Biological model of motion integration and segmentation based on form cues

Émilien Tlapale, Guillaume S. Masson, Pierre Kornprobst

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Biological model of motion integration and segmentation based on form cues

Émilien Tlapale — Guillaume S. Masson — Pierre Kornprobst

N° 6293

Septembre 2007

Thème BIO

ISSN 0249-6399 ISRN INRIA/RR--6293--FR+ENG

Biological model of motion integration and segmentation based on form cues

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

Thème BIO — Systèmes biologiques
Projet Odyssée

Rapport de recherche n° 6293 — Septembre 2007 — 44 pages

Abstract: Active vision is an essential part of every biological organism possessing an eye system: posture, eye movements, visual research, ... All require a motion percept to operate. At the basis of active vision lays the ability to calculate movements of objects in the scene, at least on a sufficient level to react correctly.

In this report we present a model of motion integration and segmentation in the first visual cortex areas. Specifically we modeled the first two cortex areas involved in motion processing in the primate: V1 and MT. To be able to process motion correctly a visual system also need to deal with form information. We investigate how form cues coming from the ventral pathway can be used by the V1/MT dorsal pathway to solve some perception problems.

By using a recurrent dynamical system between the V1 and MT layers we are able to find out psychophysical results such as motion integration and center-surround effects due to the feedback connections, or end-of-line and 2D features detectors thanks to the shunting inhibition. We propose to modulate this system by a form information coming from the ventral stream and are thus able to explain asymmetric center-surround effects as well as motion segmentation and segregation between extrinsic and intrinsic junctions.

Key-words: motion estimation, optical flow, bio-inspired vision, MT area, feedbacks, center-surround integration, shunting inhibition

[*] INRIA Sophia Antipolis, Projet Odyssée, 2004 Route des Lucioles, 06902 Sophia Antipolis
[†] INCM, Equipe DyVA UMR 6193 CNRS-Université de la Méditerranée 31, chemin Joseph Aiguier 13402 Marseille cedex

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : +33 4 92 38 77 77 — Télécopie : +33 4 92 38 77 65

Modèle biologique d'intégration et de segmentation du mouvement basé sur la forme

Résumé : La vision active est une partie essentielle de tout organisme biologique possédant un système visuel: la posture, les mouvement des yeux, la recherche visuelle, ... tous nécessitent un percept de mouvement pour fonctionner. À la base de la vision active, le calcul des mouvements des objets de la scène est priordial. Au moins à un niveau suffisant pour pouvoir réagir correctement.

Dans ce rapport nous présentons un modèle d'intégration et de segmentation du mouvement dans les premières aires du cortex visuel. En particulier nous modélisons deux zones impliquées dans le traitement du mouvement chez le macaque: V1 et MT. Pour être capable de traiter correctement le mouvement, le système visuel a aussi besoin de manipuler des informations de forme. Nous étudions comment les informations de forme venant du système ventral peuvent être utilisées dans les zones V1/MT du système dorsal pour résoudre des problèmes de perception.

En utilisant un système dynamique avec des rétroactions entre les aires V1 et MT nous sommes capable de retrouver des résultats psychophysiques tels que l'intégration du mouvement et les effets de voisinage dus à la rétroaction, ou bien des détecteurs de fin de ligne et d'information univoque grace à l'inhibition latérale. Nous proposons une modulation du système par une information de forme venant du système ventral mettant ainsi en évidence des effets de voisinage asymétrique de même qu'une segmentation du mouvement et des discriminations entre jonctions extrinsèques et intrinsèques.

Mots-clés : estimation du mouvement, flot optique, vision bio-inspirée, aire MT, rétroaction, intégration centre-périphérie, inhibition latérale

Contents

Introduction

Humans and animals have a vision system strongly correlated with either reflex or voluntary actions. Be it either to research preys highly camouflaged, to avoid collisions while running at high speed or to smash a tennis ball. Different kinds of information can be extracted from the scene depending on both the eye structure and the processing neural system. In this report we are focusing on motion processing in the primate early visual system, particularly on the first steps leading to a perception of motion.

Psychophysics and physiology are two main approaches for the understanding of the visual system. Psychophysics studies the subjective human perception when presenting a given stimulus. Physiology focuses on physical response coming from electrophysiology and brain imaging.

Combining bottom-up data from physiology and top-down data from psychophysics, models try to associate a functional role to the physical areas involved in motion processing, and to connect them in a coherent system in order to simulate the correct subjective perception.

In this report we focus on how motion information transfered by interconnected neurons from the eyes to inside the brain is processed by different spatially segregated group of neurons each having some specific functional role. Indeed the visual cortex can be divided into areas by both physiological criterion as done by Brodmann and functional criterion revealed by brain imaging.

In Chapter 1 we review the relevant psychophysical and physiological data used in motion processing. Some of the existing models for motion processing are described in Chapter 2. We present our model of motion integration and segmentation controlled by form cues in Chapter 3 together with its physiological justifications. Results for various classical psychophysical stimuli are then presented in Chapter 4.

This work was presented as a poster presentation at ECVP2007, the European Conference on Visual Perception, in August 2007. The poster as well as several animations of both stimuli and model results are available at http://emilien.tlapale.com.

Chapter 1

What do we know from the visual system?

1.1 Psychophysics, the indirect way to investigate the visual system

Psychophysics experiments reveal the perception of subjects when presenting visual stimuli. In Chapters 2 and 3 we will see how this ever increasing data can be used to construct a model explaining a maximum number of experiments. We thereafter present stimuli relevant to motion processing.

1.1.1 The aperture problem

A local motion is a motion which can be calculated using a very limited neighborhood at each position. As we will see in section 1.2 local motions are the first motion element extracted by our visual cortex, on them the computation will later be done. Local motions also serves as the basis for many computer vision algorithms dealing with motion.

Local motions however are ambiguous and lead to the well known *aperture problem* (Wallach, 1935) which has also been the point of departure for motion processing in computer vision. As an illustration, in Figure 1.1 the stimulus contain a simple bar translating behind an aperture. We cannot figure out without seeing the bar end of lines which is the real bar motion. Motion from the unambiguous end of lines needs to be integrated towards the center of the bar, so we need a more global view of the scene. When motion like this bar only contains a one-dimensional information our perception generally select the slowest motion, an effect called the *prior on low velocities*.

The delay of propagation from unambiguous to ambiguous regions is not instantaneous, it is dependant of the distance between the two regions. Moreover, Masson et al. (2000)

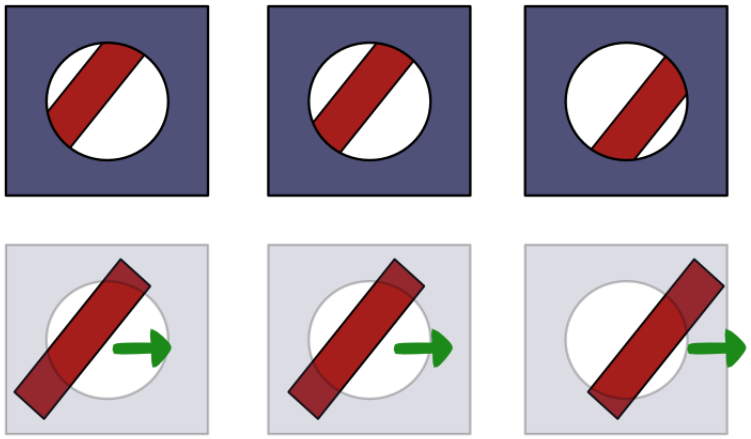


Figure 1.1: The aperture problem: when 2D information such as end of lines are not available, motion is ambiguous. The slowest motion is then perceived. When the aperture is removed, the end of lines give information.

show that there is always an initial delay for the integration, independently of the distance to the ambiguous region. The reason for this delay is still unclear, it might come from a feedback loop or any other motion integration process.

1.1.2 Motion integration

Other classical stimuli where motion ambiguity should be resolved are the *Barber Poles*. They can be seen as bars equally spaced moving in the same direction but through an invisible rectangular aperture. Figure 1.2 shows a single frame of such kind of stimulus for two different aperture: a square and a rectangle. On the square we mostly see a motion perpendicular to the bar but other motions such as rightward or downward in our example are possible. The problem is that even if the end of lines are available they do not give a useful information since they came from the invisible aperture.

As long as the aperture is asymmetric we see a motion collinear to the longer dimension (Wallach, 1935). Wallach (1976) proposed that the greater number of terminators on the longer borders explain the perceived motion. Other models propose the unambiguous information to propagate on the ambiguous line (Hildreth and Koch, 1987).

Figure 1.2: The Barber Pole stimulus: A moving pattern made of isospaced bars is seen through an invisible aperture. On a square with can see different motions but the direction perpendicular to the bars is preferred. With an asymmetric aperture such as a rectangle, the perceived motion is collinear to the longest dimension.

1.1.3 Extrinsic junctions

Yet another problem of local motions are the *extrinsic junctions*. A junction is a spatial feature that gives unambiguous 2D motion such as intersections or end of lines. We need to segregate between intrinsic and extrinsic junctions:

- Intrinsic junctions give a real motion which can be used on the global motion processing.
- Extrinsic junctions are only an artefact which must be ignored.

As an example of extrinsic junctions let us consider two bars moving in opposite directions (see Figure 1.3 (a)–(c)). Seen locally their intersection seems to move upward, however this information is not globally consistent.

The Figure 1.3 also describe another stimulus where a discrimination between intrinsic and extrinsic junction change the perception. In the first case there are no occluders or the occluder is invisible as in a Barber Pole, in the second case we display the occluder and the end of lines become extrinsic. Our model presented in Chapter 3 is able to process some extrinsic junctions like those in the Chopstick illusions to get results similar to our percept (see Section 4.2 for the results).

Shimojo et al. (1989) show that extrinsic corners have a suppressed response in comparison to intrinsic ones. This is confirmed at the neuronal level by Pack et al. (2004).

1.1.4 Multistability

Various multistable stimuli exists such as plaids or projections of rotating cylinders (Bradley et al., 1998). However there is no data available on multistability at the neuronal level and it is mostly psychophysically studied (Hupe and Rubin, 2003; Gepshtein and Kubovy, 2005). The switching process is mostly ignored and only the statistically dominant percept is studied. Multistability does not feet in a pure feedforward model of motion processing.

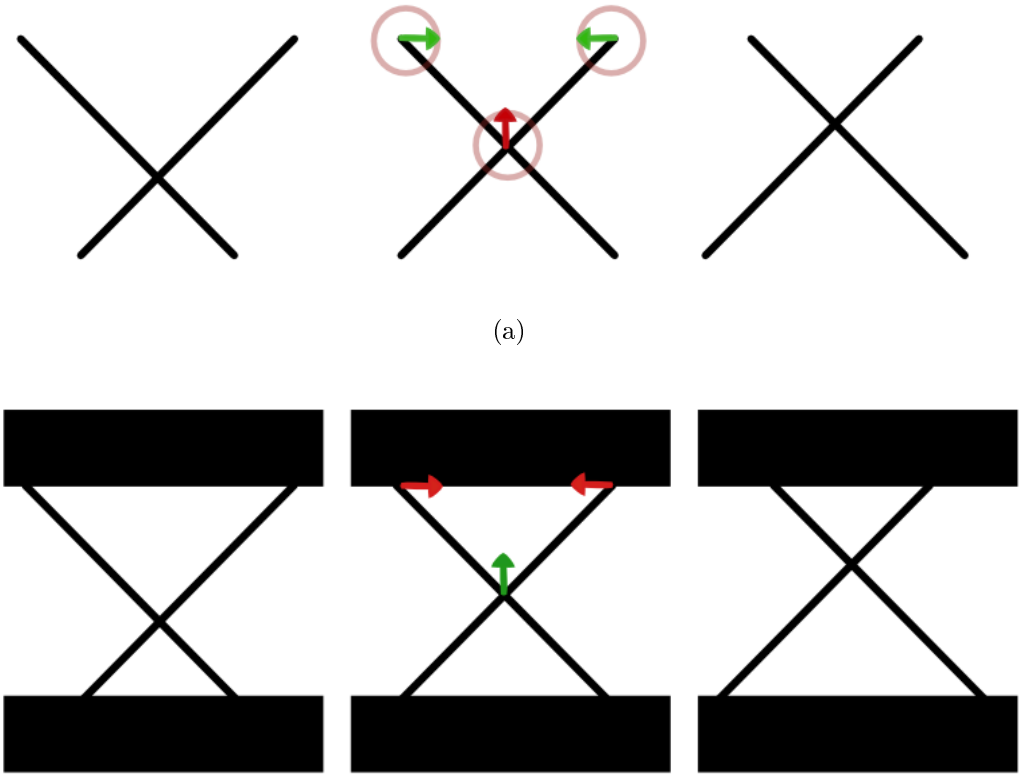


Figure 1.3: Chopstick illusion: (a) When two crossing bars are moving in opposite directions the ends of lines give a valid motion locally. However heir intersection is an extrinsic junction which locally produce an upward incoherent motion. Extrinsic junctions must be ignored. (b) As soon as visible occluders are added we perceive a single coherent motion in the direction of the motion defined by the bars intersection.

The crossed Barber Pole displayed in Figure 1.4 is a classical multistable stimulus. Main perceived motions are horizontal and vertical as well as orthogonal to the occluded plaid.

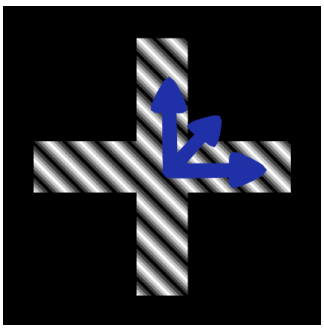


Figure 1.4: Crossed Barber Pole: Three main stable percept exists on a crossed Barber Pole, i.e. a plaid pattern moving through a crossed aperture: horizontal, vertical or a perpendicularly to the plaid lines.

1.2 Physiological data, a brain architecture for motion

1.2.1 Visual pathways

The two classical main pathways in the brain are the *form* and the *motion* streams. The former is concerned with static objects features, the later with moving objects and motion integration. Both receive their input from the eye through the LGN and the V1 cortex area (see Figure 1.5.

V1 and V2 compute both the form and the motion and so are part of the two pathways. Processing motion follows V3, MT and MST. In MT motion is integrated and segmented. Form information goes through V2 followed by V4 and IT.

As an example the general connectivity of the MT inputs is described in the Figure 1.6. The majority of input to MT come from V1 layer, particularly from the layer 4B (Born and Bradley, 2005). The V1 neurons connected to MT are highly direction-selective (Movshon and Newsome, 1996).

MT is essentially a parvocellular layer and should be color blind but MT cells can be drove with ecoluminance stimuli. The magnocellular path to MT is confirmed by Nassi et al. (2006). When the V1 area is removed, MT still remains visually responsive (Rodman et al., 1989; Girard et al., 1992), the visual information come from the superior colliculus in this case.

In this work we focus on the V1 and MT areas for motion processing. V3 and V3A are also likely to play a role in the coherent motion (Braddick et al., 2001) but their role is still unclear.

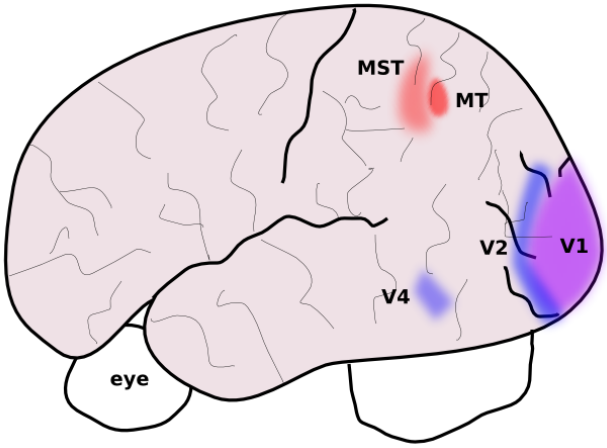


Figure 1.5: The two main pathways in visual motion processing: the form stream, in blue, analyze static features, the motion stream, in red, is specialized on dynamics processing.

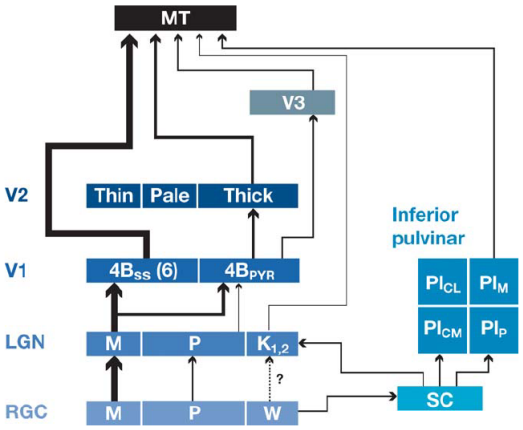


Figure 1.6: Major routes to MT. Line thickness is proportional to the magnitude of the inputs. Omitted weaker inputs include V3A, VP, PIP and many subcortical inputs. Backward connections are ignored on the schema. From Born and Bradley (2005)

1.2.2 Feedback to V1

In the previous section we described feedforward connections going farther from the eyes. But in various brain regions feedback play an important role, particularly in the visual system (Sillito et al., 2006). MT for instance connects back to V1 and V2 (Rockland and Knutson, 2000).

Angelucci and Bullier (2003) demonstrate how the feedback connections from MT to V1 are very important for the center/surround mechanism on wide areas. Indeed the feedback connections are much faster than the horizontal one (0.1-0.2 m/s versus 2-6 m/s according to Grinvald et al. (1994)). The MT surface from which a V1 neuron receive feedback data is larger than the MT surface to which it send data (Angelucci et al., 2002). See Figure 1.7 for the importance of feedback on V1.

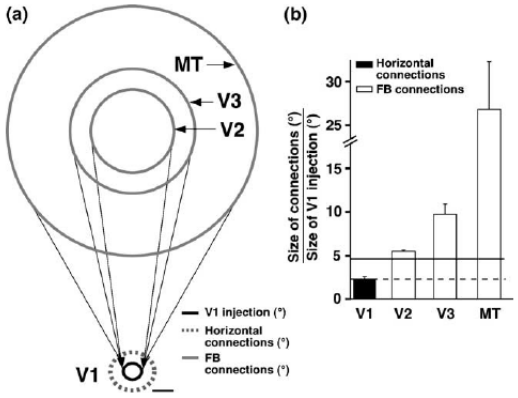


Figure 1.7: Extent of feedback (FB) connections to V1 in visual field coordinates. (a) Visual field map of a CTB injection site in V1 (black circle) and resulting labeled fields of cells of origin of feedback connections in the lower layers of extrastriate cortical areas V2, V3 and MT. The visuotopic extent of V1 horizontal connections (dashed gray circle) arising from the same V1 injection is also shown for comparison. Scale bar: 2.5°. (b) Population means of the relative visuotopic extent (gray circle diameter/black circle diameter) of connections (open bars) in the lower layers of areas V2, V3 or MT, and of V1 horizontal connections (filled bar). From Angelucci and Bullier (2003).

1.2.3 Spatial and tuning properties

The *receptive field* of a neuron can be defined as the visual field area where a stimulus induce a response. Most cortical areas dealing with visual information are retinotopically organized: Given two neurons in a small neighborhood, their receptive field should also be close one to the other.

The different cortical areas forming the visual system are hierarchically organized: upper layers, those farther from the eyes, having wider receptive fields. Typical values for the

receptive fields are: 1 degree for V1 cells or 6-10 degrees for MT cells. However those values strongly depends of the eccentricity, particularly for high layers such as MT (see Figure 1.8). The MST area having an even steeper shape. There are more cells processing the information near the fovea and they have a narrower receptive field (Mestre et al., 2001).

It should be observed that the receptive field size is dynamic and also depends on contrast: at low contrast the receptive field is wider Angelucci and Bullier (2003); Pack et al. (2005); Huang et al. (2007). This effect is a way to increase the signal-to-noise ratio.

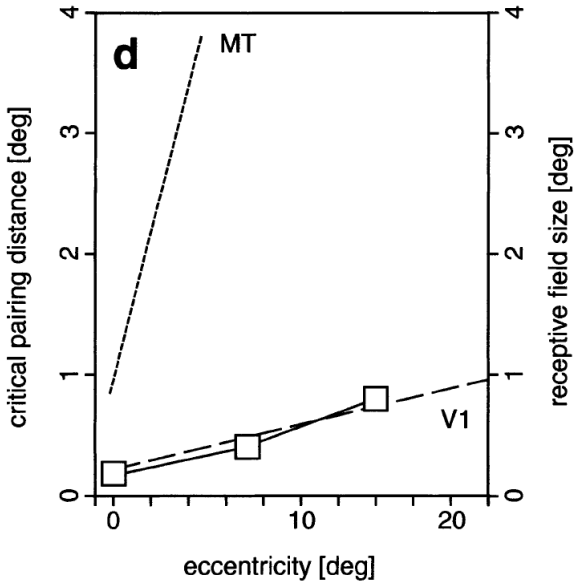


Figure 1.8: The *dashed lines* are the relation between the receptive field sizes of V1 and MT neurons and eccentricity. The *solid line with squares* displays the average critical pairing distance as a function of eccentricity. The paired dots suppression seems to occurs at the V1 level. From Mestre et al. (2001).

Speed and Direction Selectivity

Both V1 and MT have direction tuned neurons but MT neurons specifically show a strong inhibition in the anti-preferred direction: Directionally selective responses is 30% in V1 and 92% in MT according to Snowden et al. (1991).

The neurons found on those areas also are speed selective although nobody knows exactly how it works. For instance, in Masson et al. (1999) it is shown that V1 prefers low speed (there are no V1 cells detecting high speed) but MT prefers high speeds (see also Figure 1.9). The question is: From where does MT take its informations from? The V1 cells encoding 1 degree per frame do not directly connect to the MT cells encoding 1 degree per frame.

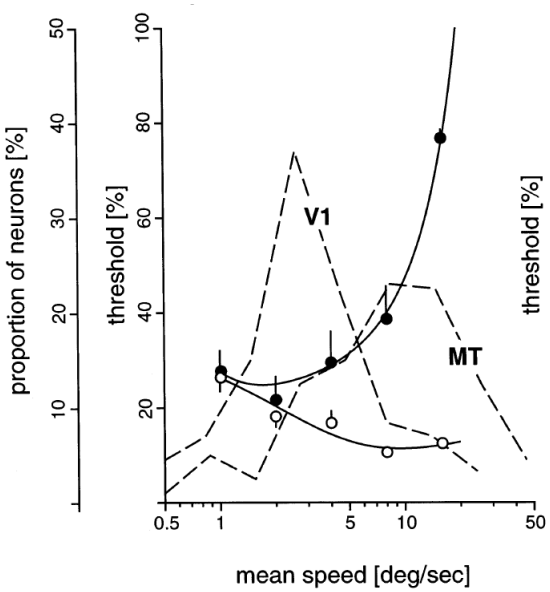


Figure 1.9: The *dashed lines* indicates the speed sensitivities of the V1 and MT neuron populations. V1 are low velocities tuned, MT neurons cover a broader velocity space. MT cells take their inputs from V1, so how could they encode higher velocities? From Masson et al. (1999)

Binocular Disparity

Two-thirds of the cells found in MT are highly disparity selective. These cells also show a strongly columnar organization (DeAngelis and Newsome, 1999) (see also Figure 1.10). Disparity tuned neurons will respond better when they are stimulated by a motion on a plane which could be at the fixation depth but also in nearer of farer.

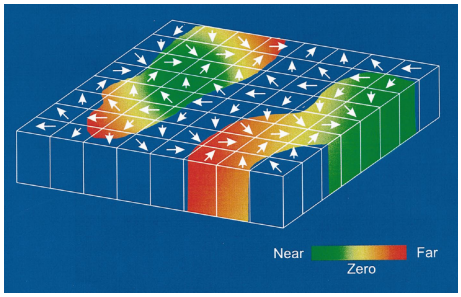


Figure 1.10: Schematic summary of the functional architecture of MT, with regard to binocular disparity and direction of motion. MT cells show a strongly columnar organization. From DeAngelis and Newsome (1999)

Center-Surround Interaction

A single MT neuron can dynamically have an excitatory on inhibitory surround depending upon the contrast Pack et al. (2005) or upon the presence of an ambiguity Huang et al. (2007). When there is an ambiguity such as in in a square contour with only 1D motion information, the MT neurons integrates. When we replace the edge by dots moving in the same directions the neuron is suppressive. Thus both diffusion and inhibition could be done in MT. Responses of a single MT neuron at high and low contrast is displayed in Figure 1.11 (a).

V1 neurons also show a center-surround suppression, the greatest in layer 4B, the weakest in layer 6 Sceniak et al. (2001). The feedback from MT is probably at the origin of the behaviour (see Section 1.2.2).

Receptive Field Structure

Although it is often assumed that the center and the surround can be represented by two concentric circles this is generally not the case. Half of the neurons have an *asymmetric* surround with one preferred region. The rest consists of *bilateral symmetric* neurons with a pair of surrounding regions on opposite sides and *symmetric* surrounds Xiao et al. (1995, 1997). See also Figure 1.11 (c).

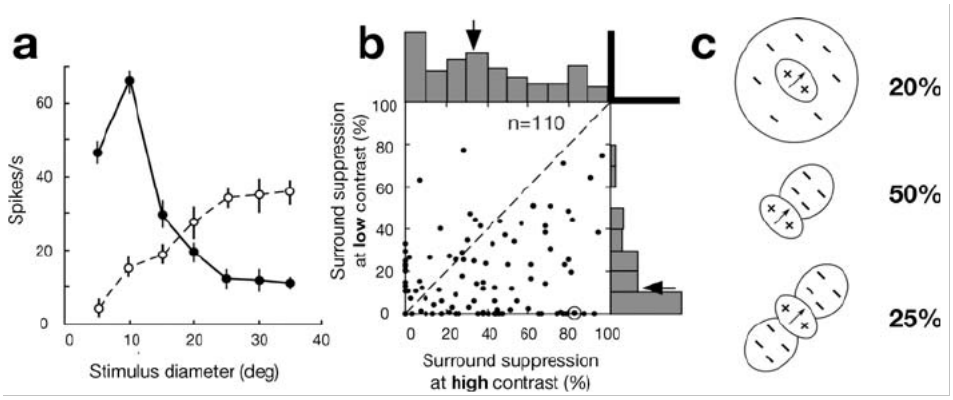


Figure 1.11: Center-surround interactions in MT. (a) An MT neurons shows an inhibitive surround for a high contrast stimulus (filled circles). The same neuron has an excitatory surround for a low contrast stimulus (empty circles). (b) MT neurons showing the strength of surround suppression measured at both high and low contrast. Surround suppression was quantified as the percent reduction in response between the largest dot patch and the stimulus eliciting the maximal response. (c) Asymmetries in the spatial organization of the suppressive surround. Neurons whose receptive fields have circularly symmetric surrounds are postulated to underlie figure-ground segregation. The first- and second-order derivatives can be used to determine surface tilt and curvature. From Born and Bradley (2005)

In the V1 area Angelucci and Bullier (2003) show how the center-surround mechanism come both from the horizontal V1 to V1 connection for the small surround and from a feedback from upper areas (e.g. MT) to V1 for larger surround.

The source of MT surrounds remains unclear and it is hard to explain it using imaging: MST is located deeper in the cortex and in fMRI one cannot distinguish between MT and MST. A suggestion we made in our model is that the center-surround might come from V1 (see Section 3.2.3).

MT Input: Subunits

In the paired dots experiment a dot is moving inside the excitatory area of an MT cell receptive field. We then add another dot next to the first moving in the opposite direction. If the two points are in a sufficiently distanced, more than 0.4°, the observer see a transparent motion. When the two points are close to each other, less than 0.4°, we see no motion.

The experiment of paired dots moving in opposite direction shows no transparency supposing a strong inhibition. This suppression seems not to occur in the V1 area since they do not have a very strong direction selectivity. On the contrary in MT there is a very strong suppression in the anti-preferred direction but MT receptive fields are much wider. The concept of subunits has been proposed to solve this problem, no one really know what they are but their origin seems to be at the V1 level (see Figure 1.8).

1.2.4 Form modulation

Using V1 local velocities, MT do integration and segmentation in order to solve the aperture problem and to give a correct de-noised estimation of the image velocities. Segmentation is required not to diffuse the velocities on two different objects for instance.

In Pack et al. (2004) it is shown that MT neurons are more responsive to unambiguous information, such as corners velocities than to ambiguous 1D signals. It is also showed that unambiguous but false signal, such as the velocities at the junctions of two objects moving in different directions is suppressed, thus involving a form-control.

Chapter 2

Existing models

Several models try to explain the percept given by psychophysical data based in a more or less rigorous way on the underlying brain physiology.

2.1 Simoncelli and Heeger's model

Simoncelli and Heeger (1998)'s model is a feedforward system combining V1 complex cells responses to create an MT cell (see Figure 2.1 (b)). Figure 2.1 (a) describe the V1 complex cell model. Note that the half-squaring rectification followed by the divisive normalization, i.e. the shunting inhibition, is similar to Nowlan and Sejnowski (1993) soft-maximization. This shunting inhibition and the whole linear/non-linear process is used in Bayerl and Neumann (2007)'s model as well as in our model (see Section 3.2.2).

2.2 Bayerl and Neumann's model

2.2.1 Overview

Bayerl and Neumann's model (Bayerl, 2005; Bayerl and Neumann, 2007) feed a two layers V1/MT recurrent system by a Reichardt-like local motion detector.

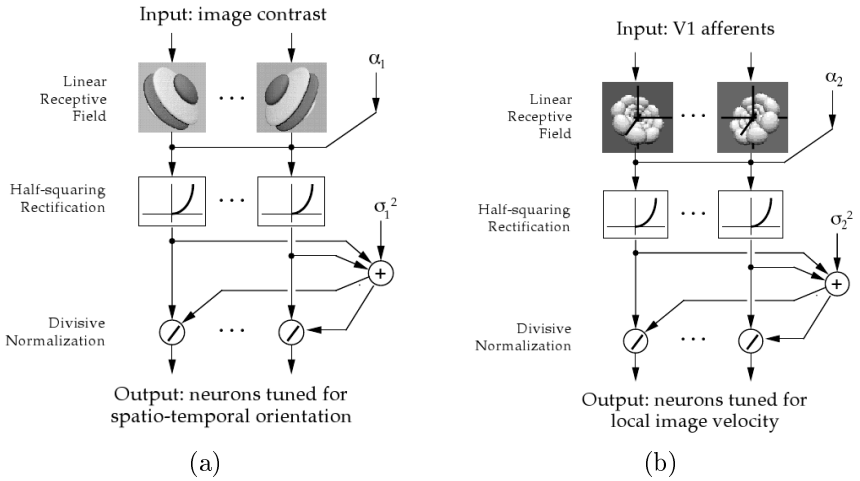


Figure 2.1: Simoncelli and Heeger's model. The V1 and MT cells follow the linear/non-linear pattern with a non-linearity introduced by the shunting inhibition.

$$p_0(t,x,v) = c_3(t,x,v) \tag{2.1}$$
$$p_1(t,x,v) = p_0\,(1 + C_f\,p_5(t,x,v)) \tag{2.2}$$
$$p_2(t,x,v) = (p_1(t,x,v))^{\beta} \overset{v}{*} G_{\sigma_{v1}} \tag{2.3}$$
$$p_3(t,x,v) = \frac{p_2(t,x,v)}{\sum_{u\in V} p_2(t,x,u)} \tag{2.4}$$
$$p_4(t,x,v) = (p_3(t,x,v))^{\beta} \overset{x}{*} G_{\sigma_{x2}} \overset{v}{*} G_{\sigma_{v2}} \tag{2.5}$$
$$p_5(t,x,v) = \frac{p_4(t,x,v)}{\sum_{u\in V} p_4(t,x,u)} \tag{2.6}$$

where the motion detector p_0 is defined as follow:

$$c_1(t,x,\alpha) = \frac{I_{t,x} * \partial_\alpha^2 G_\sigma}{\sum_\beta |I_{t,x} * \partial_\beta^2 G_\sigma| * G_\sigma + \epsilon} \tag{2.7}$$

$$c_2^+(t,x,v) = (\sum_\alpha c_1(t,x,\alpha) \cdot c_1(t+1,x+v,\alpha) * G_\sigma \tag{2.8}$$

$$c_2^-(t,x,v) = (\sum_\alpha c_1(t+1,x,\alpha) \cdot c_1(t,x+v,\alpha) * G_\sigma \tag{2.9}$$

$$c_3(t,x,v) = \frac{[c_2^+(t,x,v)]_+ - 0.5[c_2^-(t,x,v)]_+}{1 + [c_2^-(t,x,v)]_+} \tag{2.10}$$

where ∂_α represent the directional derivative in direction α and $[x]+$ is zero if x is negative, x otherwise.

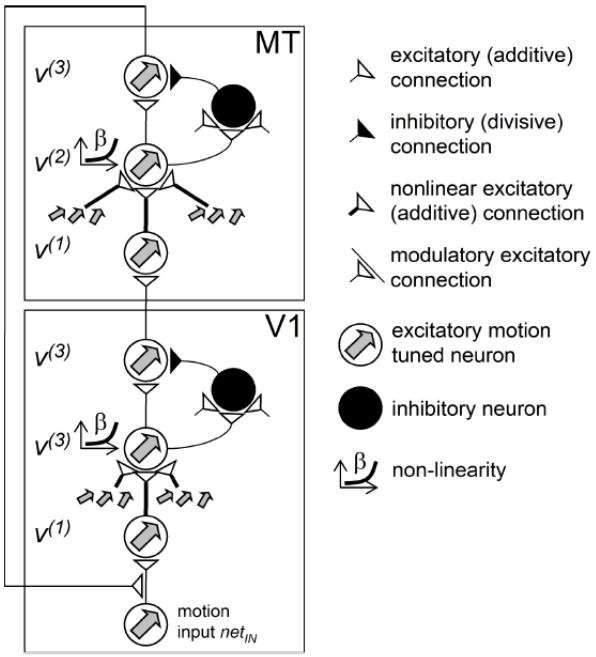


Figure 2.2: Overview of Bayerl's model. Layers V1 and MT have a similar three equations structures. $v^{(1)}$ combines the input and the feedback, $v^{(2)}$ is the feedforward integration, $v^{(3)}$ do a lateral shunting inhibition. Differences are the receptive field sizes and the input/feedback connections.

2.6 implements a shunting inhibition, similar to the *soft-maximization* found in Nowlan and Sejnowski (1994):

$$\mathcal{M}(p(t,x,v)) = \frac{e^{a\,p(t,x,v)}}{\sum_{u\in V} e^{a\,p(t,x,u)}} \tag{2.11}$$

The main difference being the use of an exponential by Nowlan and Sejnowski (1994) and a $\beta = 2$ exponent in 2.5 by Bayerl as in Simoncelli and Heeger (1998) described in the Section 2.1.

2.2.2 Discussion

With no a priori on low velocities described in Section 1.1.1, the model is not able to give the correct percept on 1D-only information where the aperture problem is present. The prior on small velocities is also necessary to explain type II plaids and the rhombus stimuli (Weiss, 1998).

→ in our model (see 3.2.1 we add this prior which can be seen as a prior in a Bayesian model.

A proposition to discriminate between intrinsic and extrinsic junction (see Section 1.1.3) is to ignore them. As an extension to their model, Bayerl and Neumann propose to build X- and T-junctions detectors and use them to ignore the information at the spatial position where they are found. This is similar to the three colors detectors of Weiss (1998).

Such a form modulation despite being able to give a correct percept for the chopstick illusion without occluders will however give the same percept of two motions for the crossing bars with occluders (see Section 1.1.3).

→ In Section 3.2.4 we propose another kind of form feature which can be integrated in the motion stream in order to get motions similar to our percept.

To get a usable optical flow, the authors propose a weighted summation of the response to all velocities at each point. This simple mechanism to get a perceived motion is not found in MT and destroy any transparent motion analysis on the data. For the transparency the authors propose to duplicate the whole system with motion opponency between the two.

→ As we will see in 4.5 the classical shunting inhibition of Nowlan and Sejnowski (1993) or Bayerl (2005) is not able to process transparent motion. We suggest an isotropic shunting inhibition to handle transparent motion.

Chapter 3

Proposed model

In our model we propose a functional role for the early motion processing cortical areas V1 and MT presented in Section 1.2. Our model should be able to solve the aperture problem defined in Section 1.1.1 by using correct motion integration presented in Section 1.1.2 coming from a V1/MT recurrent system. Moreover a form modulation coming from the ventral pathway (see Sections 1.2.1 and 1.2.4) will explain stimuli like the Chopstick illusions (see Section 1.1.3) or the stimulus of Huang et al. (2007).

3.1 Overview of the proposed model

We model eye input as a discretized sequence of gray-scaled images. Discretized space domain can be seen as a consequence of finite rods and cones. Although the question is far from being answered, some hints (Crick and Koch, 2003; Sacks, 2004) may even point toward a discretized consciousness.

In our model we want to describe the modulation of form into motion. We implemented a V1/MT structure together with form information from a V2 area. At each point we describe the possible motions.

Since V1 and MT neurons are retinotopically organized and each tuned to a specific velocity speed and direction, it seems reasonable to model the result of our model at every time step using a volume as shown in Figure 3.1 (b).

Let us comment Figure 3.2 which describes the general architecture. We use two kinds of V1 populations: *component cells* which react only to local motion and have the aperture problem and *pattern cells* which receive feedforward input from V1 component cells but also feedback input from MT. Although initial experiments only show pattern cells in V1 (Movshon and Newsome, 1996), recent papers by Tinsley et al. (2003) and Guo et al. (2004) both show the existence of V1 component cells which may be hidden by the anesthesia.

The MT area receives its input from the V1 pattern cells but at a coarse scale as implied by MT larger receptive fields. This integration into MT is modulated by form information

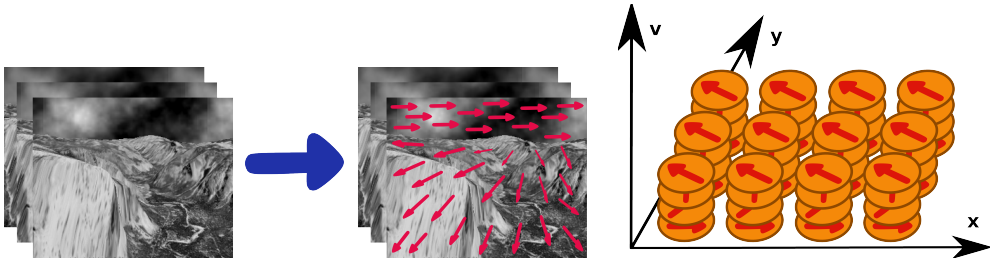


Figure 3.1: Model's goals. (a) Giving a sequence of gray-scale image our goal is to compute the velocity at every point. (b) Since multiple motion can exist at a single point and because neurons are velocity tuned we model our result as a cube at each time frame.

coming from a V2 area which is part of the form stream. The feedback from MT into V1 is a simple bijection, we did not model the existing divergence seen in Section 1.2.2 because it is not a useful feature in our model.

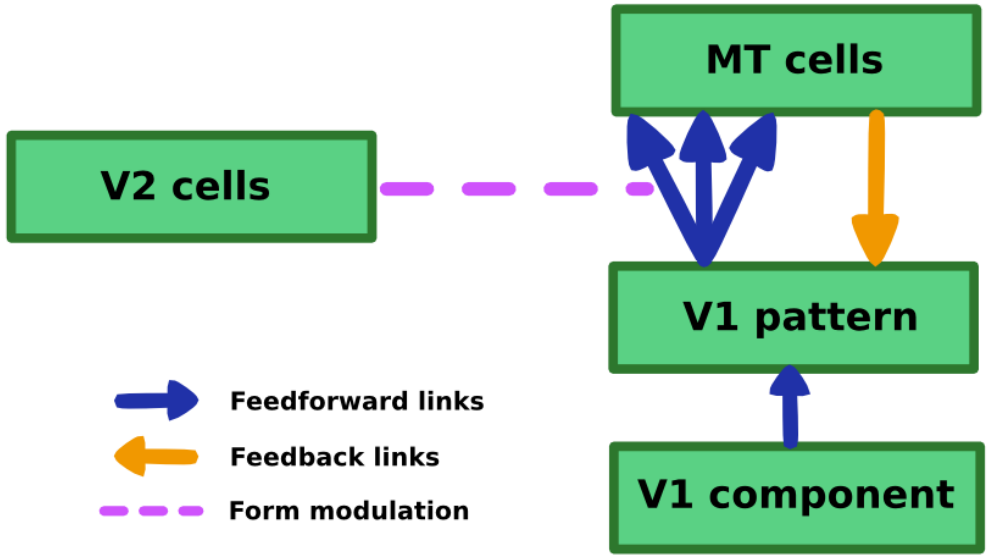


Figure 3.2: System overview. We model two kinds of V1 cells, the so-called component and pattern cells, MT cells and a form modulation coming from V2 cells. V1 component do not receive any feedback from MT, as opposed to the V1 pattern cells. V2 cells directly modulate the integration of V1 cells into MT. The feedback from MT to the V1 pattern cells will diffuse the global motion information.

3.2 Detailed presentation of the proposed model

3.2.1 Formalization

Input is described by a gray-level frame sequence

$$I : (t, x) \in (\mathbb{R}^+, X) \rightarrow I(t, x) \tag{3.1}$$

where x denotes a two-dimensional spatial coordinate in X.

Given a cortical layer i, we denote by $p_i(t, x, v)$ the activity at time t and position v of the neurons tuned for velocity v.

$$p_i : (t, x, v) \in (\mathbb{R}^+, X, V) \rightarrow p_i(t, x, v) \tag{3.2}$$

where V is the set of admissible velocities that we consider. This is equivalent to a polar coordinate (θ, ρ) found in the visual cortex where a neuron is tuned for a particular speed and a direction (see Section 1.2.3). $p_i(t, x, v)$ gives the cells activation likelihood for the cortical layer p_i. Note that by allowing more than a given velocity to occur at a given point we not only follow biological data but also allow detection of transparent motion. Modern computer vision also proposes multiple motion detection (Bergen et al., 1992).

The model is defined by the following set of equations:

$$p_0(t, x, v) = \text{Reichardt detectors}, \tag{3.3}$$

$$p_1(t, x, v) = G_\sigma(|v|)\, p_0(t, x, v), \tag{3.4}$$

$$p_2(t, x, v) = p_1(t, x, v)(1 + c_f p_3(t - 1, v, x)), \tag{3.5}$$

$$p_3(t, x, v) = \int_X G_\sigma(x - \xi)\pi(t, \Theta(\vec{\xi x}), \xi) p_1(t, v, \xi) \mathrm{d}\xi. \tag{3.6}$$

The p_1, p_2 and p_3 layers respectively contains the V1 component cells, the V1 pattern cells and the MT cells. p_0 is the same Reichardt motion detector as in Bayerl and Neumann (2007)'s model an described in Section 2.2. However we modulate the result by a prior on small velocities in equation (3.4). The prior acts as a selector in case of ambiguous motion and explaining psychophysical motion perceptions (Weiss and Adelson, 1998) and already described in Section 1.1.1. It is also used for type II plaids or the rhombus stimulus. The prior is a simple Gaussian of the speed velocity, thus $G_\sigma(|v|)$.

The V1 pattern cells represented by p_2 (3.5) receive feedforward input from the V1 component cells, p_1 but also a feedback input form MT. They are the finer scale velocity output in our model having a kind of memory from their MT feedback.

MT cells have a coarser level, as implied by their larger receptive field (see Section 1.2.3), represented in (3.6) by the spatial integration over X. The p_2 layer information is however not directly integrated into p_3 but modulated by object information coming from the form stream, the V2 layer modeled by π (3.8). The π (3.8) function gives at each point and for each direction a weight of integration according to the static form information (see Section 3.2.4).

3.2.2 Shunting inhibition

In addition to equations (3.3) to (3.6), we use a shunting inhibition at the end of each layer calculus, similar to Nowlan and Sejnowski (1994), Simoncelli and Heeger (1998) or Bayerl (2005) and described in the Chapter 2. Denoting $\hat{p}$ the inhibited version of p we have:

$$\hat{p}(t,x,v) = \frac{p^2(t,x,v)}{\sum_u p^2(t,u,x) \overset{v}{*} G_\sigma} \tag{3.7}$$

The small difference being the use of a blur instead of a sum to model biological connections between neurons at different retinotopic position and tuned to different velocities.

As we will see in Section 4.5 a shunting inhibition isotropic for the speed directions is not able to process transparent motion. Thus a Gaussian on speed direction might be required.

3.2.3 Motion integration

Motion integration is done in the V1/MT feedback loop. Figure 3.3 shows how the diffusion in V1 and even MT cells is obtained via feedback. Indeed, feedback connections are much faster than horizontal ones and serve as basis for the center/surround mechanism (Grinvald et al., 1994; Angelucci and Bullier, 2003). Moreover MT horizontal connections appear weak and a V1 path seems a reasonable hypothesis for even MT diffusion.

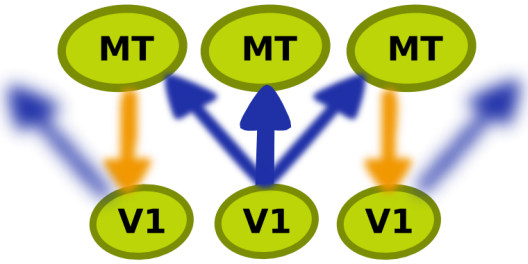


Figure 3.3: Motion diffusion in our model is done via recurrent connections between the V1 pattern cells and the MT layer.

As shown by Figure 3.4 motion information is propagated from unambiguous 2D cues, such as end of lines, to ambiguous regions (see Section 1.1.1). Furthermore the end of lines are not detected as such on the first frame. It takes a time to see that the motion in the center of the bar is ambiguous, this phenomenon is also pointed out in psychophysical experiments (Masson et al., 2000).

3.2.4 Form modulation

In (3.6) we used the function $\pi(t,d,x)$ as an integration weight for a given direction in a specific spatial point. Instead of having a simple isotropic integration at each point we define

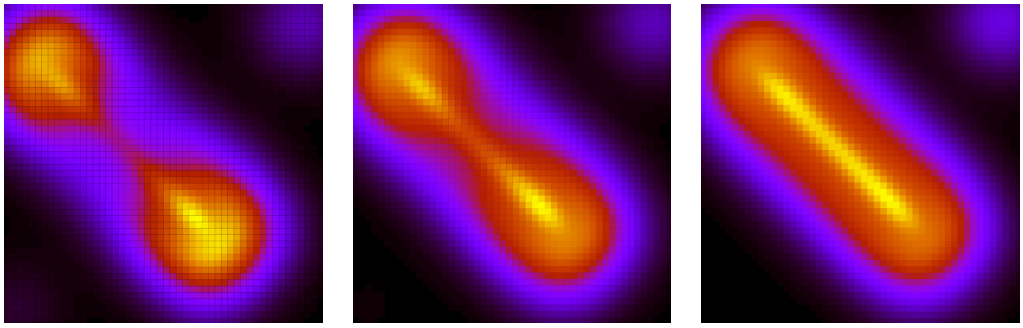


Figure 3.4: Three successive snapshots taken from the response of MT layer cells tuned to the bar speed. Motion diffusion propagate the unambiguous 2D signal to the center of the bar.

an isotropic integration according to the luminosities. For each spatial location of a given input frame, we consider all the directions. For each direction we compare the luminosity and integrate stronger if the luminosities are similar, lower if there is a high contrast. Thus we obtain an integration pattern at each point.

The direction of figure, π, is defined as follows:

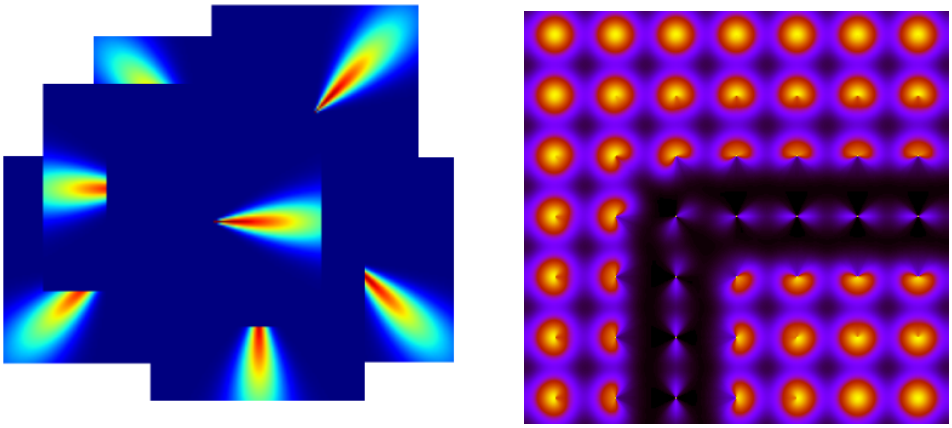
$$\pi(t,d,x) = \int_X G_\sigma(\xi - x)\, G_\sigma(d - \Theta(\overrightarrow{x\xi}))\, G_\sigma(S(\xi,x))\mathrm{d}\xi. \tag{3.8}$$

Therefore we have a product of three Gaussians:

- The first Gaussian on the distance between x and ξ.
- The second on the angular distance between d and the angle of $\overrightarrow{x\xi}$ (polar coordinate).
- The third is the stimulus similarity between the point at x and the one at ξ.

Combining spatial and angular distance gives us directional Gaussians modeling direction similarities neurons (see Figure 3.5 (a)). The V2 layer for a frame containing a corner is shown in Figure 3.5 (b).
Since our simple form correlator π gives an isotropic high response on homogeneous areas and an oriented lower response on contrast areas, and because we use π as a weight on the integration into MT, it is coherent with the observation made by Tadin et al. (2003) or Pack et al. (2005): Lower contrast implies larger integration area.

Previous studies used junctions detectors such as three-colors detectors (Weiss, 1998) X- and T-detectors (Bayerl, 2005). Those junctions where considered as intrinsic and ignored in further calculus. Our *direction of figure* π is also different from the one in Sajda and Baek (2004) who infers the owner object for each border point. Although simpler our direction of figure is able to give cues leading to successful motion perception. Figure 3.6 shows how the extrinsic junction leading to a false motion perception is eliminated while true motion is correctly diffused along two crossing bars.



(a) (b)

Figure 3.5: Illustration of the π function. (a) Directional neighborhood combining spatial and angular distance used to compute the π function. (b) A representation of the direction of figure, π, for a given frame. The representation use isospaced sample showing the form of motion integration that will be used in MT. We calculate the similarity among each directions at each point of the input frame and integrate on a given direction if it contains similar luminosities.

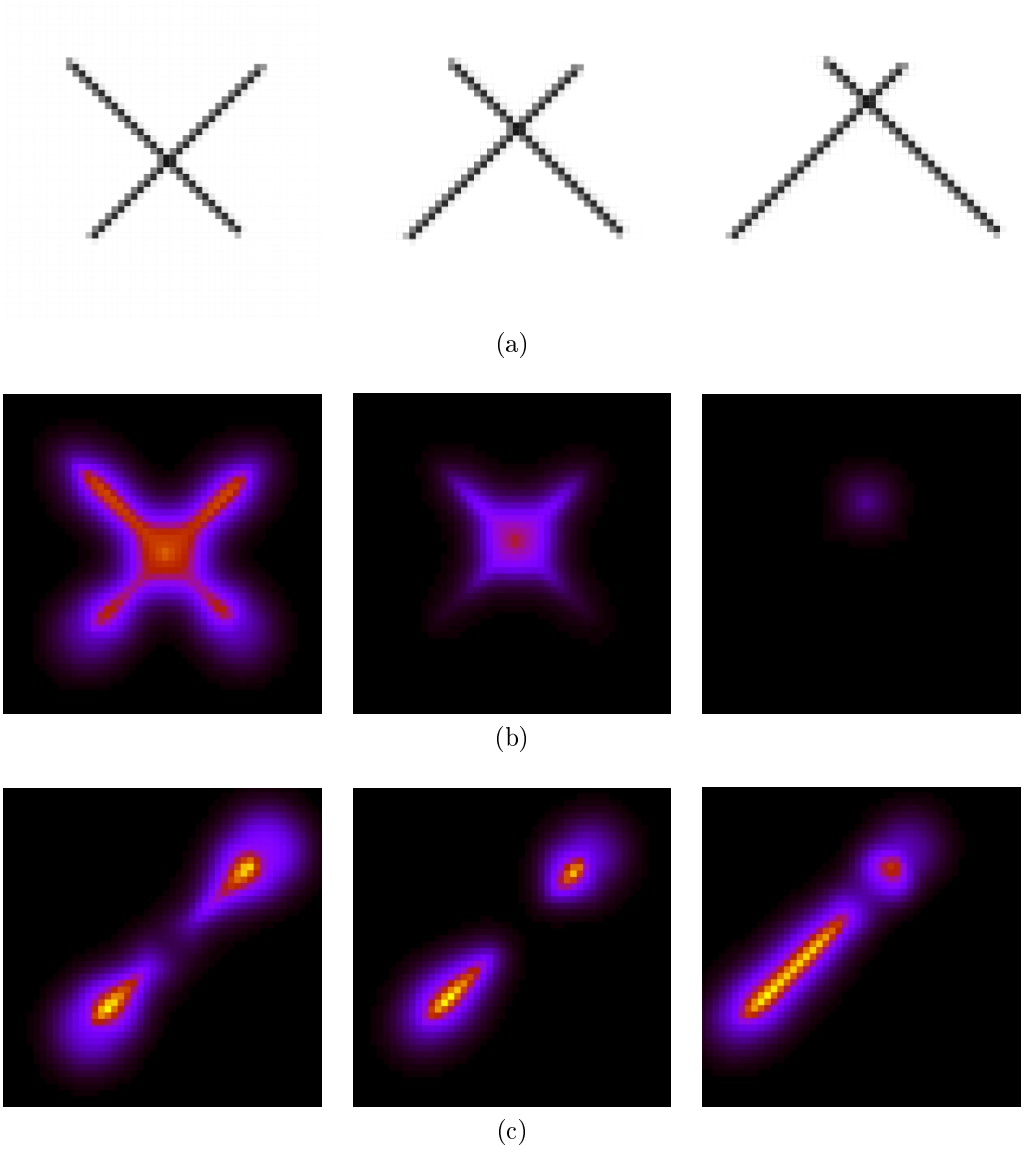


Figure 3.6: Model results on crossing bars. (a) The input stimulus consists of a two crossing bars: one with a leftward motion, the other with a rightward one. Locally we have an extrinsic junction where the bar crosses thus an upward false motion. (b) Showing the upward-tuned MT cells activation we can see that our model is able to discriminate this kind of motion. As time goes on the motion is restricted and lowered. (c) Correct motion are diffused correctly along the bars similarly to a model without form modulation. The three frames show the activation of the leftward-tuned MT cells.

Chapter 4

Results

In this chapter we present results of our model when applied to classical psychophysical stimuli such as Barber Poles or the chopstick illusion in order to demonstrates its biological plausibility. Last stimuli presented are more complex.

4.1 Barber Poles

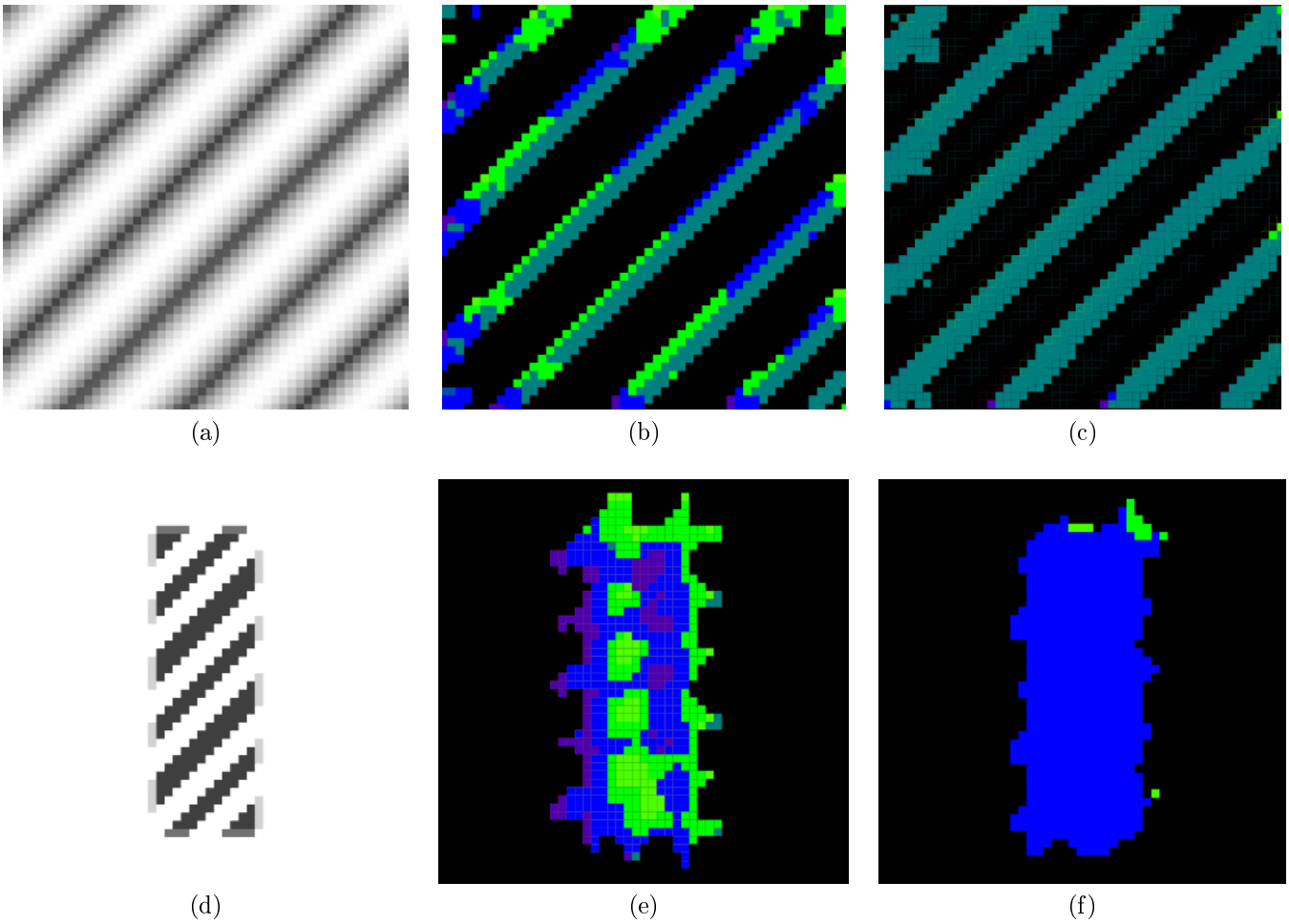
As explained in Section 1.1.2, Barber Poles can show the integration process we model in equation (3.6). In Figure 4.1 (a)–(b) we first experiment our model on a plaid viewed through a squared aperture thus without being able to locally discriminate between ambiguous motions. The prior on small velocities described in Section 1.1.1 and modeled by (3.4) explain the percept. In Figure 4.1 (d)–(f) the classical asymmetric Barber Pole is tested and we get results similar to those obtained in psychophysical experiments.

4.2 Chopstick illusions

The Chopstick illusion presented in Section 1.1.3 gives the opportunity to test the form modulation of our system in order to discriminate between extrinsic and intrinsic junctions. We already shown in Figure 3.6 that the model is able to ignore the extrinsic junction of two crossing bars. Results for the principal direction are redrawn on Figure 4.2.

On Figure 4.3 we display the maximum velocity at each point for the Chopstick illusion with occluders. As in our perception the model then detects a single vertical motion for the crossing bars. Note that there is also a motion detection in our model for the occluders.

Cells direction tuning color map

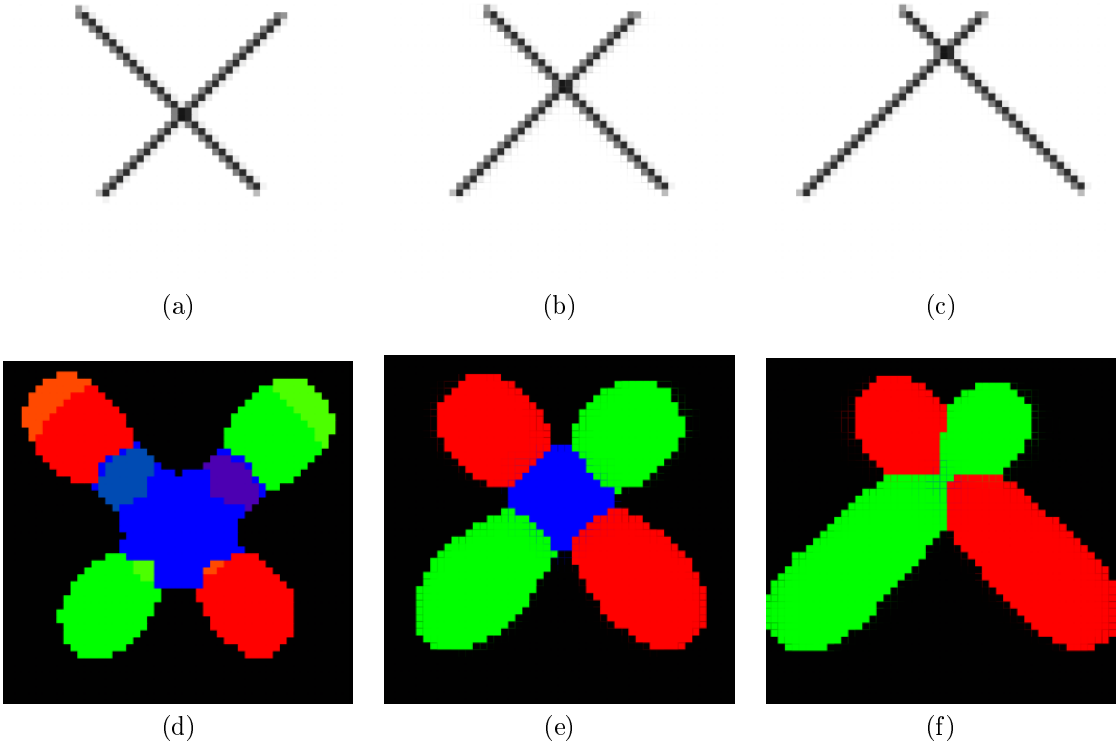


(a) (b) (c)

(d) (e) (f)

Figure 4.1: V1 cells direction results. As we only display the direction response for the V1 cells, we use a color circle to show the direction of the cell with the more important response. (a) A moving grating viewed through an invisible rectangular aperture. (b) First frame response for the V1 pattern cells. (c) V1 pattern cells response after 7 frames.

Cells direction tuning color map

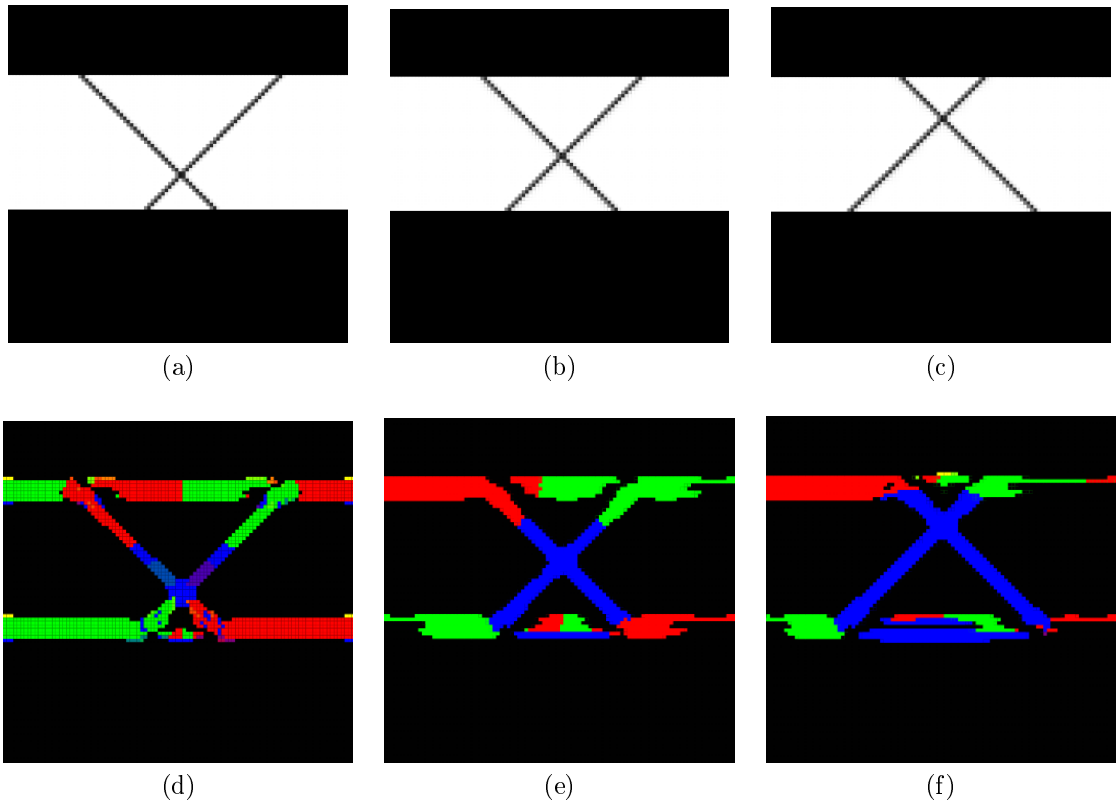


(a) (b) (c)

(d) (e) (f)

Figure 4.2: Chopstick illusion. On the non occluded Chopstick illusion we perceive two crossing bars. Their intersection leads to an extrinsic junction having a vertical motion ignored by perception as well as our model. We display the direction of the speed cells with higher response at each point.

Cells direction tuning color map



(a) (b) (c)

(d) (e) (f)

Figure 4.3: Occluded chopstick. The stimulus is the same as the crossing bars used in Figure 3.6 but the end of lines are occluded. Percept is then a single vertical motion instead of two opposite horizontal motions. In our model the previously extrinsic junction also leads the integration although some motion is apparent on the occluders. We display the direction of the speed cells with higher response at each point.

4.3 Yosemite

In order to show that our proposed model is indeed able to calculate motion streams on complex sequences we start by an experiment on the classical Yosemite sequence.

This sequence is a synthetic video simulating a motion inside mountains with clouds moving in opposite direction opposite to the camera. Results are shown in Figure 4.4. At first we got some incoherent motion but finally the perceived motion is extracted.

On Figure 4.4 (f) we see that a correct motion field is extracted, coherent with our perception. The clouds are coherently moving to the right, we see the point in the camera motion direction as a black spot with no motion, the green regions are moving to the left, the red ones are moving to the right and some in between encodes left-down or right-down motions. Also note a very precise form segmentation.

Some errors also occurs such as incoherency on the borders of our sequence. The observed staircase effect is due to the discretization of the velocity space. We are only using 24 possible velocities.

4.4 Taxi sequence

The taxi sequence is another video classically used in computer vision. It includes ambiguous reflections and motion direction changes. As seen in Figure 4.5 the three moving cars and the moving person are segmented and their motion correctly interpreted by our model.

4.5 Transparent trees

We now consider a transparent motion sequence in order to see the capacities of our model. The stimuli consists of two forest photographies combined with a pixel value average and each having different translating motion (see Figure 4.6 (a)–(c)). Our model create patches of the two frames velocities each having either one or the other direction. Although we only display the maximum direction this is true for the two principal velocities.

What can be concluded is that the classical shunting inhibition used in Nowlan and Sejnowski (1993) or Bayerl (2005) is not adapted to transparent motion. We suggest that this is due to the isotropy of the inhibition. Further experiments may use an anisotropic inhibition with maximum effect for velocities with similar directions. This is also confirms psychophysical experiments showing an elevated response for cells with a surround with opposite direction and inhibition for cells with same direction in the surround.

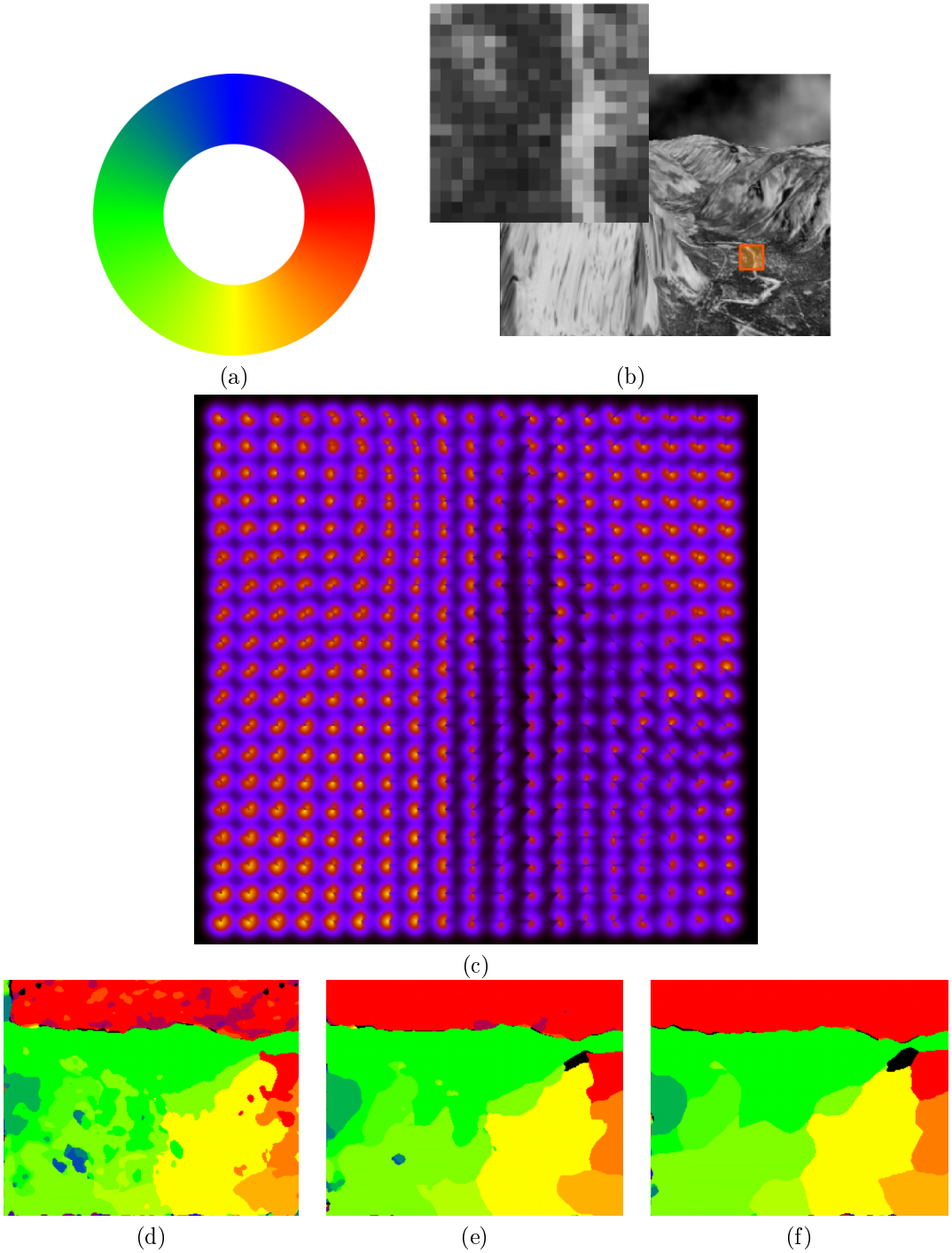
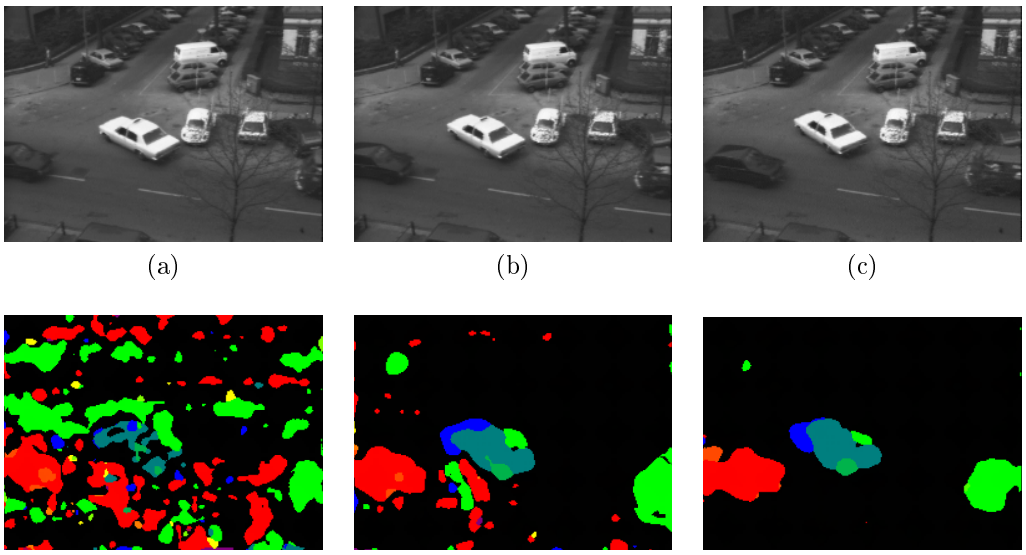


Figure 4.4: Results of our model on the Yosemite sequence in the V1 pattern cells. (a) As we only display the direction of motion, and not the speed, we use this circle color code. (b) We display the motion form detector only for a small area. (c) The ventral stream output for this small square. (d) At first various motion, including inadequate ones are found on the figure. (e) Two frames later the motion field is smoothing and removing ambiguous motions. (f) Finally the motion field is coherent with our perceived motion.

Cells direction tuning color map

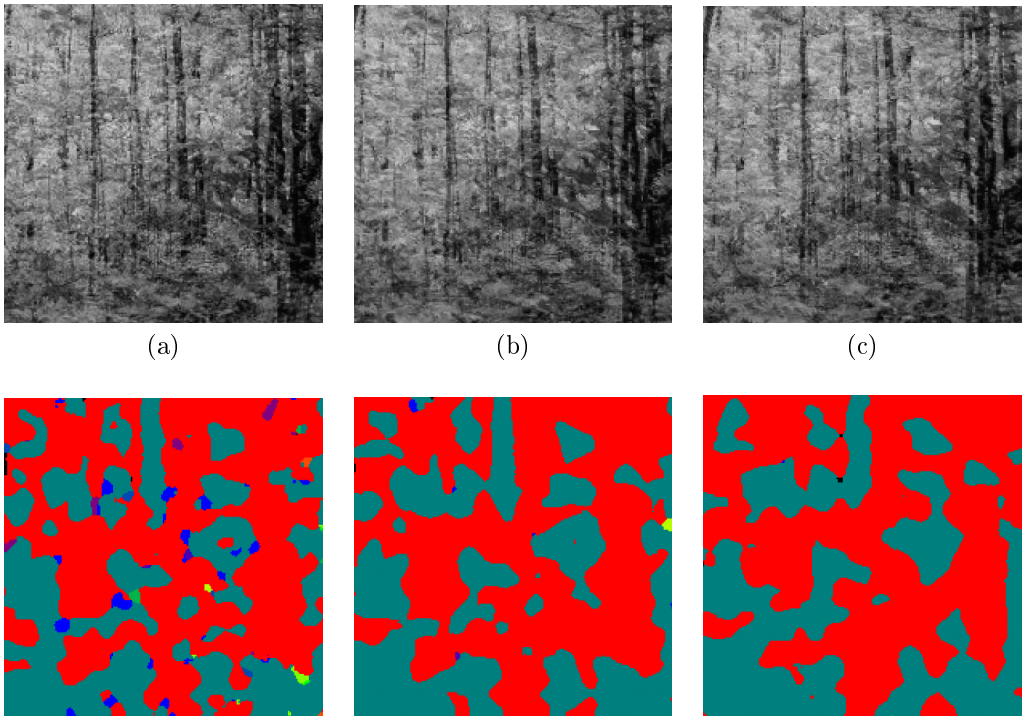


(a) (b) (c)

(d) (e) (f)

Figure 4.5: Results on the taxi sequence. Images (a)–(c) show a video sequence of three cars moving and a single person walking. In (d)–(f) we display the direction with maximal response at each location for the V1 area.

Cells direction tuning color map



(a) (b) (c)

Figure 4.6: Results on transparent. (a)–(c) Stimulus consists of two translating images combined to obtain a transparent motion. (d)–(f) Our model does not see the two motions at each point but one win thus forming patches with one of the two possible motions. This is due to the isotropic shunting inhibition. Maximum direction of the MT layers are show there.

Chapter 5

Conclusion

Based on physiological, psychophysical and previous modeling studies we proposed a model of motion processing in the primate visual cortex based of form cues.

We used a recurrent system for the V1/MT motion processing system and suggested a biological explanation for the feedback and the MT subunits based on V1 pattern cells found by different groups.

By using a novel form cue not depending on geometrical junctions we were able to explain a greater number of psychophysical illusions such as the Chopstick.

The proposed system also explain how the center-surround effects both isotropic and anisotropic come from the recurrent system together with the form cues. We also obtain a spatial integration whose area depends on contrast thus conforming to psychophysical experiments.

Future experiments will test anisotropic shunting inhibition to check whether it works on transparent motion.

Acknowledgments

This work was partially supported by the EC IP project FP6-015879, FACETS.

Bibliography

A. Angelucci and J. Bullier. Reaching beyond the classical receptive field of v1 neurons: horizontal or feedback axons? *J Physiol Paris*, 97(2–3):141–154, Mar–May 2003. doi: 10.1016/j.jphysparis.2003.09.001.

A. Angelucci, J. B. Levitt, E. J. S. Walton, J. M. Hupe, J. Bullier, and J. S. Lund. Circuits for local and global signal integration in primary visual cortex. *Journal of Neuroscience*, 22(19):8633, 2002.

P. Bayerl. *A model of visual motion perception.* PhD thesis, Universität Ulm. Fakultät für Informatik, 2005.

P. Bayerl and H. Neumann. Disambiguating visual motion by form-motion interaction—a computational model. *International Journal of Computer Vision*, 72(1):27–45, 2007.

J. R. Bergen, P. J. Burt, R. Hingorani, S. Peleg, D. S. R. Center, and N. J. Princeton. A three-frame algorithm for estimating two-component image motion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 14(9):886–896, 1992.

R. T. Born and D. C. Bradley. Structure and function of visual area mt. *Annu Rev Neurosci*, 28:157–189, 2005. doi: 10.1146/annurev.neuro.26.041002.131052.

O. J. Braddick, J. Wattam-Bell, J. Atkinson, and R. Turner. Brain areas sensitive to coherent visual motion. *Perception*, 30(1):61–72, 2001.

D. C. Bradley, G. C. Chang, and R. A. Andersen. Encoding of three-dimensional structure-from-motion by primate area mt neurons. *Nature*, 392(6677):714–7, 1998.

F. Crick and C. Koch. A framework for consciousness. *Nature Neuroscience*, 6(2):119–126, 2003.

G. C. DeAngelis and W. T. Newsome. Organization of disparity-selective neurons in macaque area mt. *J Neuroscience*, 19(4):1398–1415, Feb 1999.

S. Gepshtein and M. Kubovy. Stability and change in perception: spatial organization in temporal context. *Experimental Brain Research*, 160(4):487–495, 2005.

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. *J Neurophysiol*, 67(6):1437–1446, Jun 1992.

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. *Journal of Neuroscience*, 14(5):2545–2568, 1994.

K. Guo, P. J. Benson, and C. Blakemore. Pattern motion is present in V 1 of awake but not anaesthetized monkeys. *European Journal of Neuroscience*, 19(4):1055–1066, 2004.

E. C. Hildreth and C. Koch. The analysis of visual motion: From computational theory to neuronal mechanisms. *Annual Review of Neuroscience*, 10(1):477–533, 1987.

X. Huang, T. D. Albright, and G. R. Stoner. Adaptive surround modulation in cortical area mt. *Neuron*, 53:761–770, March 2007.

J. M. Hupe and N. Rubin. The dynamics of bi-stable alternation in ambiguous motion displays: a fresh look at plaids. *Vision Research*, 43(5):531–548, 2003.

G. S. Masson, D. R. Mestre, and L. S. Stone. Speed tuning of motion segmentation and discrimination. *Vision Research*, 39(26):4297–4308, Oct 1999.

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. *Visual Neuroscience*, 17(05):753–767, 2000.

D. R. Mestre, G. S. Masson, and L. S. Stone. Spatial scale of motion segmentation from speed cues. *Vision Research*, 41(21):2697–2713, September 2001.

J. A. Movshon and W. T. Newsome. Visual response properties of striate cortical neurons projecting to area mt in macaque monkeys. *Journal of Neuroscience*, 16(23):7733–7741, 1996.

J. J. Nassi, D. C. Lyon, and E. M. Callaway. The parvocellular lgn provides a robust disynaptic input to the visual motion area mt. *Neuron*, 50(2):319–327, 2006.

S. J. Nowlan and T. Sejnowski. Filter selection model for generating visual motion signals. *Advances in Neural Information Processing Systems*, 5:369–376, 1993.

S. J. Nowlan and T. J. Sejnowski. Filter selection model for motion segmentation and velocity integration. *J. Opt. Soc. Am. A*, 11(12):3177–3199, 1994.

C. C. Pack, A. J. Gartland, and R. T. Born. Integration of contour and terminator signals in visual area mt of alert macaque. *Journal of Neuroscience*, 24(13):3268–3280, 2004.

C. C. Pack, J. N. Hunter, and R. T. Born. Contrast dependence of suppressive influences in cortical area mt of alert macaque. *J Neurophysiology*, 93(3):1809–1815, Mar 2005. doi: 10.1152/jn.00629.2004.

K. S. Rockland and T. Knutson. Feedback connections from area MT of the squirrel monkey to areas V 1 and V 2. *The Journal of Comparative Neurology*, 425(3):345–368, 2000.

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. *J Neurosci*, 9(6):2033–2050, Jun 1989.

O. Sacks. In the river of consciousness. *New York Review*, 51(1), 2004.

P. Sajda and K. Baek. Integration of form and motion within a generative model of visual cortex. *Neural Networks*, 17(5-6):809–821, 2004.

M. P. Sceniak, M. J. Hawken, and R. Shapley. Visual Spatial Characterization of Macaque V1 Neurons. *Journal of Neurophysiology*, 85(5):1873–1887, 2001.

S. Shimojo, G. H. Silverman, and K. Nakayama. Occlusion and the solution to the aperture problem for motion. *Vision Res*, 29(5):619–26, 1989.

A. M. Sillito, J. Cudeirob, and H. E. Jonesa. Always returning: feedback and sensory processing in visual cortex and thalamus. *Trends in Neuroscience*, 29(9):307–316, June 2006. doi: doi:10.1016/j.tins.2006.05.001.

EP Simoncelli and DJ Heeger. Model of Neuronal Responses in Visual Area MT. *Vision Research*, 38(5):743–761, 1998.

R. J. Snowden, S. Treue, R. G. Erickson, and R. A. Andersen. The response of area mt and v1 neurons to transparent motion. *The Journal of Neuroscience*, 11(9):2768–2785, Sep 1991.

D. Tadin, J. S. Lappin, L. A. Gilroy, and R. Blake. Perceptual consequences of centre-surround antagonism in visual motion processing. *Nature*, 424(6946):312–5, 2003.

C. J. Tinsley, B. S. Webb, N. E. Barraclough, C. J. Vincent, A. Parker, and A. M. Derrington. The Nature of V1 Neural Responses to 2D Moving Patterns Depends on Receptive-Field Structure in the Marmoset Monkey. *Journal of Neurophysiology*, 90(2):930–937, 2003.

H. Wallach. Über visuell wahrgenommene bewegungsrichtung. *Psychological Research*, 20 (1):325–380, 1935.

H. Wallach. *On Perception*. Quadrangle/New York Times Book Co., 1976.

Y. Weiss. *Bayesian motion estimation and segmentation*. PhD thesis, Massachusetts Institute of Technology, 1998.

Y. Weiss and E. H. Adelson. Slow and smooth: A Bayesian theory for the combination of local motion signals in human vision. *Center for Biological and Computational Learning Paper*, 1998.

D. Xiao, S. Raiguel, V. Marcar, J. Koenderink, and G. A. Orban. Spatial heterogeneity of inhibitory surrounds in the middle temporal visual area. *Proceedings of the National Academy of Sciences*, 92(24):11303–11306, 1995.

D. K. Xiao, S. Raiguel, V. Marcar, and G. A. Orban. The spatial distribution of the antagonistic surround of mt/v5 neurons. *Cereb Cortex*, 7(7):662–77, 1997.

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

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

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