modulation in the lateral intraparietal area during attentive fixation and free gaze'. In: *Cerebral Cortex* 12.3 (2002), pp. 234–245.

29 J.-B. Bernard, E. Tlapale, A. Calabrese, E. Castet and P. Kornprobst. 'SolairePDF, un logiciel d'aide a la lecture de documents PDF pour les patients basse vision'. In: *Congres de la Societe Française d'Ophtalmologie*. 2008.

30 J.-B. Bernard, E. Tlapale, G. Faure, E. Castet and P. Kornprobst. 'Navisio: Towards an integrated reading aid system for low vision patients'. In: *Proceedings of the Workshop on Computer Vision Applications for the Visually Impaired*. 2008.

31 J.B. Bernard, A.-C. Scherlen and E. Castet. 'Page mode reading with simulated scotomas: a modest effect of interline spacing on reading speed'. In: *Vision research* 47.28 (2007), 3447–3459.

32 J. Berzhanskaya, S. Grossberg and E Mingolla. 'Laminar cortical dynamics of visual form and motion interactions during coherent object motion perception'. In: *Spatial Vision* 20.4 (2007), pp. 337–395.

33 M.J. Black and P. Rangarajan. 'On the unification of line processes, outlier rejection, and robust statistics with applications in early vision'. In: *IJCV* 19.1 (1996).

34 R.T. Born and D.C. Bradley. 'Structure and Function of Visual Area MT'. In: *Annu. Rev. Neurosci* 28 (2005), pp. 157–189.

35 R.T. Born, C.C. Pack, C. Ponce and S. Yi. 'Temporal Evolution of 2--Dimensional Direction Signals Used to Guide Eye Movements'. In: *Journal of Neurophysiology* 95 (2006), pp. 284–300.

36 W.H. Bosking, Y. Zhang, B. Schofield and D. Fitzpatrick. 'Orientation Selectivity and the Arrangement of Horizontal Connections in Tree Shrew Striate Cortex'. In: *The Journal of Neuroscience* 17.6 (1997), pp. 2112–2127.

37 J.D. Bouecke, E. Tlapale, P. Kornprobst and H. Neumann. 'Neural Mechanisms of Motion Detection, Integration, and Segregation: From Biology to Artificial Image Processing Systems'. In: *EURASIP, special issue on Biologically inspired signal processing: Analysis, algorithms, and applications* (2011). Article ID 781561.

38 A. Bowers, A.M.Y. Cheong and J.E. Lovie-Kitchin. 'Reading with optical magnifiers: page navigation strategies and difficulties'. In: *Optometry and vision science: official publication of the American Academy of Optometry* 84.1 (2007), p. 9.

39 A.R. Bowers, R.L. Woods and E. Peli. 'Preferred retinal locus and reading rate with four dynamic text presentation formats'. In: *Optometry & Vision Science* 81.3 (2004), p. 205.

40 L. Bowns. 'Evidence for a feature tracking explanation of why type II plaids move in the vector sum direction at short durations'. In: *Vision Research* 36.22 (1996), pp. 3685–3694.

41 D.C. Bradley and M.S. Goyal. 'Velocity computation in the primate visual system'. In: *Nature Reviews Neuroscience* 9.9 (2008), pp. 686–695.

42 G. Bradski. 'The Opencv Library'. In: *Dr. Dobb's Journal of Software Tools* (2000).

43 P.C. Bressloff, N.W. Bressloff and J.D. Cowan. 'Dynamical mechanism for sharp orientation tuning in an integrate-and-fire model of a cortical hypercolumn'. In: *Neural computation* 12.11 (2000), pp. 2473–2511.

44 K.H. Britten. 'The middle temporal area: motion processing and the link to perception'. In: *The visual neurosciences*. Vol. 2. 2003, pp. 1203–1216.

45 T. Brox and J. Malik. 'Large displacement optical flow: descriptor matching in variational motion estimation'. In: *pami* (2010).

46 A. Buades, B. Coll and J.-M Morel. 'Neighborhood filters and PDE's'. In: *Numerische Mathematik* 105.1 (2006), pp. 1–34.

47 J. Bullier. 'Integrated model of visual processing'. In: *Brain Res. Reviews* 36 (2001), pp. 96–107.

48 M.A. Bullimore and I.L. Bailey. 'Reading and eye movements in age-related maculopathy'. In: *Optometry & Vision Science* 72.2 (1995), p. 125.

49 C. Castellanos Sanchez, B. Girau and F. Alexandre. 'A connectionist approach for visual perception of motion'. In: *Brain Inspired Cognitive Systems (BICS 2004)* (2004), 1–7.

50 E. Castet, V. Charton and A. Dufour. 'The extrinsic/intrinsic classification of two-dimensional motion signals with barber-pole stimuli'. In: *Vision Research* 39.5 (1999), pp. 915–932.

51 E. Castet, J. Lorenceau, M. Shiffrar and C. Bonnet. 'Perceived speed of moving lines depends on orientation, length, speed and luminance'. In: *Vision Research* 33 (1993), pp. 1921–1921.

52 E. Castet and J. Zanker. 'Long-range interactions in the spatial integration of motion signals'. In: *Spatial Vision* 12.3 (1999), 287–307.

53 E. Cesmeli, D.T. Lindsey and D.L. Wang. 'An oscillatory correlation model of human motion perception'. In: *Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on*. Vol. 4. 2000.

54 E. Cesmeli and D.L. Wang. 'Motion segmentation based on motion/brightness integration and oscillatory correlation'. In: *Neural Networks, IEEE Transactions on* 11.4 (2002), pp. 935–947.

55 E. Cesmeli, D.L. Wang, D.T. Lindsey and J.T. Todd. 'Motion segmentation using temporal block matching and LEGION'. In: *Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Joint Conference on*. Vol. 3. IEEE. 2002, pp. 2069–2074.

56 J. Chey, S. Grossberg and E. Mingolla. 'Neural dynamics of motion processing and speed discrimination'. In: *Vision Research* 38 (1997), pp. 2769–2786.

57 S.T.L. Chung. 'The effect of letter spacing on reading speed in central and peripheral vision'. In: *Investigative Ophthalmology & Visual Science* 43.4 (2002), p. 1270.

58 S.T.L. Chung and J.S. Mansfield. 'Does reading mixed-polarity text improve reading speed in peripheral vision'. In: *Proceedings of the Vision* 99 (), 249–250.

59 M.M. Churchland and S.G. Lisberger. 'Shifts in the population response in the middle temporal visual area parallel perceptual and motor illusions produced by apparent motion'. In: *Journal of Neuroscience* 21.23 (2001), p. 9387.

60 P.S. Churchland and T.J. Sejnowski. *The computational brain*. MIT Press, 1992.

61 Stephen Coombes. 'Waves, bumps, and patterns in neural fields theories'. In: *Biological Cybernetics* 93.2 (2005), 91–108.

62 G.-H. Cottet and M. El Ayyadi. 'A Volterra type model for image processing'. In: *IEEE tip* 7.3 (1998).

63 M.D. Crossland, L.E. Culham and G.S. Rubin. 'Fixation stability and reading speed in patients with newly developed macular disease'. In: *Ophthalmic and Physiological Optics* 24.4 (2004), 327–333.

64 P. Dayan and L.F. Abbott. *Theoretical Neuroscience : Computational and Mathematical Modeling of Neural Systems*. MIT Press, 2001.

65 G.C. DeAngelis and W.T. Newsome. 'Organization of disparity-selective neurons in macaque area MT'. In: *Journal of Neuroscience* 19.4 (1999), p. 1398.

66 G. Deco and P. Roland. 'The role of multi-area interactions for the computation of apparent motion'. In: *NeuroImage* 51.3 (2010), pp. 1018–1026.

67 P. Degond and S. Mas-Gallic. 'The Weighted Particle Method for Convection-Diffusion Equations'. In: *Mathematics of Computation* 53.188 (1989), pp. 485–525.

68 K.G. Derpanis and J.M. Gryn. 'Three-dimensional nth derivative of Gaussian separable steerable filters'. In: *IEEE International Conference on Image Processing*. Vol. 3. 2005, pp. 553–556.

69 R. Dubner and S.M. Zeki. 'Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey.' In: *Brain Research* 35.2 (1971), p. 528.

70 R. Edwards. 'Approximation of neural network dynamics by reaction-diffusion equations'. In: *Mathematical Methods in the Applied Sciences* 19 (1996), pp. 651–677.

71 W. Enkelmann. 'Investigation of multigrid algorithms for the estimation of optical flow fields in image sequences'. In: *Workshop on Computer Vision, Graphics, and Image Processing (CVGIP)* 43 (1988), pp. 150–177.

72 B. Ermentrout. 'Neural networks as spatio-temporal pattern-forming systems'. In: *Reports on Progress in Physics* 61 (1998), 353–430.

73 M.-J. Escobar. 'Bio-Inspired Models for Motion Estimation and Analysis: Human action recognition and motion integration'. PhD thesis. Universite de Nice Sophia-Antipolis, 2009.

74 M.-J. Escobar, G. S. Masson and P. Kornprobst. 'A Simple Mechanism to Reproduce the Neural Solution of the Aperture Problem in Monkey Area MT'. In: *Neurocomp 2008*. 2008.

75 O. Faugeras, F. Grimbert and J.-J. Slotine. 'Abolute stability and complete synchronization in a class of neural fields models'. In: *SIAM App Math* 61.1 (2008), 205–250.

76 G. Faye and O. Faugeras. 'Some theoretical and numerical results for delayed neural field equations'. In: *Physica D* 239.9 (2009). Special issue on Mathematical Neuroscience., 561–578.

77 D.J. Felleman and D.C. Van Essen. 'Distributed hierarchical processing in the primate cerebral cortex'. In: *Cerebral cortex* 1.1 (1991), p. 1.

78 C.L. Fennema and W.B. Thompson. 'Velocity determination in scenes containing several moving objects'. In: *Computer Graphics and Image Processing* 9.4 (1979), pp. 301–315.

79 V.P. Ferrera and H.R. Wilson. 'Perceived direction of moving two-dimensional patterns'. In: *Vision Research* 30.2 (1990), pp. 273–287.

80 D.J. Field, A. Hayes and R.F. Hess. 'Contour integration by the human visual system: evidence for a local "association field"'. In: *Vision Research* 33.2 (1993), pp. 173–193.

81 G.D. Field, J.L. Gauthier, A. Sher, M. Greschner, T.A. Machado, L.H. Jepson, J. Shlens, D.E. Gunning, K. Mathieson, W. Dabrowski, Paninski L., Litke A.M. and Chichilnisky E.J. 'Functional connectivity in the retina at the resolution of photoreceptors'. In: *Nature* 467.7316 (2010), 673–677.

82 I. Fine, D.I.A. MacLeod and G.M. Boynton. 'Surface segmentation based on the luminance and color statistics of natural scenes'. In: *Journal of the Optical Society of America A* 20.7 (2003), pp. 1283–1291.

83 J. Fix, N. Rouger and F. Alexandre. '2010'. In: *Cognitive Computation* ().

84 R.A. Frazor and W.S. Geisler. 'Local luminance and contrast in natural images'. In: *Vision Research* 46.10 (2006), pp. 1585–1598.

85 W.T. Freeman and E.H. Adelson. 'The design and use of steerable filters'. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13.9 (1991), pp. 891–906.

86 W.S. Geisler. 'Visual perception and the statistical properties of natural scenes'. In: *Annu. Rev. Psychol.* 59 (2008), pp. 167–192.

87 W.S. Geisler, D.G. Albrecht and A.M. Crane. 'Responses of Neurons in Primary Visual Cortex to Transient Changes in Local Contrast and Luminance'. In: *Journal of Neuroscience* 27.19 (2007), p. 5063.

88 S. Georges, P. Series, Y. Fregnac and J. Lorenceau. 'Orientation dependent modulation of apparent speed: psychophysical evidence'. In: *Vision Research* 42.25 (2002), pp. 2757–2772.

89 M.A. Giese. *Dynamic Neural Field Theory for Motion Perception*. Springer, 1998.

90 P. Girard, P.A. Salin and J. Bullier. 'Response selectivity of neurons in area MT of the macaque monkey during reversible inactivation of area V1'. In: *Journal of Neurophysiology* 67.6 (1992), p. 1437.

91 T. Gollisch and M. Meister. 'Rapid neural coding in the retina with relative spike latencies'. In: *Science* 319.5866 (2008), p. 1108.

92 A. Gorea and J. Lorenceau. 'Directional performances with moving plaids: component-related and plaid-related processing modes coexist'. In: *Spatial vision* 5.4 (1991), pp. 231–252.

93 A. Grinvald, E. E. Lieke, R. D. Frostig and R. Hildesheim. 'Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex'. In: *Journal of Neuroscience* 14.5 (1994), 2545–2568.

94 S. Grossberg and E. Mingolla. 'Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading'. In: *Psychological review* 92.2 (1985), 173–211.

95 S. Grossberg, E. Mingolla and C. Pack. 'A neural model of motion processing and visual navigation by cortical area MST'. In: *Cerebral Cortex* 9.8 (1999), pp. 878–895.

96 S. Grossberg, E. Mingolla and L. Viswanathan. 'Neural dynamics of motion integration and segmentation within and across apertures'. In: *Vision Research* 41.19 (2001), pp. 2521–2553.

97 P.G. Grossenbacher and C.T. Lovelace. 'Mechanisms of synesthesia: Cognitive and physiological constraints'. In: *Trends in Cognitive Sciences* 5.1 (2001), pp. 36–41.

98 N.M. Grzywacz and A.L. Yuille. 'Theories for the visual perception of local velocity and coherent motion'. In: *Computational models of visual processing*. Chap. 16, pp. 231–252.

99 G.A Hahn, D. Penka, C. Gehrlich, A. Messias, M. Weismann, L. Hyvärinen, M. Leinonen, M. Feely, G. Rubin, C. Dauxerre and *others*. 'New standardised texts for assessing reading performance in four European languages'. In: *British Journal of Ophthalmology* 90.4 (2006), p. 480.

100 S. Harland, G.E. Legge and A. Luebker. 'Psychophysics of reading – XVII. Low-vision performance with four types of electronically magnified text'. In: *Optometry & Vision Science* 75.3 (1998), p. 183. ISSN: 1040-5488.

101 C.M. Harrison. 'Low-vision reading aids: reading as a pleasurable experience'. In: *Personal and Ubiquitous Computing* 8.3 (2004), 213–220.

102 D.J. Heeger. 'Optical Flow Using Spatiotemporal Filters'. In: *International Journal of Computer Vision (IJCV)* 1.4 (1988), pp. 279–302.

103 M.A. Heroux, R.A. Bartlett, V.E. Howle, R.J. Hoekstra, J.J. Hu, T.G. Kolda, R.B. Lehoucq, K.R. Long, R.P. Pawlowski, E.T. Phipps, A.G. Salinger, H.K. Thornquist, R.S. Tuminaro, J.M. Willenbring, A.Williams and K.S. Stanley. 'An overview of the Trilinos project'. In: *ACM Trans. Math. Softw.* 31.3 (2005), pp. 397–423.

104 R.F. Hess and D.J. Field. 'Integration of contours: new insights'. In: *Trends in Cognitive Sciences* 3.12 (1999), pp. 480–486.

105 R.F. Hess, A. Hayes and D.J. Field. 'Contour integration and cortical processing'. In: *Journal of Physiology-Paris* 97.2-3 (2003), pp. 105–119.

106 E. Hildreth. 'The measurement of visual motion'. PhD thesis. MIT, 1983.

107 E.C. Hildreth. 'The detection of intensity changes by computer and biological vision systems'. In: *Workshop on Computer Vision, Graphics, and Image Processing (CVGIP)* 22 (1983), pp. 1–27.

108 E.C. Hildreth and C. Koch. 'The analysis of visual motion: From computational theory to neuronal mechanisms'. In: *Annual Review of Neuroscience* 10.1 (1987), pp. 477–533.

109 A.L. Hodgkin and A.F. Huxley. 'A quantitative description of membrane current and its application to conduction and excitation in nerve'. In: *The Journal of physiology* 117.4 (1952), p. 500.

110 C.E. Holt. 'Does timing of axon outgrowth influence initial retinotectal topography in Xenopus?' In: *Journal of Neuroscience* 4.4 (1984), pp. 1130–1152.

111 B.K. Horn and B.G. Schunck. 'Determining Optical Flow'. In: *Artificial Intelligence* 17 (1981), pp. 185–203.

112 T. Hosoya, S.A. Baccus and M. Meister. 'Dynamic predictive coding by the retina'. In: *American Journal of Ophthalmology* 140.5 (2005), pp. 969–969.

113 X. Huang, T.D. Albright and G.R. Stoner. 'Adaptive Surround Modulation in Cortical Area MT'. In: *Neuron* 53 (2007), pp. 761–770.

114 X. Huang, T.D. Albright and G.R. Stoner. 'Stimulus Dependency and Mechanisms of Surround Modulation in Cortical Area MT'. In: *Journal of Neuroscience* 28.51 (2008).

115 D.H. Hubel and T.N. Wiesel. 'Receptive fields and functional architecture of monkey striate cortex'. In: *The Journal of Physiology* 195.1 (1968), p. 215.

116 D.H. Hubel and T.N. Wiesel. 'Receptive fields, binocular interaction and functional architecture in the cat visual cortex.' In: *J Physiol* 160 (1962), pp. 106–154.

117 J.M. Hupe and N. Rubin. 'The dynamics of bi-stable alternation in ambiguous motion displays: a fresh look at plaids'. In: *Vision Research* 43.5 (2003), pp. 531–548.

118 U.J. Ilg. 'The role of areas MT and MST in coding of visual motion underlying the execution of smooth pursuit'. In: *Vision Research* 48 (2008), 2062–2069.

119 A.K. Jain and B. Yu. 'Document representation and its application to page decomposition'. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 20.3 (2002), 294–308.

120 E. Jones, T. Oliphant, P. Peterson et al. *SciPy: Open source scientific tools for Python*. 2001–. URL: http://www.scipy.org/.

121 M.K. Kapadia, M. Ito, C.D. Gilbert and G. Westheimer. 'Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys'. In: *Neuron* 50 (1995), pp. 35–41.

122 M. Kinoshita and H. Komatsu. 'Neural Representation of the Luminance and Brightness of a Uniform Surface in the Macaque Primary Visual Cortex'. In: *Journal of Neurophysiology* 86.5 (2001), pp. 2559–2570.

123 H.-K. Ko, M. Poletti and M. Rucci. 'Microsaccades precisely relocate gaze in a high visual acuity task'. In: *Nature Neuroscience* 13 (2010), 1549–1553.

124 J.J. Koenderink and A.J. van Doorn. 'Representation of local geometry in the visual system'. In: *Biological Cybernetics* 55 (1987), 367–375.

125 T.L. Kooi. 'Local direction of edge motion causes and abolishes the barberpole illusion'. In: *Vision Research* 33.16 (1993), pp. 2347–2351.

126 R.J. Krauzlis. 'Eye movements'. In: *Fundamental Neuroscience*. 2008, 775–792.

127 R.J. Krauzlis. 'The control of voluntary eye movements: new perspectives'. In: *The Neuroscientist* (2005), 124–137.

128 B. Krekelberg, R.J.A. van Wezel and T.D. Albright. 'Interactions between speed and contrast tuning in the middle temporal area: Implications for the neural code for speed'. In: *Journal of Neuroscience* 26.35 (2006), p. 8988.

129 M. Kumar, A. Paepcke and T. Winograd. 'EyePoint: practical pointing and selection using gaze and keyboard'. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM. 2007, pp. 421–430.

130 V.A.F. Lamme and P.R. Roelfsema. 'The distinct modes of vision offered by feedforward and recurrent proce ssing'. In: *Trends in Neurosciences* 23.11 (2000), pp. 571–579.

131 T. Lee and D. Mumford. 'Hierarchical Bayesian inference in the visual cortex'. In: *Journal of the Optical Society of America A* 20.7 (2003).

132 T.S. Lee and M. Nguyen. 'Dynamics of subjective contour formation in the early visual cortex'. In: *Proceedings of the National Academy of Sciences* 98.4 (2001), p. 1907.

133 G.E. Legge. *Psychophysics of reading in normal and low vision*. CRC, 2006. ISBN: 0805843280.

134 G.E. Legge, D.G. Pelli, G.S. Rubin and M.M. Schleske. 'Psychophysics of reading – 1. Normal vision'. In: *Vision Research* 25.2 (1985), 239–252.

135 G.E. Legge, J.A. Ross, A. Luebker and J.M. Lamay. 'Psychophysics of reading – VIII. The Minnesota low-vision reading test'. In: *Optometry & Vision Science* 66.12 (1989), p. 843.

136 G.E. Legge and G.S. Rubin. 'Psychophysics of reading – IV. Wavelength effects in normal and low vision'. In: *Journal of the Optical Society of America A* 3.1 (1986), 40–51.

137 G.E. Legge, G.S. Rubin, D.G. Pelli and M.M. Schleske. 'Psychophysics of reading – II. Low vision'. In: *Vision research* 25.2 (1985), 253–265.

138 P. Lennie and J.A. Movshon. 'Coding of color and form in the geniculostriate visual pathway (invited review)'. In: *Journal of the Optical Society of America A* 22.10 (2005), pp. 2013–2033.

139 L. Liden and E. Mingolla. 'Monocular occlusion cues alter the influence of terminator motion in the barber pole phenomenon'. In: *Vision Research* 38.24 (1998), 3883–3898.

140 S.G. Lisberger, E.J. Morris and L. Tychsen. 'Visual motion processing and sensory-motor integration for smooth pursuit eye movements'. In: *Annual Review of Neuroscience* 10.1 (1987), pp. 97–129.

141 M.S. Livingstone and B.R. Conway. 'Contrast affects speed tuning, space-time slant, and receptive-field organization of simple cells in macaque V1'. In: *Journal of neurophysiology* 97.1 (2007), p. 849.

142 J. Lorenceau and D. Alais. 'Form constraints in motion binding'. In: *Nature Neuroscience* 4 (2001), pp. 745–751.

143 J. Lorenceau and M. Shiffrar. 'The influence of terminators on motion integration across space'. In: *Vision Research* 32.2 (1992), pp. 263–273.

144 J. Lorenceau, M. Shiffrar, N. Wells and E. Castet. 'Different motion sensitive units are involved in recovering the direction of moving lines'. In: *Vision Research* 33 (1993), pp. 1207–1207.

145 J. Lorenceau and L. Zago. 'Cooperative and competitive spatial interactions in motion integration'. In: *Visual Neuroscience* 16.04 (1999), pp. 755–770.

146 B. Lucas and T. Kanade. 'An Iterative Image Registration Technique with an Application to Stereo Vision'. In: *International Joint Conference on Artificial Intelligence*. 1981, pp. 674–679.

147 G. Löffler and H.S. Orbach. 'Computing feature motion without feature detectors: A model for terminator motion without end-stopped cells'. In: *Vision Research* 39.4 (1998), pp. 859–871.

148 V. Mante and M. Carandini. 'Mapping of stimulus energy in primary visual cortex'. In: *Journal of neurophysiology* 94.1 (2005), p. 788.

149 V. Mante, R.A. Frazor, V. Bonin, W.S. Geisler and M. Carandini. 'Independence of luminance and contrast in natural scenes and in the early visual system'. In: *Nature Neuroscience* 8.12 (2005), pp. 1690–1697.

150 G. Masson, D. Mestre and L. Stone. 'Speed tuning of motion segmentation and discrimination'. In: *Vision Research* 39 (1999), pp. 4297–4308.

151 G.S. Masson and E. Castet. 'Parallel motion processing for the initiation of short-latency ocular following in humans'. In: *The journal of neuroscience* 22.12 (2002), pp. 5147–5163.

152 G.S. Masson and U.J. Ilg, eds. *Dynamics of Visual Motion Processing*. 1st ed. Neuronal, Behavioral, and Computational Approaches. Springer Verlag, 2010.

153 G.S. Masson, Y. Rybarczyk, E. Castet and D.R. Mestre. 'Temporal dynamics of motion integration for the initiation of tracking eye movements at ultra-short latencies'. In: *Visual Neuroscience* 17.05 (2000), pp. 753–767.

154 G.S. Masson and L.S. Stone. 'From following edges to pursuing objects'. In: *Journal of neurophysiology* 88.5 (2002), pp. 2869–2873.

155 JH Maunsell and JR Gibson. 'Visual response latencies in striate cortex of the macaque monkey'. In: *Journal of Neurophysiology* 68.4 (1992), pp. 1332–1344.

156 R. Miikkulainen, J.A. Bednar, Y. Choe and J. Sirosh. *Computational maps in the visual cortex*. Citeseer, 2005. ISBN: 0387220240.

157 A. Montagnini, P. Mamassian, L. Perrinet, E. Castet and G.S. Masson. 'Bayesian modeling of dynamic motion integration'. In: *Journal of Physiology-Paris* 101.1-3 (2007), pp. 64–77.

158 A. Montagnini, M. Spering and G.S. Masson. 'Predicting 2D Target Velocity Cannot Help 2D Motion Integration for Smooth Pursuit Initiation'. In: *J Neurophysiol* 96 (2006), pp. 3545–3550.

159 J. Moran and R. Desimone. 'Selective attention gates visual processing in the extrastriate cortex'. In: *Science* 229.4715 (1985), p. 782.

160 M.J. Morgan and E. Castet. 'The aperture problem in stereopsis'. In: *Vision Research* 37.19 (1997), pp. 2737–2744.

161 JA Movshon, EH Adelson, MS Gizzi and WT Newsome. 'The analysis of moving patterns'. In: *Pattern recognition mechanisms* (1986), pp. 117–151.

162 J.A. Movshon, E.H. Adelson, M.S. Gizzi and W.T. Newsome. 'The analysis of visual moving patterns'. In: *Pattern recognition mechanisms* (1986), pp. 117–151.

163 H.H. Nagel and W. Enkelmann. 'An investigation of smoothness constraint for the estimation of displacement vector fields from image sequences'. In: 8 (1986), pp. 565–593.

164 K. Nakayama and G.H. Silverman. 'The aperture problem. II. Spatial integration of velocity information along contours'. In: *Vision Research* 28.6 (1988), pp. 747–753.

165 J.I. Nelson and B.J. Frost. 'Intracortical facilitation among co-oriented, co-axially aligned simple cells in cat striate cortex'. In: *Experimental Brain Research* 61.1 (1985), pp. 54–61.

166 Tal Nir, Alfred M. Bruckstein and Ron Kimmel. 'Over-Parameterized Variational Optical Flow'. In: *International Journal of Computer Vision* 76.2 (2008), pp. 205–216.

167 S.J. Nowlan and T.J. Sejnowski. 'Filter selection model for motion segmentation and velocity integration'. In: *Journal of the Optical Society of America A* 11.12 (1994), pp. 3177–3199.

168 M. Otte and H.H. Nagel. 'Optical Flow Estimation: Advances and Comparisons'. In: *Proceedings of the 3rd European Conference on Computer Vision*. Ed. by Jan-Olof Eklundh. Vol. 800. 1994, pp. 51–70.

169 C.C. Pack, V.K. Berezovskii and R.T. Born. 'Dynamic properties of neurons in cortical area MT in alert and anaesthetized macaque monkeys'. In: *Nature* 414.6866 (2001), pp. 905–908.

170 C.C. Pack and R.T. Born. 'Integration of motion signals over regions of uniform luminance by MT neurons in the alert macaque'. In: *Journal of Vision* 2.7 (2002), 412a.

171 C.C. Pack and R.T. Born. 'Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain'. In: *Nature* 409 (2001), pp. 1040–1042.

172 C.C. Pack, A.J. Gartland and R.T. Born. 'Integration of Contour and Terminator Signals in Visual Area MT of Alert Macaque'. In: *The Journal of Neuroscience* 24.13 (2004), 3268–3280.

173 C.C. Pack, S. Grossberg and E. Mingolla. 'A neural model of smooth pursuit control and motion perception by cortical area MT'. In: *Journal of Cognitive Neuroscience* 13.1 (2001), 102–120.

174 C.C. Pack, J.N. Hunter and R.T. Born. 'Contrast Dependence of Suppressive Influences in Cortical Area MT of Alert Macaque'. In: *J Neurophysiol* 93 (2005), pp. 1809–1815.

175 C.C. Pack, J.N. Hunter and R.T. Born. 'Contrast Dependence of Suppressive Influences in Cortical Area MT of Alert Macaque'. In: *Journal of Neurophysiology* 93.3 (2005), 1809–1815.

176 C.C. Pack, M.S. Livingstone, K.R. Duffy and R.T. Born. 'End-Stopping and the Aperture Problem: Two-Dimensional Motion Signals in Macaque V1'. In: *Neuron* 39.4 (2003), 671–680.

177 T.V. Papathomas, A. Gorea and B. Julesz. 'Two carriers for motion perception: color and luminance'. In: *Vision Research* 31.11 (1991), pp. 1883–1892.

178 M.A. Paradiso, S. Blau, X. Huang, S.P. MacEvoy, A.F. Rossi and G. Shalev. 'Lightness, filling-in, and the fundamental role of context in visual perception'. In: *Progress in Brain Research* (2006), 109–123.

179 D.R. Patzwahl and S. Treue. 'Combining spatial and feature-based attention within the receptive field of MT neurons'. In: *Vision Research* (2009).

180 P. Pearson, P. Seakins, D. Lee, N. Tucker and D. Sherwood. 'myReader: an Auto-Reader for Low Vision Users'. In: *7th IAPR Workshop on Document Analysis Systems*. 2006.

181 D.G. Pelli, M. Palomares and N.J. Majaj. 'Crowding is unlike ordinary masking: Distinguishing feature detection and integration'. In: *Journal of Vision* 4.12 (2004), p. 12.

182 D.G. Pelli, K.A. Tillman, J. Freeman, M. Su, T.D. Berger and N.J. Majaj. 'Crowding and eccentricity determine reading rate'. In: *Journal of vision* 7.2 (2007).

183 X. Peng and D.C. Van Essen. 'Peaked Encoding of Relative Luminance in Macaque Areas V1 and V2'. In: *Journal of Neurophysiology* 93.3 (2005), pp. 1620–1632.

184 D.J. Pinto, J.C. Brumberg, D.J. Simons, G.B. Ermentrout and R. Traub. 'A quantitative population model of whisker barrels: re-examining the Wilson-Cowan equations'. In: *Journal of Computational Neuroscience* 3.3 (1996), 247–264.

185 D.J. Plummer and V.S. Ramachandran. 'Perception of transparency in stationary and moving images'. In: *Spatial vision* 7.2 (1993), pp. 113–123.

186 U. Polat and D. Sagi. 'Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments'. In: *Vision Research* 33 (1993), pp. 993–993.

187 U. Polat and D. Sagi. 'The architecture of perceptual spatial interactions'. In: *Vision Research* 34.1 (1994), pp. 73–78.

188 R.P. Power and B. Moulden. 'Spatial gating effects on judged motion of gratings in apertures'. In: *Perception* 21 (1992), pp. 449–449.

189 N.J. Priebe and S.G. Lisberger. 'Estimating target speed from the population response in visual area MT'. In: *Journal of Neuroscience* 24.8 (2004), p. 1907.

190 N.J. Priebe, S.G. Lisberger and A.J. Movshon. 'Tuning for Spatiotemporal Frequency and Speed in Directionally Selective Neurons of Macaque Striate Cortex'. In: *The Journal of Neuroscience* 26.11 (2006), pp. 2941–2950.

191 V.S. Ramachandran and R.L. Gregory. 'Does colour provide an input to human motion perception?' In: *Nature* 275.5675 (1978), p. 55.

192 K. Rayner. 'Eye movements in reading and information processing'. In: *Psychological Bulletin* 85.3 (1978), 618–660.

193 W. Reichardt. 'Autokorrelationsauswertung als Funktionsprinzip des Zentralnervensystems'. In: *Zeitschrift fur Naturforschung* 12 (1957), pp. 447–457.

194 H.R. Rodman and T.D. Albright. 'Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT)'. In: *Experimental Brain Research* 75 (1989), 53–64.

195 H.R. Rodman, C.G. Gross and T.D. Albright. 'Afferent basis of visual response properties in area MT of the macaque. I. Effects of striate cortex removal'. In: *Journal of Neuroscience* 9.6 (1989), p. 2033.

196 A.F. Rossi, C.D. Rittenhouse and M.A. Paradiso. 'The Representation of Brightness in Primary Visual Cortex'. In: *Science* 273.5278 (1996), p. 1104.

197 G.S. Rubin and G.E. Legge. 'Psychophysics of reading – VI. The role of contrast in low vision'. In: *Vision Research* 29.1 (1989), 79–91.

198 E. Russell-Minda, J.W. Jutai, J.G. Strong, K.A. Campbell, D. Gold, L. Pretty and L. Wilmot. 'The legibility of typefaces for readers with low vision: A research review'. In: *Journal of Visual Impairment and Blindness* 101.7 (2007), p. 402.

199 N.C. Rust, V. Mante, E.P. Simoncelli and J.A. Movshon. 'How MT cells analyze the motion of visual patterns'. In: *Nature Neuroscience* 9 (2006), pp. 1421–1431.

200 A. Salgado and J. Sanchez. 'Temporal Constraints in Large Optical Flow Estimation'. In: *Computer Aided Systems Theory – EUROCAST 2007*. Vol. 4739. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2007, pp. 709–716.

201 Michael P. Sceniak, Dario L. Ringach, Michael J. Hawken and Robert Shapley. 'Contrast's effect on spatial summation by macaque V1 neurons'. In: *Nature Neuroscience* 2.8 (1999), pp. 733–739.

202 P. Series, S. Georges, J. Lorenceau and Y. Fregnac. 'A network view of the structure of center/surround modulations of V1 receptive field properties in visual and cortical spaces'. In: *Neurocomputing* 38 (2001), pp. 881–888.

203 P. Series, S. Georges, J. Lorenceau and Y. Fregnac. 'Orientation dependent modulation of apparent speed: a model based on the dynamics of feedforward and horizontal connectivity in v1 cortex'. In: *Vision Research* 42.25 (2002), pp. 2781–2798.

204 M. Shiffrar, X. Li and J. Lorenceau. 'Motion integration across differing image features'. In: *Vision Research* 35.15 (1995), pp. 2137–2146.

205 M. Shiffrar and J. Lorenceau. 'Increased motion linking across edges with decreased luminance contrast, edge width and duration'. In: *Vision Research* 36.14 (1996), pp. 2061–2067.

206 S. Shimojo, G.H. Silverman and K. Nakayama. 'Occlusion and the solution to the aperture problem for motion.' In: *Vision Research* 29.5 (1989), pp. 619–26.

207 A.M. Sillito, J. Cudeiro and H.E. Jones. 'Always returning: feedback and sensory processing in visual cortex and thalamus'. In: *TRENDS in Neurosciences* 29.6 (2006), pp. 307–316.

208 E.P. Simoncelli and D.J. Heeger. 'A Model of Neuronal Responses in Visual Area MT'. In: *Vision Research* 38 (1998), pp. 743–761.

209 E.P. Simoncelli and B.A. Olshausen. 'Natural image statistics and neural representation'. In: *Annual Review of Neuroscience* 24.1 (2001), pp. 1193–1216.

210 L.L. Sloan and S.J. Ryan Jr. 'Reading Aids for the Partially Sighted: A Nontechnical Explanation of Basic Optical Principles'. In: *International ophthalmology clinics* 11.1 (1971), p. 19.

211 M. Smith, N. Majaj and A. Movshon. 'Dynamics of motion signaling by neurons in macaque area MT'. In: *Nature Neuroscience* 8.2 (2005), pp. 220–228.

212 R. J. Snowden, S. Treue, R. G. Erickson and R. A. Andersen. 'The response of area MT and V1 neurons to transparent motion'. In: *The Journal of Neuroscience* 11.9 (1991), pp. 2768–2785.

213 S. Soto-Faraco, A. Kingstone and C. Spence. 'Cross-modal dynamic capture: congruency effects in the perception of motion across sensory modalities'. In: *Journal of Experimental Psychology* 30.2 (2004), pp. 330–345.

214 RW Sperry. 'Chemoaffinity in the orderly growth of nerve fiber patterns and connections'. In: *Proceedings of the National Academy of Sciences of the United States of America* 50.4 (1963), pp. 703–710.

215 R.W. Sperry. 'Effects of 180 degree rotation of the retinal field on visuomotor coordination'. In: *Journal of experimental zoology* 92 (1943), pp. 263–279.

216 D.D. Stettler, A. Das, J. Bennett and C.D. Gilbert. 'Lateral connectivity and contextual interactions in macaque primary visual cortex'. In: *Neuron* 36.4 (2002), pp. 739–750.

217 C. Stiller and J. Konrad. 'Estimating motion in image sequences'. In: *Signal Processing Magazine, IEEE* 16.4 (1999), pp. 70–91.

218 G.R. Stoner, T.D Albright and V.S. Ramachandran. 'Transparency and coherence in human motion perception'. In: *Nature* 344.6262 (1990).

219 C. Strecha and L. Van Gool. 'Motion-Stereo Integration for Depth Estimation'. In: *ECCV*. Vol. 2. 2002, pp. 170–185.

220 D. Sun, S. Roth, T.U. Darmstadt and M.J. Black. 'Secrets of Optical Flow Estimation and Their Principles'. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2010).

221 T. Tani, I. Yokoi, M. Ito, S. Tanaka and H. Komatsu. 'Functional Organization of the Cat Visual Cortex in Relation to the Representation of a Uniform Surface'. In: *Journal of Neurophysiology* 89.2 (2003), pp. 1112–1125.

222 W. Taouali, N. Rouger and F. Alexandre. 'Saccades generation: from the Visual Input to the Superior Colliculus'. In: *International Conference on Neural Computation*. 2010.

223 E. Tlapale, J.-B. Bernard, E. Castet and P. Kornprobst. *The SOLAIRE Project: A Gaze-Contingent System to Facilitate Reading for Patients with Scotomatas*. Tech. rep. RT-0326. INRIA, 2006.

224 E. Tlapale, P. Kornprobst, J.D. Bouecke, H. Neumann and G.S. Masson. 'Evaluating motion estimation models from behavioural and psychophysical data'. In: *BIONETICS, special track on Bio-Inspired Machine Vision*. 2010.

225 E. Tlapale, P. Kornprobst, G.S. Masson, J. Bouecke and H. Neumann. 'Bio-inspired motion estimation – From modelling to evaluation, can biology be a source of inspiration?' In: ***Submitted to*** *International Journal of Computer Vision* ().

226 E. Tlapale, P. Kornprobst, G.S. Masson and O. Faugeras. 'A neural field model for motion estimation'. In: ***To appear in*** *Proceedings of the Second conference "Mathematics and Image processing"*. 133–153.

227 E. Tlapale, G. S. Masson and P. Kornprobst. 'Modelling the dynamics of motion integration with a new luminance-gated diffusion mechanism'. In: *Vision Research* 50.17 (2010), pp. 1676–1692.

228 E. Tlapale, G. S. Masson and P. Kornprobst. 'Motion Integration Modulated by Form Information'. In: *Second plenary french conference on computational neuroscience*. 2008.

229 E. Tlapale, G.S. Masson and P. Kornprobst. 'A dynamical neural model of motion integration'. In: *Vision Sciences Society*. 2010.

230 E. Tlapale, G.S. Masson, T. Vieville and P. Kornprobst. *Biological model of motion integration and segmentation based on form cues*. Tech. rep. RR-6293. INRIA, 2007.

231 E. Tlapale, G.S. Masson, T. Vieville and P. Kornprobst. 'Model of motion field diffusion controlled by form cues'. In: *European Conference on Visual Perception*. 2007.

232 J. Tombran-Tink, C.J. Barnstable, J.F. Rizzo and Ebooks Corporation. *Visual Prosthesis and Ophthalmic Devices: New Hope in Sight*. Humana Press, 2007.

233 C. Torres Huitzil, B. Girau and C. Castellanos Sanchez. 'On-chip visual perception of motion: a bio-inspired connectionist model on FPGA'. In: *Neural Networks* 18.5-6 (2005), 557–565.

234 F.G. Tricomi. *Integral Equations*. Dover, 1985.

235 L.M. Vaina, M. Lemay, D.C. Bienfang, A.Y. Choi and K. Nakayama. 'Intact "biological motion" and "structure from motion" perception in a patient with impaired motion mechanisms: A case study'. In: *Visual Neuroscience* 5.04 (1990), pp. 353–369.

236 J.P.H. Van Santen and G. Sperling. 'Elaborated Reichardt detectors'. In: *Journal of the Optical Society of America A* 2.2 (1985), pp. 300–320.

237 R. Veltz and O. Faugeras. *Illusions in the Ring Model of visual orientation selectivity*. Tech. rep. Submitted to Plos Comp Bio. arXiv, 2010.

238 R. Veltz and O. Faugeras. 'Local/global analysis of the stationary solutions of some neural field equations'. In: *SIAM J. Applied Dynamical Systems* 9.3 (2010), pp. 954–998.

239 T. Vieville, S. Chemla and P. Kornprobst. 'How do high-level specifications of the brain relate to variational approaches?' In: *Journal of Physiology - Paris* 101.1-3 (2007), pp. 118–135.

240 K. Wall and P.E. Danielsson. 'A fast sequential method for polygonal approximation of digitized curves'. In: *Workshop on Computer Vision, Graphics, and Image Processing (CVGIP)* 28 (1984), pp. 220–227.

241 J.M. Wallace, L.S. Stone and G.S. Masson. 'Object Motion Computation for the Initiation of Smooth Pursuit Eye Movements in Humans'. In: *Journal of Neurophysiology* 93.4 (2005), pp. 2279–2293.

242 H. Wallach. *On perception*. Crown, 1976.

243 H. Wallach. 'Über visuell wahrgenommene Bewegungsrichtung'. In: *Psychological Research* 20.1 (1935), pp. 325–380.

244 H. Wallach, A. Weisz and P.A. Adams. 'Circles and derived figures in rotation'. In: *The American Journal of Psychology* 69.1 (1956), pp. 48–59.

245 A.B. Watson and A.J. Ahumada. 'Model of human visual-motion sensing'. In: *Journal of the Optical Society of America A* 2.2 (1985), pp. 322–342.

246 J. Weickert. 'Coherence-Enhancing Diffusion of Colour Images'. In: *7th National Symposium on Pattern Recognition and Image Analysis*. 1997.

247 J. Weickert and C. Schnorr. 'Variational optic flow computation with a spatio-temporal smoothness constraint'. In: *Journal of Mathematical Imaging and Vision* 14.3 (2001), pp. 245–255.

248 Y. Weiss and E.H. Adelson. 'Adventures with gelatinous ellipses – constraints on models of human motion analysis'. In: *Perception* 29 (2000), pp. 543–566.

249 Y. Weiss and E.H. Adelson. 'Perceptually organized EM: A framework for motion segmentation that combines information about form and motion'. In: *International Conference on Computer Vision*. 1995.

250 Y. Weiss and D.J. Fleet. 'Velocity likelihoods in biological and machine vision'. In: *Probabilistic Models of the Brain: Perception and Neural Function*. MIT Press, 2001, pp. 81–100.

251 Y. Weiss, E.P. Simoncelli and E.H. Adelson. 'Motion illusions as optimal percepts'. In: *nature neuroscience* 5.6 (2002), pp. 598–604.

252 M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers and H. Bischof. 'Anisotropic Huber-L1 Optical Flow'. In: *Proceedings of the British Machine Vision Conference (BMVC)*. 2009.

253 H.R. Wilson and J.D. Cowan. 'A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue'. In: *Biological Cybernetics* 13.2 (1973), 55–80.

254 H.R. Wilson and J.D. Cowan. 'Excitatory and inhibitory interactions in localized populations of model neurons'. In: *Biophys. J.* 12 (1972), 1–24.

255 H.R. Wilson, V.P. Ferrera and C. Yo. 'A psychophysically motivated model for two-dimensional motion perception.' In: *Visual Neuroscience* 9.1 (1992), pp. 79–97.

256 A. Wohrer and P. Kornprobst. 'Virtual Retina: A biological retina model and simulator, with contrast gain control'. In: *Journal of computational neuroscience* 26.2 (2009), 219–249. ISSN: 0929-5313.

257 T. Womelsdorf, K. Anton-Erxleben, F. Pieper and S. Treue. 'Dynamic shifts of visual receptive fields in cortical area MT by spatial attention'. In: *Nature Neuroscience* 9.9 (2006), pp. 1156–1160.

258 D.K. Xiao, S. Raiguel, V. Marcar and G.A. Orban. 'The spatial distribution of the antagonistic surround of MT/V5 neurons'. In: *Cerebral Cortex* 7.7 (1997), p. 662.

259 J. Xiao, H. Cheng, H. Sawhney, C. Rao and M. Isnardi. 'Bilateral Filtering-Based Optical Flow Estimation with Occlusion Detection'. In: *ECCV*. 2006.

260 C. Yo and H.R. Wilson. 'Perceived direction of moving two-dimensional patterns depends on duration, contrast and eccentricity.' In: *Vision Research* 32.1 (1992), pp. 135–47.

261 B. Yu and L. Zhang. 'Pulse-coupled neural networks for contour and motion matchings'. In: *Neural Networks, IEEE Transactions on* 15.5 (2004), pp. 1186–1201.

262 A.L. Yuille and N.M. Grzywacz. 'A winner-take-all mechanism based on presynaptic inhibition feedback'. In: *Neural Computation* 1.3 (1989), pp. 334–347.

263 S.M. Zeki. 'A century of cerebral achromatopsia'. In: *Brain* 113.6 (1990), p. 1721.

264 S.M. Zeki. 'Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey'. In: *The Journal of Physiology* 236.3 (1974), p. 549.

265 X. Zhang and A.A. Minai. 'Temporally sequenced intelligent block-matching and motion-segmentation using locally coupled networks'. In: *Neural Networks, IEEE Transactions on* 15.5 (2004), pp. 1202–1214.

266 J. Zihl, D. von Cramon and N. Mai. 'Selective disturbance of movement vision after bilateral brain damage'. In: *Brain* 106.2 (1983), p. 313.

Contents