Invariant models of vision between phenomenology, image statistics and neurosciences.

by

Gonzalo Sanguinetti

Thesis Directors:

Prof. Giovanna Citti
Prof. Alessandro Sarti

Academic Director:

Prof. Gregory Randall

Contents

Chapter 1

Introduction

The objective of this thesis is to develop mathematical models of vision which integrate empirical data from three different scientific areas: phenomenology of perception, cognitive neurosciences and image statistics. We mainly focus on low level vision tasks and in particular we are interested in the completion and grouping of contours. Our scope is to identify common properties and structures within the experimental data and to explain them in terms of simple invariance principles, common to the three fields.

This work could be considered as a part of a project for naturalizing phenomenology. An in-depth insight to these contemporary issues is discussed in [76]. According to this book, naturalization is the process by which phenomenology is made coherent and compatible with natural science, filling the gap between the science of nature and the science of mind. Of course this is a long term objective (a utopia perhaps) and there are different positions between the various schools of philosophers, which we are not able discuss here.

In the field of the phenomenology of perception, the integration of contours has been largely studied by the Gestalt theory since the beginning of the twentieth century [94]. The Gestalt psychology established basic "grouping laws" which are crucial in constructing a phenomenological representation of the physical world. The basic idea is that points (or previously formed visual objects) having one or several characteristics in common, are grouped together to form a new, larger visual object, a gestalt. A central concept of the Gestalt is that the grouping laws are independent from past experience. More recently, many psychophysical experiments allowed to measure the quantitative parameters of these laws underlying perception. This quantitative approach allows to formulate mathematical and computational models of Gestalt (see for example

[18]). A particular interest of this thesis is the concept of association fields introduced by Field, Hayes and Hess [29] which encodes different Gestalt principles such as good continuation and proximity.

In neurophysiology, most of the neural processing for boundary coding in highly evolved mammals, including humans, is performed by the primary visual cortex (V1). This evidence emerges from an impressive amount of experiments based mostly on measuring the neural activity by means of dye-electrode recording and cortical imaging. For example, simple cells in V1 are considered basic elements for boundary coding. Nowadays it is clear that neural computations and its image analysis strictly depend on the functional architecture of the cortex. Indeed, in V1, the fundamental structures implemented in the neural circuitry are closely related to contour grouping [41].

Finally, statistical analysis of natural images by means of information theory produces data structures that can be compared with psychophysical and neurophysiological results. Possible isomorphism's between the structure of the environment, gestalt and neural connectivity may be due to evolution and plasticity processes (an adaptation of our sensory system to the environment). Then, it might be of great interest to estimate statistical structures of natural images by analyzing large databases of pictures from arbitrary natural scenes [87].

Our aim is to formally model the experimental data provided by the three different scientific fields mentioned above within a suitable mathematical framework. The framework was introduced by Hoffman in [39], who understood that the primary visual cortex is a fiber bundle equipped with a contact structure. In [75, 73] J. Petitot and Y. Tondout reassessed the model of W. Hoffman by modelling the cortex as a jet-bundle equipped with a contact form and proving that this structure is coherent with contemporary psychophysical and neurophysiological findings. Similar ideas taking into account a differential geometry approach were introduced by S. Zucker [6]. In [82, 14] G. Citti and A. Sarti reconsidered the cortical structure as a Lie group equipped with a sub-Riemannian metric, inheriting from Hoffman the capability of modelling higher perceptual tasks by stratification of groups. In a recent article [13], P. Chossat and O. Faugeras proposed that the hypercolumns of orientations encode the structure tensor so the cortex may be modelled as a space of 2D symmetric tensors. Then, the time evolution of the activity in the cortex is governed by a Wilson-Cowan equation type [97], operating under the appropriate mathematical space which takes into account the rotation and translation symmetries.

In this setting we are able to identify the main original contributions of the thesis:

- a new empirical domain is introduced assessing the statistics of natural images;
- due to the intrinsic statistical character of the acquired data a suitable stochastic mathematical framework is introduced in the invariant geometrical settings;
- isomorphism's between the three empirical domains have been identified and modelled.

The thesis is organized as follows:

Chapter 2 starts by briefly reviewing the Gestalt theory and some of its basic laws. It describes how from a pyramidal composition of Gestalts a perceptual reality is created, formed by a discrete number of visual objects emerging from a background. Then the perceptual completion phenomenon is discussed, distinguishing between the amodal and modal completion. The first refers to the reconstruction of the "under" of the physical image (the background or an occluded object) while the second refers to illusory contours (building perceptual units indistinguishable from real stimuli). In both cases there are points belonging to more than one perceptual object at the same time. This suggests that the dimension of the phenomenological space is higher than that of the physical one. The chapter ends by describing one key psychophysical experiment related to contour organization which inspired the concept of association fields, and it is the classical result of Field, Hayes and Hess [29]. Then, a recent experiment [57] which has elucidated certain new properties of trans-axial association fields is carefully reviewed.

Chapter 3 collects various results from neurophysiological experiments. The objective is to describe the functional architecture of V1 and how neural computations strictly depend on it. First, receptive fields of simple cells are interpreted as elements for contour detection. Then the three main structures implemented by neural circuitry are described: the layered, the retinotopic and the hypercolumnar. Next, the pinwheel structure is introduced as the real topological implementation of the hypercolumns. Finally, the well known experiment of Bosking et Al. [11], which revealed the cortico-cortical connectivity pattern, is described.

Chapter 4 reviews the functional architecture of the V1 geometrical model introduced by Citti and Sarti in [14]. Following ideas first introduced by

Hoffman and Petitot, the hypercolumnar structure is described as a fiber bundle where the role of simple cells as orientation detectors leads to a notion of a cotangent bundle. The structure is also identified with the Euclidean Motion Group $SE(2)$, the 3D Lie group of transformations invariant to translations and rotations in the plane. Then, a differential structure is considered, in order to model the long range co-axial horizontal connections between hypercolumns. The structure is formalized as the Lie algebra of the $SE(2)$ with a sub-Riemannian structure. The integral curves of its generating vector fields model the co-axial association fields.

Chapter 5 presents an implementation of the perceptual completion model proposed by Citti and Sarti in [82, 14]. Within the model, an image was lifted onto a surface in the $SE(2)$ space. The completion was achieved by means of a propagation process modelled as a two step algorithm inspired by neural architectures. The original work provided proof that the algorithm converges to a diffusion driven mean curvature flow in the sub-Riemannian settings and an implementation of the mean curvature flow was proposed to provide completion. In this chapter we directly implement the two step algorithm of diffusion-concentration that permits representation of simultaneous occluded and occluding objects. The diffusion process, modelling cortical propagation, is restricted to the sub-Riemannian differential structure. The implementation of the completion algorithm is the first original contribution in the thesis. This work was published in [79, 80].

Chapter 6 links the sub-Riemannian model with a well known probabilistic approach to the perceptual completion problem introduced by D. Mumford [62]. Mumford proposed modelling the probability of the reappearance of an occluded contour in a natural scene with a particular Fokker Planck equation in the space of position and orientations. This hypothesis was the inspiration for successive models which use in different ways the same Fokker Planck equation to address propagation and enhancement of contours [95, 4, 20]. Some of the works identified the group of positions and orientations with the $SE(2)$. The novelty introduced in this Chapter is the interpretation of the Mumford's Fokker Planck operator as a natural propagation equation in the cortically inspired structure. We introduce a new time-independent fundamental solution of the Fokker Planck operator which models the lateral connectivity in the cortex. This operator is better suited as a connectivity pattern than the sub-Riemannian heat kernel which was used in Chapter 5 to propagate the cortical activity. Then, we provide an analytical approximation and a numerical computation of the fundamental solution. Finally, an interpretation of the orientation distribution as a second order tensor allowed us to compare these methods with another well known heuristic perceptual

completion algorithm, the Tensor Voting of G. Medioni [58]. Some of the results in this chapter are published in [81].

Chapter 7 deals with the computation of the distribution of oriented edges from a large data set of natural scenes and models the acquired results in sub-Riemannian settings with both deterministic and stochastic instruments. The Chapter starts by reviewing previously reported histograms of edge co-occurrences computed in natural image databases. Then, a technique for the estimation of co-occurrence probability of relative positions and orientations is used to obtain a translation-orientation invariant 3D histogram. The histogram is compared with psychophysical and neurophysiological data. From the deterministic point of view, we show that the histogram encodes the association fields. In the stochastic framework, the time independent Fokker Planck fundamental solution well fits the probability of co-occurrence. This is one of the most notable results of the thesis. Parametric identification between the Fokker Planck and the co-occurrence histogram is provided. The results of this chapter are published in [81].

In Chapter 8 the feature of scale is added to the features of boundary orientation. This leads to consider a 4 dimensional histogram and the affine group (the Lie group invariant to rotations, translations and scaling transformations). The extension of the cortical model was provided by Sarti, Citti and Petitot in [84] where a symplectic structure is introduced. The scale takes the form of the distance from a boundary. Then, it is possible to take into account the interior of the objects. Using analogous considerations to the ones in [14], two sub-Riemannian structures are used as models for the connectivity. One is inherited from the $SE(2)$ model and models the co-axial connectivity. The other is responsible for trans-axial connectivity and it is morphologically different from the previous one. Finally, a differential operator for modelling the trans-axial connectivity is introduced by analyzing natural image's statistics.

Chapter 9 is devoted to trans-axial association fields recently reported in [57] and reviewed in Section 2.4. These association fields basically describe the perceptual saliency of the parallel structure which have a morphological difference with the co-linear structure. First, the psychophysical data is re-analyzed to achieve a format compatible to a subsequent comparison with the statistics of natural images and with neurophysiological models. It is shown that the statistical distribution of parallel and co-linear edges is coherent with the psychophysical data. On the other hand, the connectivity predicted by the symplectic model is in agreement with these association fields.

The last Chapter is devoted to conclusions.

Chapter 2

Phenomenology of perception

2.1 Introduction

Visual perception is not a simple acquisition of the real stimulus, but the result of a series of complex processes able to mediate between the physical stimulus and the phenomenological organization. According to Gaetano Kanizsa, one of the main exponents of the Gestalt psychology:

> *"Perception is not a mere passive recording of information impressed upon my sensory organs by the environment. Rather, it consists of an active construction by means of which sensory data are selected, analyzed, and integrated with properties not directly noticeable but only hypothesized, deduced, or anticipated, according to available information and intellectual capacities."*

In his book "Grammatica del vedere" [46], Kanizsa supported this positions, by studying in depth the perceptual phenomena on synthetic images, focusing in particular on the phenomenical presence of boundaries without physical stimulus (as for example in the well known triangle of Kanizsa shown in Fig. 2.6), describing the mechanisms of modal and amodal completion.

In this section we briefly recall the main concepts developed by the Gestalt theory, assessing the main grouping laws: proximity, similarity, good continuation, alignment, closure.

In particular we are interested in the property of good continuation and alignment, which is the basis of the phenomena of boundary completion. This phenomenon studied in depth by Fields, Hayes and Hess [29], who

experimentally tested the ability of subjects to detect a target alignment composed of small Gabor-like patches (see Fig. 2.8). The result of their experiments is summarized in a graph of so called association fields, which illustrates the set of all possible subjective boundaries starting from a fixed point.

Finally we move on to other perceptual phenomena such as parallelism and scale.

2.2 Gestalt theory

The movement of the Gestalt theory was started by Wertheimer, Köhler and Koffka, who considered perceptual phenomena as global events, not reducible to the set of the parts, well described by von Ehrenfels:

> *"The whole is more than the sum of the parts."*

Hence "grouping" is a crucial process in visual perception: whenever points (or previously formed visual objects) have one or several characteristics in common, they become grouped and form a new, larger visual object, a gestalt.

The first concept developed by the Gestalt theory was the articulation of the visual field in figure and ground, representing the minimal structure of visual perception. In the case of ambiguous figures the problem of the selection of figure-ground is particularly evident: a cup or two profiles are perceived as a figure, following the specific figure-ground selection (Fig. 2.1).

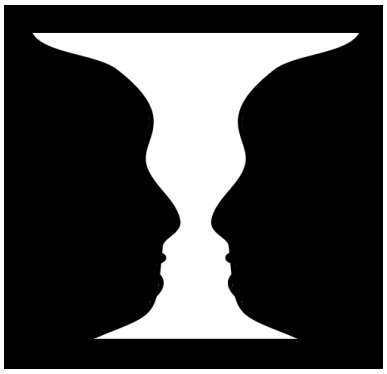


Figure 2.1: The Rubin vase. An optical illusion, which may be perceived either as a vase or as two human profiles facing each other. This illusion was created by the Danish psychologist Edgar Rubin.

The basic idea of the Gestalt theory is that there exist laws which allow emergence of figures without any mediation of past experience. These characteristics were defined as laws that describe the influence of global context in the perception of local features. Elements tend to be perceptually grouped and made salient in the following cases:

- **proximity:** Elements are close to each other, and apart enough from the rest of the elements to form a group or a cluster (see Fig. 2.2)

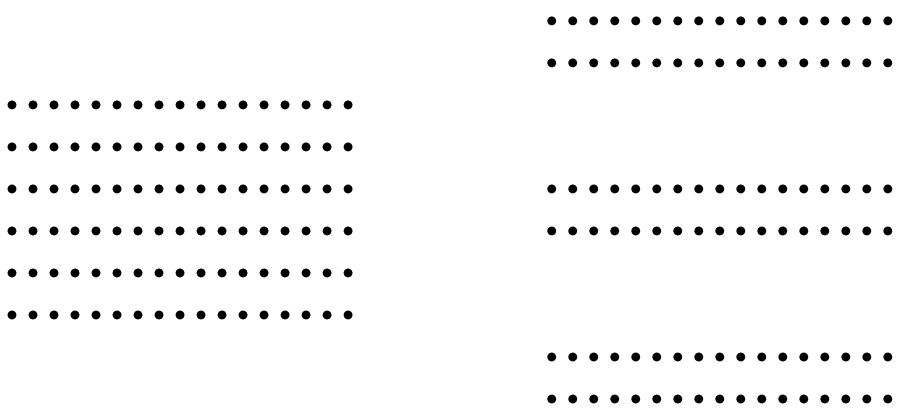


Figure 2.2: An example of grouping according to proximity. In the left and right image there are the same set of points, but on the right the points are grouped by the proximity constraint. This example was taken from [46].

- **similarity:** Elements similar in color, shape, texture or orientation are grouped together. Each one of these provides a partial gestalt (see Fig. 2.3).

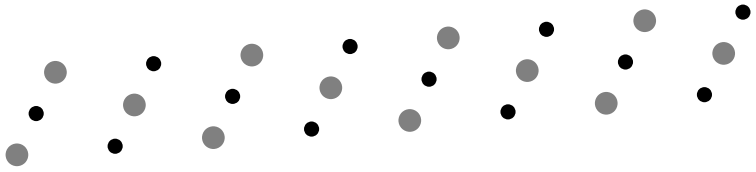


Figure 2.3: An example of grouping according to similarity. Similarity wins over proximity in this instance.

- **closure:** The perception is even stronger in case of closed contour. In this case, we have a pop up of the figure, in the sense that we perceive an object, defined by the closed contours (see Fig. 2.4).

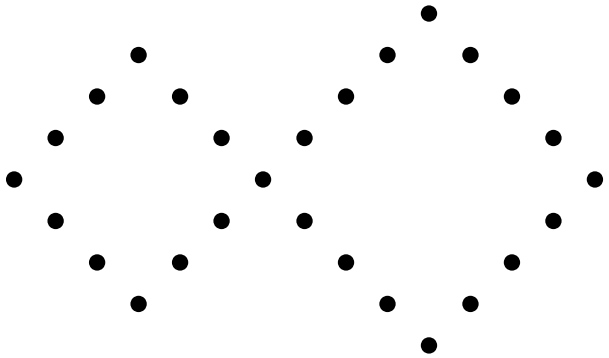


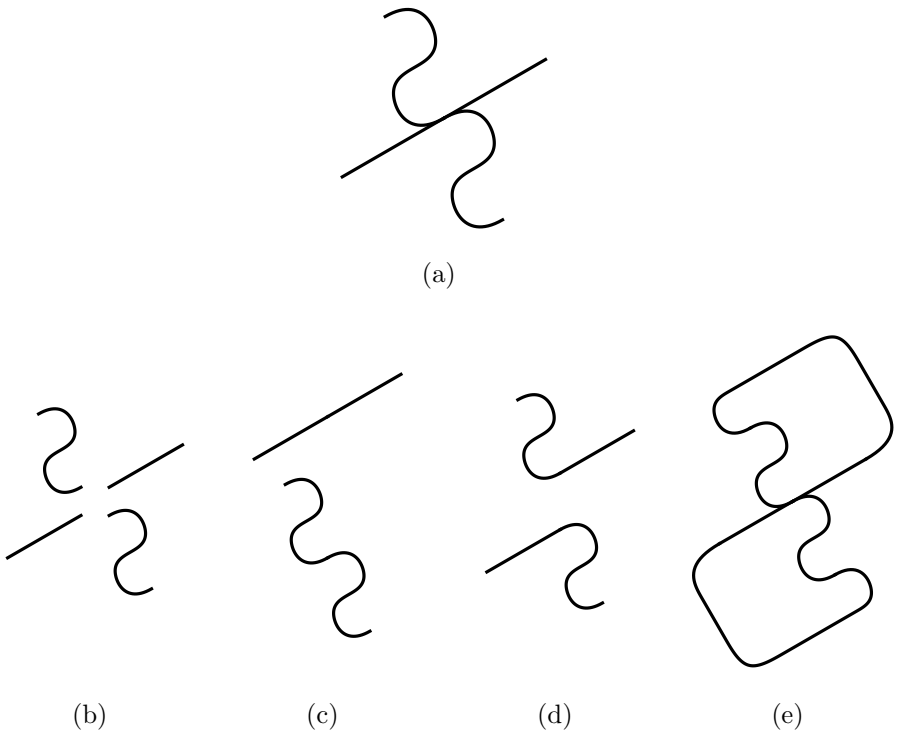
Figure 2.4: An example of grouping according to closure. This example was taken from [46].

- **good continuation and alignment:** Elements aligned or with comparable alignment tend to form a continuous curve (see Fig. 2.5).

From the above experiments we can observe that:

- the perceptual reality is formed in each time by a discrete number of objects;
- these objects do not necessarily depend on the existence of correspondent physical objects
- the segmentation of the phenomenological space between objects and background depend on precise stimulation conditions.

More than one grouping law at a time can contribute to the perception of a complex object. For example while perceiving a region enclosed by a contour several gestalt can play a role: alignment of the boundary, closure, convexity, similarity in color of the points within the interior (see Fig. 2.5). Moreover, gestalts can be applied iteratively, to the atomic input, and then to the partial gestalts already recognized.



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

Figure 2.5: This image (a) is formed by four pieces (b). It is usually interpreted as the intersection of the two curves shown in (c) instead of the configuration of the two tangent curves (d). Nevertheless if the segment is closed las in (e) the tangent configuration emerges, due to the closure law. This example is inspired from the Figures 1.30 and 1.37 of Kanizsa's book [46].

2.3 Perceptual Completion

Every time that a figure is segmented, the ground is perceived as completed "under" the figure. Since the occluded background underlies the occluding figure, it is completed without any sensory counterparts. This phenomenon is called by Kanizsa "amodal completion". Amodal completion is a fundamental process taking place every time that

- a figural unit is produced (occluding the background);
- a figure occludes another figure;
- a figure is self-occluding.

An example of amodal completion due to occlusion is presented in Fig. 2.6 left.

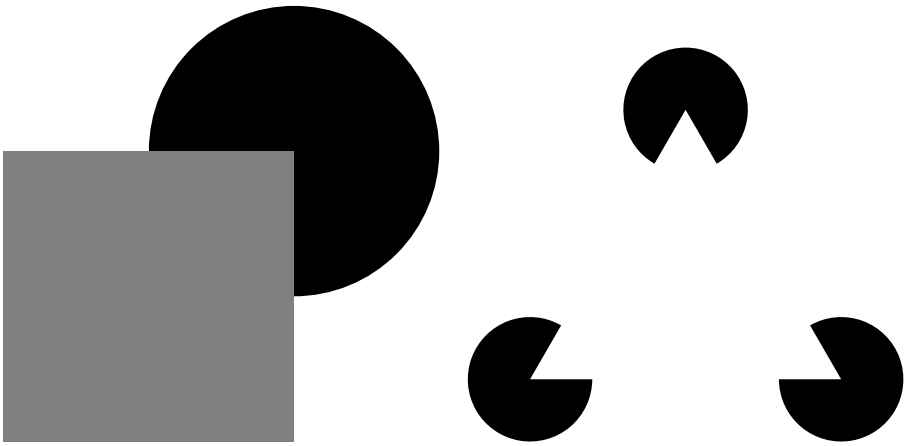


Figure 2.6: (Left) An example of *amodal completion*. The figure is perceived as a black circle occluded by the gray square. The circle is present in the visual field, but the completion is performed without an illusory contour. (Right) The Kanizsa triangle [46]. A white triangle occluding three black disks is phenomenologically perceived. There is an apparent contour separating the triangle from the figure, indeed the interior looks whiter than the background. There is also a stratification of figures, the triangle emerges and seems to be above the disks. This type of phenomenon is classified by Kanizsa as *modal completion*.

On the other hand missing parts of a figure can be completed, creating a perception that is phenomenologically indistinguishable from real stimuli. Since the completed figure is perceived with the modality of vision, the underlying process is called "modal completion". It gives rise to the well known phenomenon of illusory boundaries (or subjective contours).

It often takes place in order to complete occluding objects (in Fig. 2.6 right, the completed triangle is occluding the 3 circles).

A point made clear by the studies of phenomenology in perception is that in all circumstances of completion both the occluding and the occluded objects are perceived at the same time in the scene and therefore there are points in the input stimulus corresponding to more than one figure at the perceptual level. This suggests that the phenomenological space has a higher dimension than that of the physical space, in this case the two dimensional image.

2.4 Association Fields of coaxial elements

A central idea of the Gestalt psychology is that continuity is an essential notion in perception. More recently, partially motivated by this fact, cognitive neuroscience has renewed its attention in the representation of contours and their integration. Indeed, over the last twenty years hundreds of papers have been published in the fields of neurophysiology, psychophysics and computer sciences, combining approaches to elaborate models of how contours are perceived and integrated [28].

In order to study the perceptual integration of contours Field, Hayes, and Hess introduced in [29] two types of stimulus. Both types of stimulus consisted of a smooth path of spatially separated elements, embedded in a field of randomly oriented elements. In one type of stimulus (snakes or co-axial), the elements are arranged tangentially to a path. In the other type (ladders or trans-axial), the elements are orthogonal to the path.

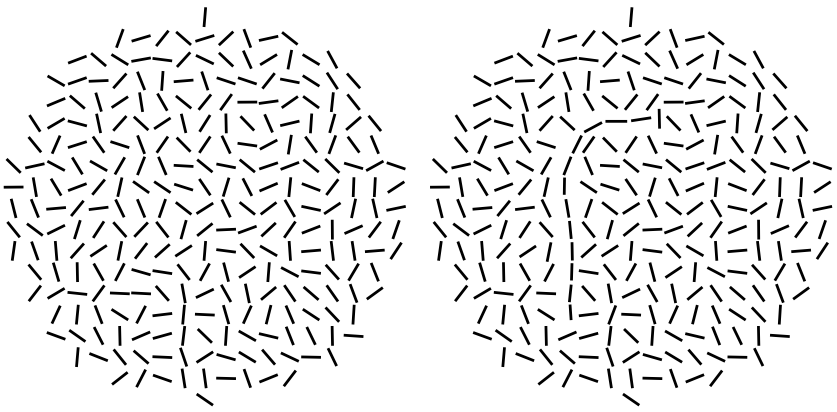


Figure 2.7: The 2 figures show grids made of oriented segments. In the left figure all the segments are randomly oriented and there is not perception of any structure while in the grid in the right figure some of them are aligned. A perceptual structure emerges and this effect is known as “pop out” or perceptual saliency.

Particularly in [29] is exploited the concept of association fields for coaxial elements. The basic idea is simple: if we consider a set of randomly distributed small segments, there is not perception of any kind of structure, but, when some of the elements are aligned a contour emerges from the background (see Fig. 2.8). A classical result from the Gestalt theory, the law of good continuation. The segments tend to be perceptually grouped and made

salient if they form a continuous contour. Earlier works studied the problem of detecting contours composed of dots or short line segments (called test elements) embedded in a field of similar elements of random position and orientation (as in Fig. 2.7). The problem of these experiments, as pointed out in [36], is that their nature (wide spatial and orientation spectrum) leaves open the possibility that simple filtering is able to resolve the task of integrating the test elements into a single contour. Therefore, the oriented elements selected in [29] were Gabor like functions (referred to as Gabor patches) which have a dot like spectrum and they approximately match the tuning properties of the simple cells in V1 (this point will be detailed in the next two chapters).

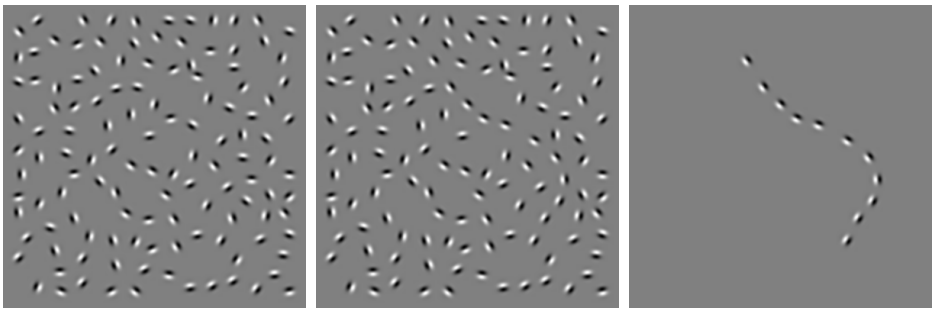


Figure 2.8: Test images of psychophysical experiments, similar to those used by Field, Hayes and Hess [29]. The experiment consists of showing to subjects a grid made of oriented Gabor patches. In some images, the grid contains elements aligned over a path and other similar patches randomly distributed. In the other cases, each oriented element is placed randomly. The task for the participant is to detect whether there are or not aligned elements in the grid (a forced choice between two alternatives 2AFC).

The experiment tested the ability of subjects to detect a target composed of small Gabor patches aligned to form a smooth contour with the *two-alternative forced choice* (2AFC) method. The 2AFC is a classical methodology used to measure human performance in psychophysical tasks, where stimuli with different target intensities have to be discriminated [47]. It signifies that an observer should indicate which one of two stimuli presented in a trial contains a target, with the target presented in only one of the two stimuli. The target was embedded in a large field of similar oriented random elements. The distances between the elements were larger than the dimension of a single element (see Fig. 2.8).

They found that the target detectability depended on several factors, most importantly, the relative orientation of adjacent contours elements. The

structure is perceived correctly if the patches are aligned tangentially to the path and the angle variation between two adjacent elements is not too large. The perception is described as a phenomenon of "pop-out" or perceptual saliency. Targets could be detected if the orientation of adjacent contours elements are small. The effect was robust over a range of element densities and presentation times.

Based upon these results Field et al. suggested that local interactions between contour elements follow specific rules and form the basis for contour integration in humans. These constraints form a local association field that govern how different oriented contours should interact to form a coherent group (see Fig 2.9).

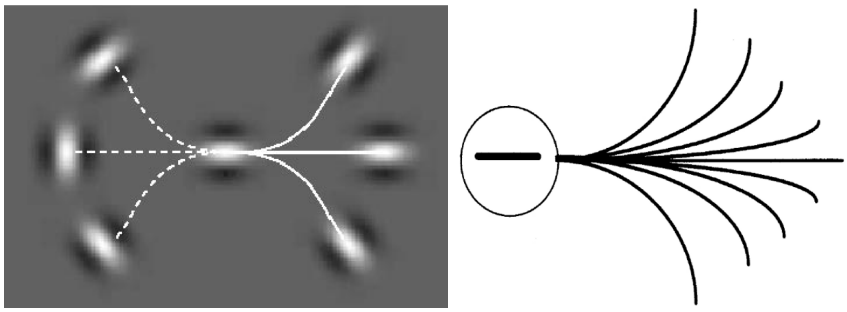


Figure 2.9: In the left figure the solid line indicates a configuration between the patches where the association exists while the dashed line indicates a configuration where it does not. The right figure shows the association field of Field, Hayes and Hess (images taken form [29]).

2.5 Association Fields of trans-axial elements

A similar but weaker effect in detectability was observed when elements were oriented perpendicular to the contour (ladder or transaxial). In a recent experiment [57], the authors carefully compared the two different types of perceptual integration mechanisms using 2AFC. The stimulus used by the authors consisted of images of odd-symmetric Gabor elements embedded into a grid, as in the previous experiment. In that experiment the snake target consisted of a configuration where the Gabor elements are arranged tangentially to a path. On the contrary, in the ladder target, the axis of the Gabor patches are perpendicular to the path (see Fig 2.10 bottom). Besides

the elements in the path there are other randomly distributed and oriented patches on the grid.

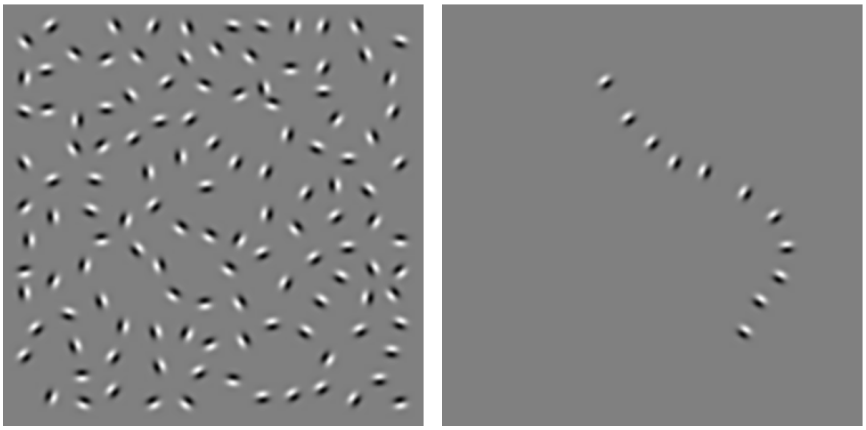


Figure 2.10: Test image of the ladder experiment, similar to those used in [29] or in [57].

Previous works demonstrated that ladders are usually harder to identify than snakes [29, 37, 8, 52, 56], but [57] was the first work to test systematically a set of parameters: the Gabor element separation, the path angle and the carrier wavelength, by using a sampling of the parameter space.

The 2AFC procedure followed by the authors in [57] consisted of presenting sequences of trials to a subject. Each sequence corresponded to a particular combination of carriers wavelength, element separation and contour type (snake or ladder). The separation between elements s is measured in degrees of the visual field and takes values between 1.09°, 1.54°, 2.18° and 3.08°. The carrier wavelength cw (measured in the same units) takes values between 0.193°, 0.273°, 0.385° and 0.545°. In a single session only the path angle α was varied and took values between 0°, 10°, 30° or 40°. The paths considered were created following the co-circularity rule: two adjacent elements in a contour are tangent to the same arc of circumference (see Fig. 2.11).

Two subjects performed the experiment. Each one was presented with a sequence of 40 trials for each combination of parameters. Each trial consisted of two stimuli presented sequentially, one containing a contour and the other only random elements. The order in which the stimuli were presented was determined randomly. The subjects were asked to indicate which stimuli had the contour, and received an auditory feedback to indicate whether the response was correct or not. Figure 2.12 summarizes the results of the experiment relating the performance level with respect to elements separation.

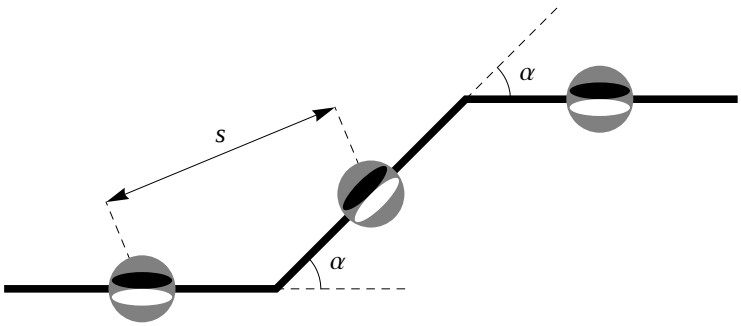


Figure 2.11: A schematic representation of part of a snake-type target used in the experiment from [57]. Here the Gabor patches are positioned tangentially to the path while in the ladders the elements are orthogonal to it. The thick solid line represents the backbone of the contour. The difference in orientation and the separation between the patches are the angle α and the scalar s respectively. The parameter s and α are chosen in such a way that the patches are tangential to the same arc of circumference.

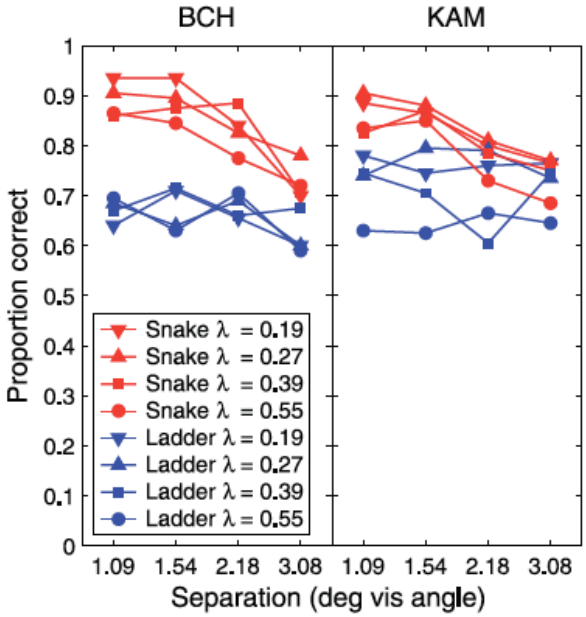


Figure 2.12: Performance levels collapsed across path angle. The different lines on the graphs show the data for snakes and ladders with different carrier wavelengths (image taken from [57]).

As a conclusion of the experiment they observed this interesting finding, summarizing an important distinction between snakes and ladders :

> *"increasing the separation between the elements had a disruptive effect on the detection of snakes but had no effect on ladders, so that as separation increased, performance on the two types converged"*

We will try in the following chapters to model the set of phenomenological behavior presented in this section with a suitable mathematical framework. To setup the framework we are going to be motivated by the neural architecture of the visual cortex, as it is outlined in the next chapter.

Chapter 3

The visual cortex

3.1 Introduction

The objective of this chapter is to introduce the basic structures of the functional architecture of the primary visual cortex (V1). The main idea is that neural computations strictly depend on the organization and connectivity of neurons in the cortex. We will consider here only the structures that are relevant to the models presented in the later chapters, those involved in the boundary coding. Receptive fields and receptive profiles of simple cells will be introduced as basic elements for boundary coding. The three fundamental structures implemented in the neural circuitry, namely the layered, retinotopic and the hypercolumnar structure will be described from the neurophysiological and functional point of view. Then the pinwheel structure will be introduced as the real topological implementation of the hypercolumnar structure. Finally the connectivity pattern between simple cells will be considered. The functional and physiological descriptions are only qualitative and almost every quantitative detail is deliberately avoided. There are many sources in literature where the readers can find information about this topic, starting from the very seminal but in-depth work of Hubel and Wiesel [42, 43, 41] to the recent manuscript by Jean Petitot [74].

3.2 The cerebral cortex and the visual pathway

The cerebral cortex is the outer sheet of neural tissue on the two cerebral hemispheres, also called the "gray matter". Most of the neurons associated

with sensory or cognitive processing are located in the cortex. The rest of the cerebral cortex mainly consists of axons connecting cortical neurons to each other, it is called the "white matter". The cortex is commonly described as being constituted by three parts: sensory, motor, and association areas. The sensorial areas are the parts of the cortex that receive sensory inputs. The visual cortex is the area that serves the sense of vision and receives the optical information from the visual path.

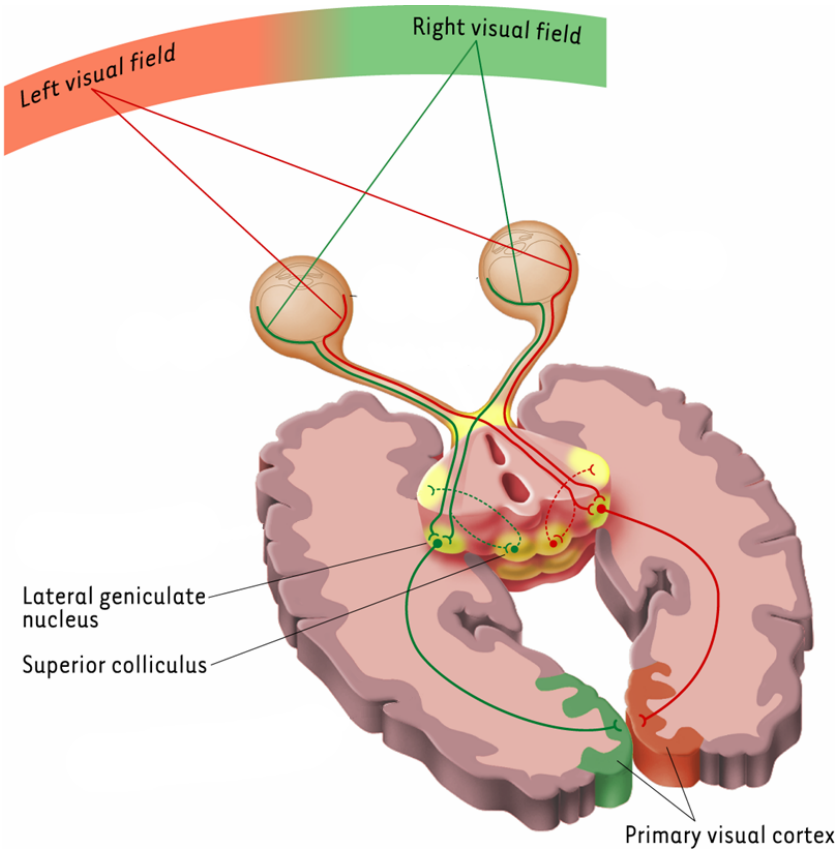


Figure 3.1: The visual pathway. The optical signal is converted into electrical signals by photo-receptors. The electrical signal is transmitted into the brain through the retino-geniculo-cortical pathway (image adapted from [35]).

Fig. 3.1 illustrates the visual pathway. Light enters the eye, reaching the retina. The retina is a curved, thin sheet of brain tissue that grows out into the eye to provide the starting point for the neural processing of visual signals. The retina is covered by more than a hundred million photoreceptors, which convert the light into electrical signals. From the photoreceptors, the signal is transmitted through a couple of neural layers. The last of the retinal

processing layers consist of ganglion cells, which send the output of the retina (in form of action potentials) away from the eye using their very long axons. The axons of the ganglion cells form the optic nerve. The optic nerve transmits the visual signals to the lateral geniculate nucleus (LGN) of the thalamus, a structure in the middle of the brain through which most sensory signals pass on their way from the sensory organs to the main sensory processing cortical areas. From the LGN, the signal goes to various destinations, the most important being the visual cortex at the back of the head, where most of the visual processing is performed.

In humans, the visual cortex contains approximately a fifth of the total cortex, which reflects the importance of visual processing in us. It consists of a number of distinct areas. The primary visual cortex (V1) is the area to which most of the retinal output first arrives and is the most widely-studied visual area.

3.3 Simple cells in V1

3.3.1 Receptive field and receptive profile

The receptive field of a visual neuron is the area of visual field in which visual stimulation influences neural responses. Considering that the visual field is reflected on the retina, it is often defined as the region of the retina where the action of light alters the firing of the neuron. Some of these cortical cells respond to light and dark spots in different sub-regions of the receptive field (called ON and OFF zones respectively) and the arrangement of these subregions can be used to predict the responses of the cell to visual stimuli. Then, we are able to define the receptive profile (RP) of a neuron as its transfer function if we think it as a filter. The RP is a function $\varphi(x, y)$ measuring the response of the cell (positive values for the ON zones and negative for the OFF zones), $\varphi : D \to \mathbb{R}$, where the domain D is the receptive field, and (x, y) are retinal coordinates.

3.3.2 The action of receptive profile on the stimulus

Visual neurons act on the visual stimulus in a very complex way including non trivial temporal dynamics, non linear responses to light intensity and contextual modulation accounting for non local behaviors. For our purpose we will consider just the so called classical receptive field [17] acting on the

stimulus with linear and local behavior as in the following. Let $I(x, y)$ be the optical signal defined on the retina (or equivalently the visual field) and $\varphi(x - x_0, y - y_0)$ be the RP of a neuron defined in a domain D centered on (x_0, y_0). As we previously mentioned, the neuron acts on I as a filter, and it computes the mean value of I on D weighted by φ:

$$I_\varphi(x_0, y_0) = \int_D I(x', y')\varphi(x' - x_0, y' - y_0)dx'dy'.$$

The response of the neuron I_φ can be interpreted as a weighted measure at the point (x_0, y_0) of the signal I. Therefore, if there is a set of neurons with RP φ covering the whole retina, a convolution between the RP and the input I is performed:

$$I_\varphi(x, y) = \int_D I(x', y')\varphi(x' - x, y' - y)dx'dy' = (I * \varphi)(x, y).$$

3.3.3 Simple cells in V1 as orientation detectors

Using electrode recordings on cats visual cortex, Hubel and Wiesel in [42] provided the first characterization of receptive fields in V1 based on their responses. Reverse correlation techniques enabled the precise recording of the RPs. In there, light and dark spots are presented at different positions of the receptive field and their responses is measured (the firing rate of the neuron). Afterwards, the correlation between the inputs and the outputs yields the transfer function of the neurons. A review of reverse correlation techniques can be found in [77]. Cells with separate subregions that respond to either light or dark spots are called simple cells. All other cells in the visual cortex that do not have separate ON/OFF zones (the majority of the cells) are called complex cells [54].

Many researchers noted that the RPs of simple cells have a remarkable resemblance to oriented partial derivatives of gaussians (DoG) up to a third order (see for example [49]). Then, if:

$$\begin{cases} \tilde{x} = x\cos\theta + y\sin\theta \\ \tilde{y} = -x\sin\theta + y\cos\theta \end{cases} \tag{3.1}$$

where the angle $\theta \in S^1$ is the orientation of the symmetry axis of the filter, an idealized model for odd-symmetric RP is (Fig. 3.2):

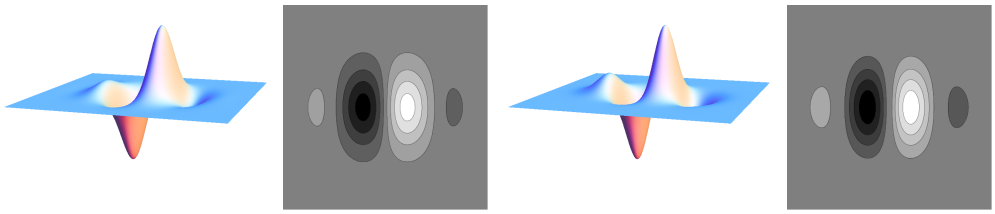
$$\varphi(x, y) = \frac{\partial^3 G_\sigma}{\partial \tilde{y}^3}(x, y) \tag{3.2}$$

and for an even-symmetric RP (Fig. 3.3):

$$\varphi(x,y) = \frac{\partial^2 G_\sigma}{\partial \tilde{y}^2}(x,y), \tag{3.3}$$

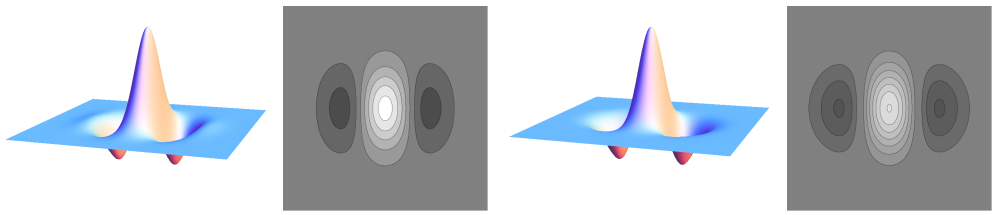
where G_σ is the 2D gaussian kernel:

$$G_\sigma(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}. \tag{3.4}$$



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

Figure 3.2: The graph (a) and some level curves (b) of a third partial derivative of gaussian (eq (3.2)). The graph (c) and some level curves (d) of the imaginary part of a 2D Gabor filter (eq 3.5). Both are classical models of simple odd-symmetric orientation cell.



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

Figure 3.3: The graph (a) and some level curves (b) of a second partial derivative of Gaussian (eq (8.9)). The graph (c) and some level curves (d) of the real part of a 2D Gabor filter (eq 3.6). Both are classical models of simple even-symmetric orientation cell.

Another classical model for the V1 RPs are the Gabor filters, proposed by Daugman [16] or Jones and Plamer [45] amongst others. A Gabor filter is a sinusoid modulated Gaussian:

$$\Psi_{\sigma,\omega,\theta}(x,y) = e^{i\omega\tilde{y}} G_\sigma(x,y),$$

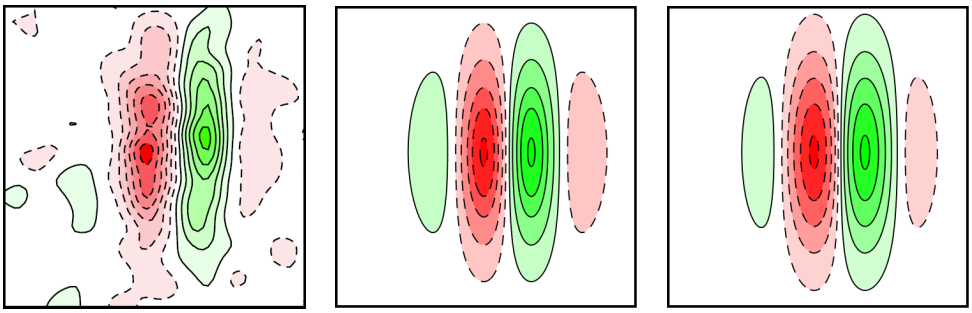
where the imaginary part models an odd-symmetric RP (Fig. 3.2b):

$$\varphi(x, y) = \mathrm{Im}(\Psi_{\sigma,\omega,\theta}) = \sin(\omega\tilde{y})G_\sigma(x, y) \tag{3.5}$$

and the real part an even one (Fig. 3.3b):

$$\varphi(x, y) = \mathrm{Real}(\Psi_{\sigma,\omega,\theta}) = \cos(\omega\tilde{y})G_\sigma(x, y). \tag{3.6}$$

Qualitatively the two models are similar even if they have very different properties from the theoretical point of view. In Daugman [16] the receptive profiles of simple cells have been obtained as a minimum of the uncertainty principle of the Heisenberg group, while in [83] they result as a minimum of the uncertainty principle in the group invariant under rotations and translations. Within this thesis the DoG model will be mostly used. In Fig. 3.4a there is shown a RP of a V1 simple cell belonging to a cat (the image was taken from [17]). In Fig. 3.4b one can see its fitting with a DoG filter of a third order and in Fig. 3.4c with an odd Gabor kernel.

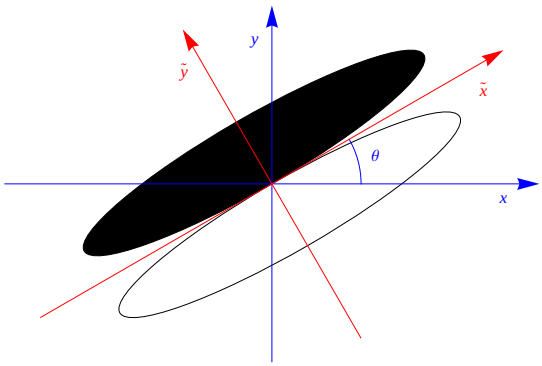


(a) (b) (c)

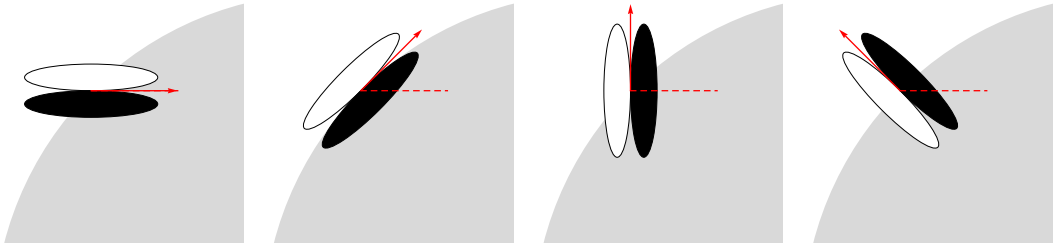
Figure 3.4: The receptive profile of a simple cell of V1. In red are visualized the positive contrast areas (ON) while in green the negative (OFF). (a) The level curves recorded in vivo with electrophysiological techniques and reconstructed by reverse correlation, the size of the image corresponds to 5 degrees of the visual field (this image was taken from [17]). (b) Fitted with a third derivative of Gaussian and (c) with the imaginary part of a Gabor kernel.

Most of the V1 simple cells are functionally involved in visual processing as orientation detectors. This means that at a certain scale, their response is a measure of the local orientation of the stimulus at a certain retinal point. The angle in which the response of the cell is maximal is called the orientation preference (OP) of the neuron. A simple cell sensing a boundary

fires maximally when the orientation of its axis is aligned with the boundary itself. Nevertheless a broad set of cells with suboptimal orientation respond to the stimulus. Then the cortex is equipped with a neural circuitry that is able to sharpen orientation tuning. With this mechanism, called non-maximal suppression, the output of the cells with suboptimal orientation is suppressed, allowing just a small set of cells optimally oriented to code for boundary orientation.



(a)



(b)

Figure 3.5: A set of simple cells schematically represented sensing of a boundary. The output is maximum when the axis (red arrow) is tangent to the boundary, but it is also not nulled for a broad set of sub-optimal orientations.

3.4 The functional architecture of V1

We refer to the spatial organization and connectivity between neurons inside a cortical area as the functional architecture. In the case of V1, three main

structures were identified by neurophysiologists: the layered, the retinotopic and the hypercolumnar structures.

3.4.1 The layered structure

The *layered* structure indicates that the cortex is formed by 6 horizontal layers and a number of sublayers. The sublayer 4C is where most of the axons from the LGN arrive and were the concentration of oriented cells is higher.

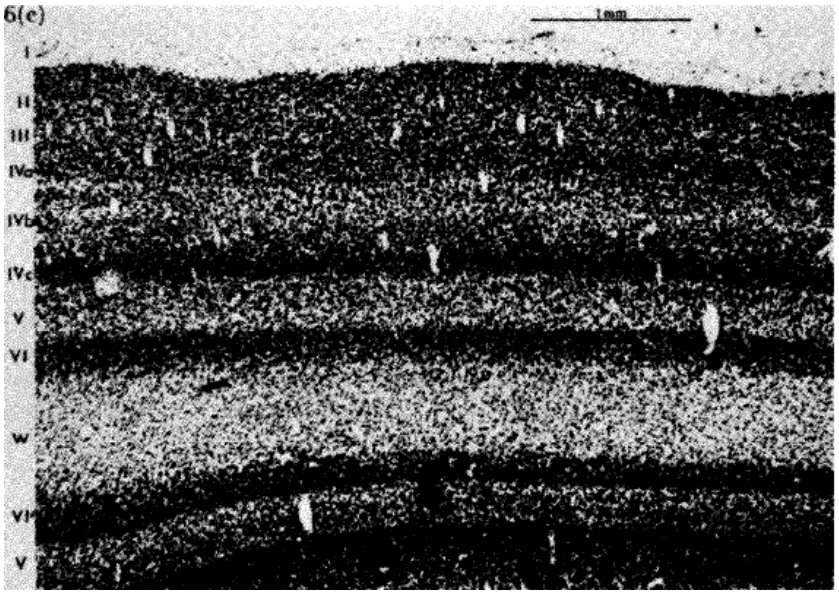


Figure 3.6: The layered structure of the primary visual cortex as steined and photographed by Nissl (image taken from [41]). The picture reveals quite clearly the different layers I to VI.

3.4.2 The retinotopic mapping

The *retinotopy* is a particular kind of topographic organization implying that there exists a topology preserving mapping from the retina to the cortex. The position of the center of the simple cell receptive fields form an ordered sampling mosaic that covers a portion of the visual field. Because of this orderly arrangement, which emerges from the spatial specificity of connections between neurons in the retina and in the cortex, cells in each structure can be seen as forming a map of the visual field (also called a retinotopic map, or a visuotopic map). In other words, what is near in the retina is near in the

cortex. Retinotopic maps are found in many mammalian brains. The specific size, number, and spatial arrangement of these maps differ considerably between species. From a quantitative point of view, a model for the mapping well fitting the experimental data is based on a logarithmic conformal map [1].

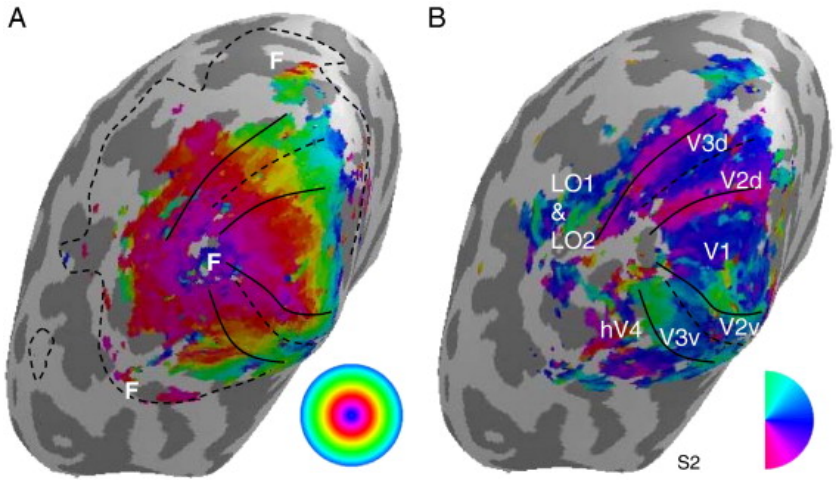


Figure 3.7: The retinotopic structure of the primary visual cortex recorded by FMRI (image taken from [68]). The mapping is visualized on a reconstructed anatomical volume. On the left the radial mapping is shown while on the right the polar angle is visualized.

3.4.3 The hypercolumnar structure

Hubel and Wiesel discovered that while penetrating the cortex perpendicularly and recording the RPs of the simple cells, their OPs remain constant, no matter the depth. Not only the OP, but also the ocular dominance, the direction of movement preference and other features are similar in each column. This is why the cortex can be thought of as a 2D layer with respect to orientation coding. A tangential penetration in the superficial layers of the cortex reveals that the RPs of cells close to each other strongly overlap while the OP varies smoothly[1] generating the orientation hypercolumnar structure. Fig 3.8 shows the classical cube model of Hubel an Wiesel [43] that summarizes the experimental results.

[1]the only exceptions are singular points where cells around them are arranged radially varying their preferred angle, the so called pinwheels.

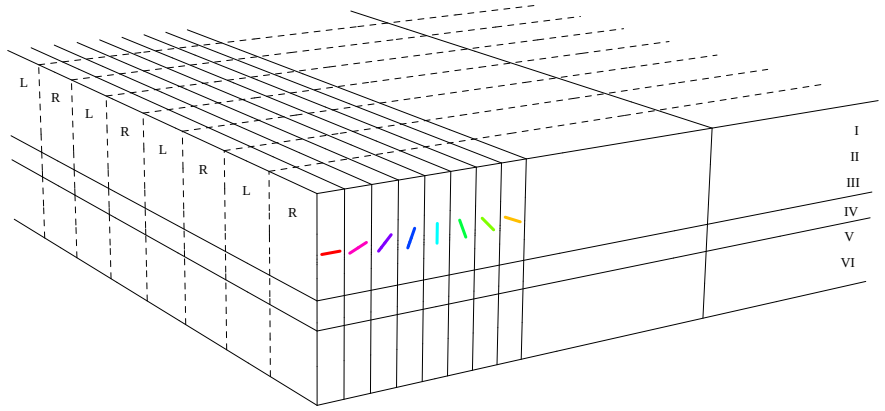


Figure 3.8: A simplified version of the classical Hubel and Wiesel cube scheme of the primary visual cortex (cytochrome-oxidase blobs are not shown). Cells belonging to the same column share similar RP characteristics (almost identical receptive fields, same OP and ocular dominance). The orientation hypercolumns are arranged tangentially to the cortical sheet. Moving across the cortex the OP varies while the RP strongly overlap. Oriented bars colored with a polar color codes are used to represent the OPs (the hue represent the angle).

The cortical structure is largely redundant. It means that at a certain scale and resolution, for each point of the retina (x, y) there exists a whole set of neurons in V1 responding maximally to every possible local orientation θ. Since ideally the position on the retina takes values in the plane $\mathbb{R}^2$ and the orientation preference in the circle S^1, the visual cortex domain can be locally modelled as the product space $\mathbb{R}^2 \times S^1$. Each point (x, y, θ) of this 3D space, represents a column of cells in the cortex associated to a retinal position (x, y), all of which are tuned to the orientation given by the angle θ.

Fig 3.9 shows a schematic representation of the visual cortex. The hypercolumns are draw vertically. The different colors represent different orientations. The coordinates (x, y, θ) of this 3D space isomorphic to $\mathbb{R}^2 \times S^1$ are the parameters of the RPs: (x, y) is the retinotopic position and θ the tuning angle.

The fundamental consideration here is that V1 is modelled as the 3D space of positions and orientations, while as mentioned earlier the cortex is essentially a 2D layer. Therefore, a dimensional reduction problem must be faced. The strong redundancy of the cortex (receptive fields of close neurons highly overlap) allows for the codifying 3D information in a 2D structure. The visual

system has implemented the dimensional collapse in the pinwheel structure described bellow.

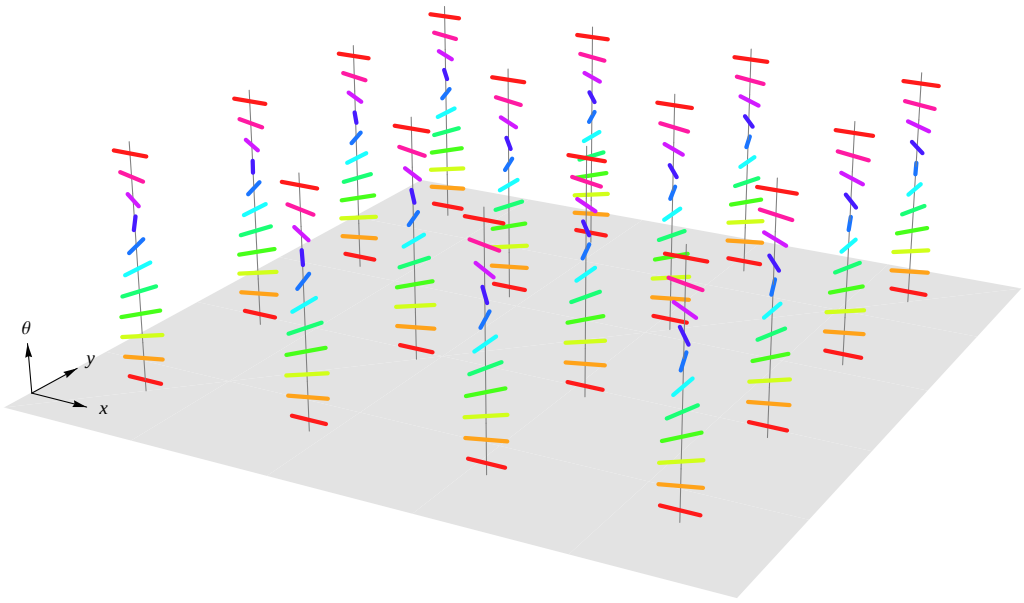


Figure 3.9: The visual cortex modelled as a set of hypercolumns. Over each retinotopic point (x, y) there is a set of cells coding for the set of orientations $\{\theta \in S^1\}$ and generating the 3D space $\mathbb{R}^2 \times S^1$. Each bar represents a possible orientation. The color coded map is the same one used in Fig. 3.8.

3.4.4 The pinwheel structure

3.4.4.1 The experiment of Bonhöffer and Grinvald

As we have seen, neurons lying below each other in the visual cortex share the same orientation tuning among other functional properties. Across the cortical surface, the OPs change gradually forming an orientation map. In order to determine the precise layout of the iso-orientation domains of the cortex, the *in-vivo intrinsic-signal optical imaging* techniques were developed in the early nineties by Bonhöffer and Grinvald [9]. This brain imaging technique introduced a real revolutionary progress within the study of the cortical organization and function. It allowed for the acquisition of activity images for the population of cells from the superficial layers of V1 revealing the global map of orientations. Previously the only acquisition techniques were the recordings of a few individual neurons using micro-electrodes or the

2-deoxyglucose method which permits post-mortem visualization of brain activity areas. A very illustrative analogy with events that occurred within meteorology was produced by Petitot in [74]. Passing from the recording of individual cells to the activity of a cortical area can be compared to changing from sending balloons to the use of satellite images.

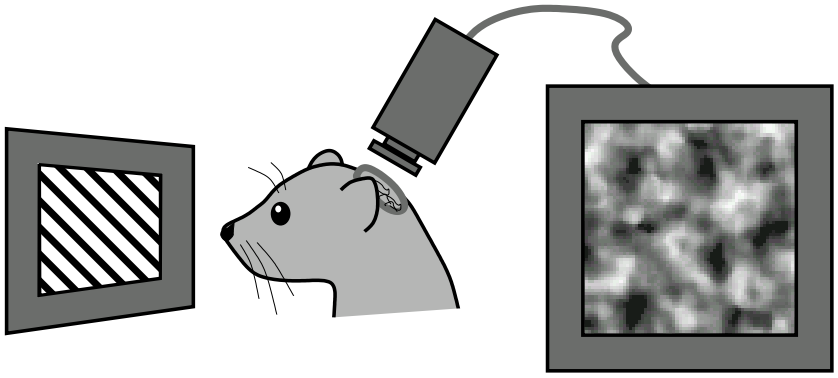


Figure 3.10: Optical imaging (image taken from [59]). Images are taken from the exposed cortex of the animal, which is illuminated with an infrared light (605 nm). The images are acquired with a video camera while the animal is visually stimulated with moving gratings projected onto a screen. The signal to noise ratio of the functional maps is improved by averaging several stimulus sessions.

The experiment of Bonhöffer and Grinvald is described as follows (a detailed technical description of the method can be found in [34]). High contrast oriented and moving gratings consisting of square wave images with constant frequency at a given angle (translating in both directions along the wave vector) are presented to an animal as a visual stimulus. A hole in the crane above its primary visual cortex is opened and the window is illuminated with an infrared light. Small but different absorptions of light patterns are observed depending on the grating's orientations and are registered with an optical video camera. The intensity of the images recorded depends on the neural activity of the neuron population under each point. The images may be improved therefore averaging an increased realization of the experiment. A simple post processing of these images is made by taking the difference between patterns corresponding to orthogonal gratings, cutting the low frequencies and normalizing. Thus, precise iso-orientation maps, i.e. the areas that best respond to one orientation, are obtained. Fig. 3.11 from [15] shows the patterns obtained in a real experiment for eight different orientations. The areas in dark forming the round patches correspond to regions where the

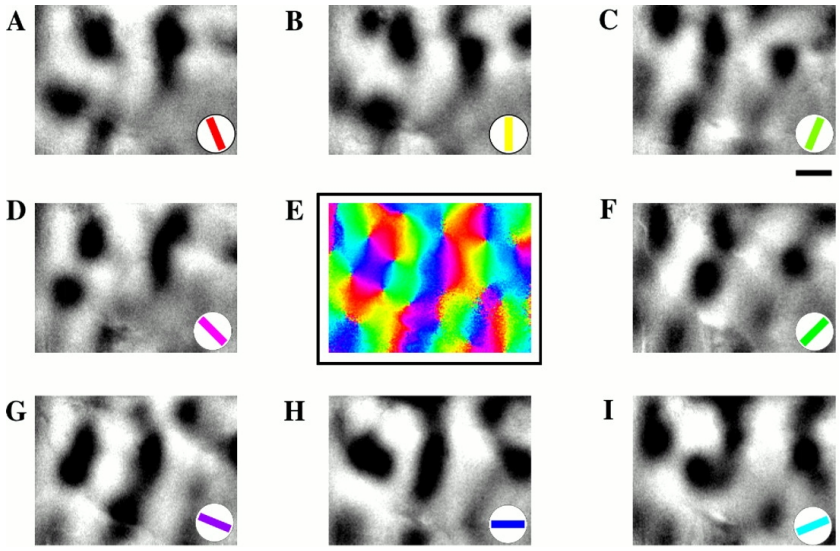


Figure 3.11: Orientation preference maps (image taken from [15]). The gray level images show the cortical responses to stimulation at the orientation indicated by the colored bars. Dark regions correspond to strong neuronal response. The central panel is obtained combining all the eight images; the color indicates the preferred orientation.

response to the stimulus is maximal. In the corner of each image, an oriented colored bar indicates the orientation of the grating.

All the orientation maps can be combined using a color scale (the central image in Fig. 3.11). Each color corresponds to the best orientation stimulus. In Fig. 3.12 the orientation maps of a significant portion of a tree shrew's V1 is shown (from [11]). In the figure one is able to identify 2 emerging structures:

- regular zones where the color changes gradually and obviously correspond to the orientation hypercolumns recorded with the electrodes.
- singular points where all colors appear once and only once around them. These points arranged like spokes of a wheel are called pinwheels. All the orientations are present radially varying near the center in both clockwise and counter-clockwise form.

The orientations recorded belong to the interval $[0, \pi)$, seeing as the technique does not permit for the distinction of cells with opposite contrasts (this becomes clear just by looking at the set of possible stimulus). This implies

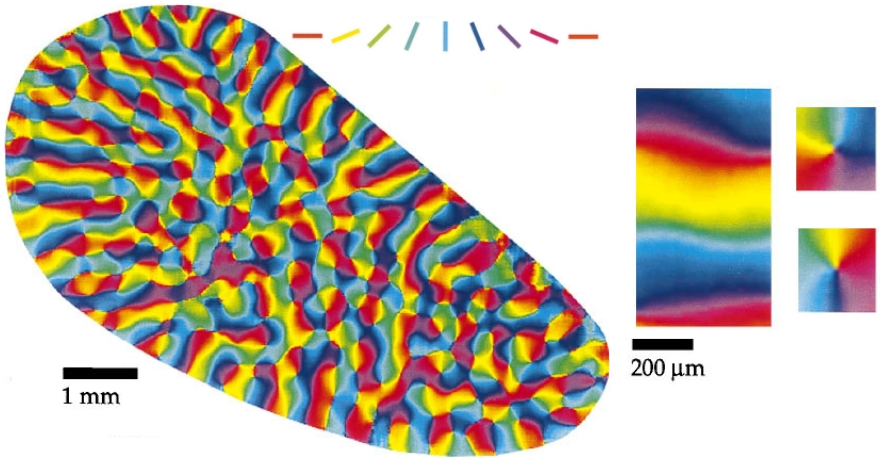


Figure 3.12: Orientation preference maps of tree shrew's visual cortex (image taken from [11]). Orientation preference of each location is color-coded as usual. Portions of the orientation preference map shown on the left are enlarged to demonstrate that the orientation preference maps contained both linear zones and pinwheel arrangements.

that points opposed with respect to the pinwheel's center have perpendicular orientations.

3.4.4.2 The structure of pinwheels at a fine scale

Conventional optical imaging has a lateral spatial resolution of 50 μm at best, and a very limited depth resolution, so it is not capable of resolving the fine scale structure of the pinwheel centers. Below this spatial scale, conventional optical imaging is not able to decide whether neurons selective for different orientations were arranged randomly near the pinwheel center or might be perfectly segregated. Therefore the question of whether the pinwheels really exists at a neuronal scale or whether they were a consequence of the acquisition technique remained open for many years. The two-photon calcium imaging technique based on confocal microscopy allowed Ohki et Al [66] to analyze the functional maps at single-cell resolution in three dimensions. Firstly, preferred orientation maps of a cat's visual cortex were recorded with intrinsic-signal optical imaging in order to target the injection of a calcium indicator dye to a pinwheel center. This injection typically labelled thousands of neurons near the pinwheel. The confocal imaging allowed for the simultaneous measurement of the visual responses of tens to hundreds of labelled neurons within a given optical cross-section parallel to the cortical surface. This measurements were

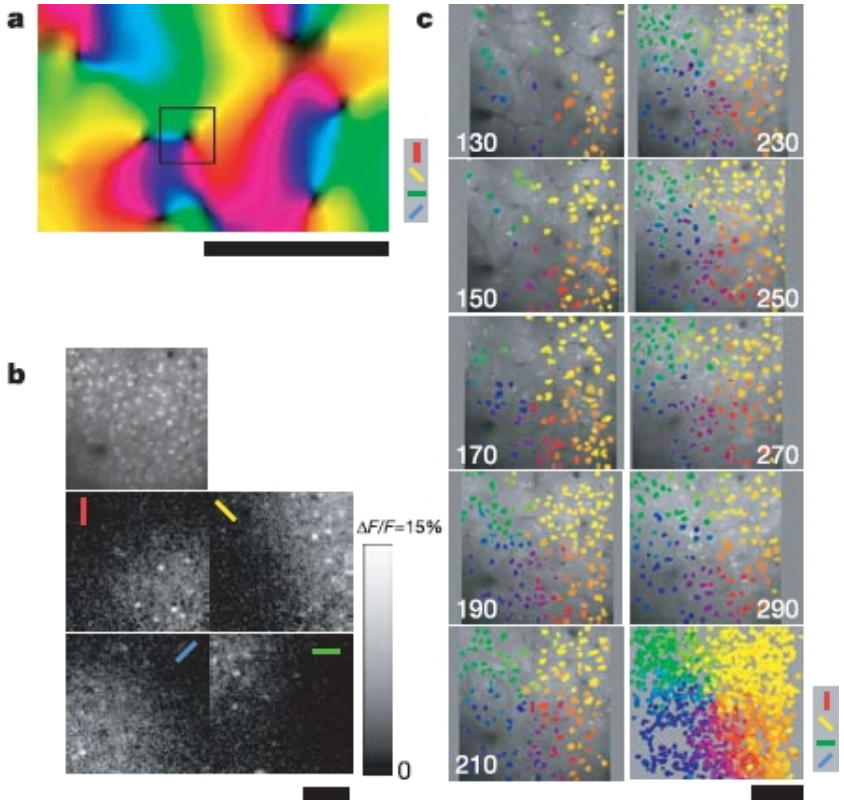


Figure 3.13: Pinwheels in 3D at a fine-scale (image taken from [66]). (a) An orientation map obtained with the classical optical signal. As usual hue corresponds to the best orientation. Darker colors in the pinwheel centers represent less selective responses. b) Two-photon calcium imaging. The square region is about $250\mu m$ deep. The four bottom panels show the maps relative to the orientations indicated. The upper panel is the average image. (c) Cell-based orientation maps from nine different depths indicated in μm. The last panel shows the overlay of images from all nine depths.

repeated at different depths. The results obtained are presented in Fig. 3.13.

This experiment definitively confirmed that pinwheels exist as a real neurophysiological structure.

3.4.5 The pattern of lateral connectivity

In previous sections we have seen how the visual system breaks up the visual signal and represents it in arrays of neurons, each one coding an individual component of the image. An interesting question is how these neurons communicate to each other.

To this aim, a fundamental neurophysiological finding was the discovery of the lateral connectivity in the cortex, also called horizontal or intrinsic connectivity. These experiments (see for example [11, 33]) revealed the existence of connections parallel to the cortical surface that run several millimeters (6 to 8 mm in the visual cortex). These long range connections allow for communication between cells of similar OP belonging to different hypercolumns with non overlapping receptive fields. The linked cells not only share the angle of tuning, but also the axis corresponding to the orientation is roughly the same. The experiment performed by D. Fitzpatrick and his collaborators provided a particularly illustrative result (see Fig 3.14). They injected a chemical tracer into a small area of the visual cortex of a tree shrew. The tracer was propagated through the lateral connections and the resulting image was combined with the orientation maps obtained with optical imaging. It can be observed that in the immediate vicinity of each neuron, the connections are relatively isotropic, but over larger distances they closely follow the OPs.

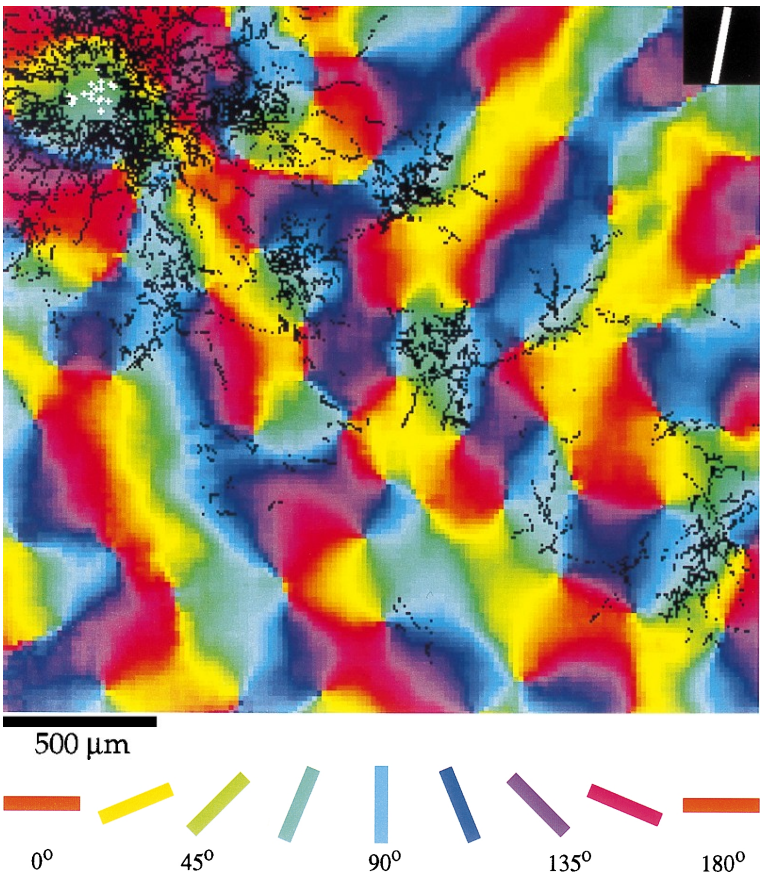


Figure 3.14: Long range connections (image taken from [11]). The white dots indicate where the chemical tracer (biocytin) has been injected. The tracer is propagated through the lateral connections to points shaded in black. These locations are plotted together with the orientation map.

Chapter 4

A cortical model in SE(2)

4.1 Introduction

This section is devoted to mathematically modelling the cortical space introduced in the previous section. We are mainly interested in the structure of the cortex, which is responsible for the functionality of the cortex itself.

We start by reviewing the definition and some basic properties of the Lie group theory, which take into account the symmetry of the cortex. Then, we recall that the hypercolumnar structure of the primary visual cortex has been modelled as a fiber bundle in a seminal paper [39] by W. Hoffman, developed by J. Petitot and Y. Tondout in [75, 73] as a jet bundle (with the Heisenberg symmetries) and by G. Citti and A. Sarti [82, 14] as the principle fiber bundle of the Lie group SE(2). In this Chapter the last model is presented.

Each hypercolumn of simple cells, defined at a retinal point (x, y), forms a mono-dimensional group of rotations. This structure is identically repeated for every point of the retina, and can therefore be considered as a fiber of the 3-dimensional principal bundle. This bundle can be identified with a cotangent bundle, as its elements are orientation detectors. In this description the neural process of maxima selection due to intra-cortical short range inhibitory connections lifts boundaries or curves present on the retinal plane to curves in V1, and whole retinal images to surfaces in V1. Then a differential structure is introduced into this group, in order to model the long range horizontal connections between different hypercolumns. This structure is formalized as the Lie algebra with a sub-Riemannian structure (see [82, 14]). The integral curves of its generating vector fields will be interpreted as a mathematical representation of the long range connection (see 3.4.5 in the previous section)

and their 2D retinal projection will model the association fields of Field, Hayes and Hess (see [29]).

4.2 The notion of Lie Group and Tangent Space

In this section we provide some basic definitions of the Lie group theory, as it is the essential framework utilized in this thesis. All definitions can be found in standard mathematical textbooks (for example [69]). A very accessible version for non mathematicians can be found in J. Koenderink's book [50].

Definition 1. *An internal operation $\cdot$ on a set G is a function which associates to each couple of elements (a, b) of G another element denoted as $a \cdot b$ of G. This condition is also called **Closure**: for all a, $b \in G$, $a \cdot b \in G$. To qualify the set G as a **group**, the operation must satisfy three requirements known as the group axioms:*

- ***Associativity.*** *For all a, b, $c \in G$, $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.*
- ***Identity element.*** *There exists an element $e \in G$, such that for all $a \in G$, $e \cdot a = a \cdot e = a$.*
- ***Inverse element.*** *For all $a \in G$, there exists an element $a^{-1} \in G$ such that $a \cdot a^{-1} = a^{-1} \cdot a = e$, where e is the identity element.*

Groups fulfilling the property of commutativity: for all a, $b \in G$, $a \cdot b = b \cdot a$, are called **Abelian** or **commutative**.

A **subgroup** is a subset G' of G satisfying the group axioms.

The distinguishing feature of a Lie group is that it also carries the structure of a smooth manifold, so that the group operation is continuous.

Definition 2. *A **Lie group** is a group which also carries the structure of a manifold in such a way that both the group operation*

$$m : G \times G \to G, \;\; m(g, h) = g \cdot h, \;\; g, h \in G,$$

and the inversion

$$i : G \to G, \;\; i(g) = g^{-1}, \;\; g \in G,$$

are smooth maps between manifolds.

Examples of Lie Groups are :

- the Euclidean space $\mathbb{R}^n$, with the usual sum as group law;
- the set of square matrices $n \times n$, with determinant different from 0. In this set we consider the standard product of matrices, and the existence of inverse is ensured by the condition on the determinant. Note that this group is not commutative;
- the circle S^1 of angles mod 2π, with the standard sum of angles;
- the group of rotations and translations on the plane to be described in detail in the following section.

Definition 3. *Let C be a smooth curve on a manifold $\mathcal{M}$, parameterized by $\gamma : I \to \mathcal{M}$, where I is a subinterval of $\mathbb{R}$. In local coordinates $x = (x^1, \ldots, x^n)$, C is given by n smooth functions $\gamma(s) = (\gamma^1(s), \ldots, \gamma^n(s))$ of the real variable s. At each point $x = \gamma(s)$ of C, the curve has a* ***tangent vector***

$$\boldsymbol{X}|_x = \dot{\gamma}(s) = \dot{\gamma}^1(s)\frac{\partial}{\partial x^1} + \cdots + \dot{\gamma}^n(s)\frac{\partial}{\partial x^n}.$$

By this definition it becomes clear that tangent vectors can be identified with directional derivatives in the direction:

$$\vec{X}|_x = (\dot{\gamma}^1, \ldots, \dot{\gamma}^n)$$

if applied to smooth functions $f : \mathcal{M} \to \mathbb{R}$. In particular

$$\mathbf{X}f = < \vec{X}, \nabla f > .$$

Within this thesis we will identify differential operators and vectors, hence use indifferently the two notations.

Definition 4. *The collection of all possible tangent vectors to all possible tangent curves passing through a point $x \in \mathcal{M}$ is called the* ***tangent space*** *to $\mathcal{M}$ at x, and is denoted as $T\mathcal{M}|_x$.*

Definition 5. *An* ***integral curve*** *of a vector field $\vec{X}$ is a smooth parametrized curve $x = \gamma(s)$ whose tangent vector at any point coincides with the value of $\vec{X}$ at the same points for all s:*

$$\dot{\gamma}(s) = \vec{X}|_{\gamma(s)}.$$

An integral curve is called **maximal integral curve** if it is not contained in a longer integral curve.

Definition 6. *For two differential operators* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$*, their* ***Lie bracket*** *(or* ***commutator****) is defined by their actions on functions* $f : \mathcal{M} \to \mathbb{R}$*:*

$$[\boldsymbol{X}, \boldsymbol{Y}]f = \boldsymbol{X}(\boldsymbol{Y}f) - \boldsymbol{Y}(\boldsymbol{X}f).$$

Note that the Lie Bracket is a measurement of the non-commutativity of the operators: it is defined as the difference of applying them in reverse order. In particular $[X, Y]$ is identically 0 if X and Y commute.

If $\mathbf{X}$ and $\mathbf{Y}$ are first order operators:

$$\begin{aligned}\mathbf{X} &= a_1\partial_{x^1} + \cdots + a_n\partial_{x^n}\\ \mathbf{Y} &= b_1\partial_{x^1} + \cdots + b_n\partial_{x^n}\end{aligned}$$

a simple computation ensures that:

$$[\mathbf{X}, \mathbf{Y}]f = \left(\mathbf{X}b_1\partial_{x^1} + \cdots + \mathbf{X}b_n\partial_{x^n} - \mathbf{Y}a_1\partial_{x^1} - \cdots - \mathbf{Y}a_n\partial_{x^n}\right) f$$

is a first derivative.

Definition 7. *Let* G *be a Lie group. For any element* $g \in G$*, we define the* ***left-multiplication*** *(or* ***left-translation****)* $L_g : G \to G$ *by:*

$$L_g(h) = g \cdot h, \quad \text{for all } h \in G,$$

where $\cdot$ *denotes the group operation in* G*.*

Definition 8. *An operator* $\boldsymbol{X}$ *on* G *is called* ***left-invariant*** *if:*

$$\boldsymbol{X}(f o L_g) = (\boldsymbol{X}f) o L_g, \qquad \text{for all } g \in G.$$

Definition 9. *The* ***Lie algebra*** *of a Lie group* G *is the vector space of all left-invariant vector fields on* G*.*

Intuitively the Lie algebra associated to a Lie group encodes its differential structure, and it is identified as the tangent space at the origin.

4.3 The visual cortex is a fiber bundle

As proposed by Hofmann [39], the mathematical structure ideally modelling the retinotopic and hypercolumnar structure is called the fiber bundle or fibration.

In the simplest case[1] a fiber bundle is defined by two differentiable manifolds B, called the base manifold of the fiber bundle and F, called the fiber of the bundle. In this simplest case, the total space of the fiber bundle will be identified by $B \times F$. In addition, within the fiber bundle there is also defined a natural projection

$$\pi : B \times F \to B.$$

In [75, 73], Petitot and Tondut reconsidered the fiber bundle structure and proposed a cortical model based on jet-bundles. The bundle is associated to the 1-form:

$$-pdx + dy,$$

where $p \in \mathbb{R}$ is the angular coefficient of the tangent to the boundaries. This theory leads to Lie symmetries of a Heisenberg group. In this formal framework the so called base space of the fibration is the retina and there is a map associating it to each retinotopic position $(x, y) \in \mathbb{R}^2$ a fiber which is a copy of the whole possible set of angular coefficients in the plane $\{p \in \mathbb{R}\}$.

In [82, 14], Citti and Sarti proposed a model of the principle fiber bundle of the $SE(2)$ group. In this context the 1-form characterizing the bundle can be expressed as:

$$-\sin\theta dx + \cos\theta dy$$

This bundle became the standard method for describing the cortical space also used in the book of J. Petitot [74].

[1] A fiber bundle is defined by two differentiable manifolds M and C, a group G, and a projection π. C and M are called, respectively, base space and total space. The total space is locally described as a cartesian product $C = M \times G$, meaning that to every point $m \in M$ is associated a whole copy of the group G, called the fiber. The function π is a surjective differential map, which locally acts as:

$$\pi : M \times G \to M, \;\; \pi(m, g) = m,$$

where g is an element of G. In the model presented here, the base space is implemented in the retinal space and the total space in the cortical space. Particularly the group G of rotations to the point $m = (x, y)$ is implemented in an hypercolumn over the same point.

Fig 3.9 can be now interpreted as a schematic representation of the visual cortex rearranged as a fiber bundle. The hypercolumns are draw vertically and are called fibers. The different colors represent different orientations.

4.3.1 The cotangent bundle

As mentioned before, the functionality of the odd receptive fields has been interpreted as gradient indicators and two similar models were described: the derivative of Gaussian and the Gabor filters. Let us choose the Gabor model for the following deductions (for the case of DoG computations are straightforward). The imaginary part of a Gabor filter with orientation θ has the expression:

$$\varphi(x,y) = \frac{1}{2\pi\sigma^2}\sin(k\tilde{y})e^{-\frac{\tilde{x}^2+\tilde{y}^2}{\sigma^2}}$$

where

$$\begin{cases} \tilde{x} = x\cos\theta + y\sin\theta \\ \tilde{y} = -x\sin\theta + y\cos\theta \end{cases}.$$

The map $(x,y) \to (\tilde{x},\tilde{y})$ is a rotation of angle θ. Then, the function φ can be approximated (up to a multiplicative constant) by

$$\frac{2\sin(k\tilde{y})}{k\sigma^2}e^{-\frac{\tilde{x}^2+\tilde{y}^2}{\sigma^2}} \simeq \frac{2\tilde{y}}{\sigma^2}e^{-\frac{\tilde{x}^2+\tilde{y}^2}{\sigma^2}} = \partial_{\tilde{y}}e^{-\frac{\tilde{x}^2+\tilde{y}^2}{\sigma^2}}.$$

A derivative in the direction $\tilde{y}$ can be expressed in the original variables (x,y,θ) as a directional derivative in the direction of the vector $(-\sin\theta,\cos\theta)$:

$$\partial_{\tilde{y}}I = -\sin\theta\partial_x I + \cos\theta\partial_y I = <(-\sin\theta,\cos\theta),\nabla I>$$

This derivative applied to a function I expresses the projection of the gradient in the direction $(-\sin\theta,\cos\theta)$. Since the gradient is an element of the tangent plane, the vector $(-\sin\theta,\cos\theta)$, which acts on it will be considered as an element of the cotangent plane, and represented as 1- form

$$-\sin\theta dx + \cos\theta dy. \tag{4.1}$$

The quantity $<(-\sin\theta,\cos\theta),\nabla I>$ will be a maximum when $(-\sin\theta,\cos\theta)$ is parallel to ∇I or equivalently when it is perpendicular to a level line of the image I (see Fig. 4.3). Therefore, a simple cell with receptive field centered at the point (x,y) can be identified with a 1-form which selects the direction θ of a boundary at this point.

4.4 The visual cortex as the $SE(2)$ group

The Rototranslation group is the fundamental mathematical structure used in this thesis. In the literature it is also known as the *2D Eculidean motion group* $SE(2)$. It is the 3D group of rigid motions in the plane or equivalently the group of elements invariant to rotations and translations. The aim of this section is to show that the visual cortex at a certain level is naturally modelled as the Rototranslation group with a sub-Riemannian metric.

4.4.1 The group law

In the previous section we have described the functional architecture of the visual cortex and in particular we have seen how the fibration resulting from the vector product $\mathbb{R}^2 \times S^1$ is the space of parameters of simple V1 cells. The parametrization is given by their retinotopic position (x_1, y_1) and their orientation preference θ. Another way of thinking with regards to this space is illustrated in Fig. 4.1. The half-white/half-black circles represent oriented receptive profiles of odd simple cells, where the angle of the axis is the angle of tuning. Every possible receptive profile is obtained from a mother kernel by translating it of the vector (x_1, y_1) and rotating over itself by an angle θ.

We denote T_{x_1,y_1} the translation of the vector (x_1, y_1) and R_θ a rotation matrix of angle θ:

$$\mathbf{R}_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Then a general element of the $SE(2)$ group is of the form $A_{x_1,y_1,\theta} = T_{x_1,y_1} \circ R_\theta$, and applied to a point (x, y) it yields:

$$A_{x_1,y_1,\theta_1} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + \mathbf{R}_{\theta_1} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$$

All the profiles can be interpreted as: $\varphi(x1, y1, \theta_1) = \varphi_0 \circ A_{x1,y1,\theta_1}$ The set of parameters $g_1 = (x_1, y_1, \theta_1)$ form a group with the operation induced by the composition $A_{x_1,y_1,\theta_1} \circ A_{x_2,y_2,\theta_2}$. This turns out to be:

$$g_1 \cdot g_2 = (x_1, y_1, \theta_1) +_R (x_2, y_2, \theta_2) = \left(\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \mathbf{R}_{\theta_1} \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \right)^T, \theta_1 + \theta_2 \right)$$

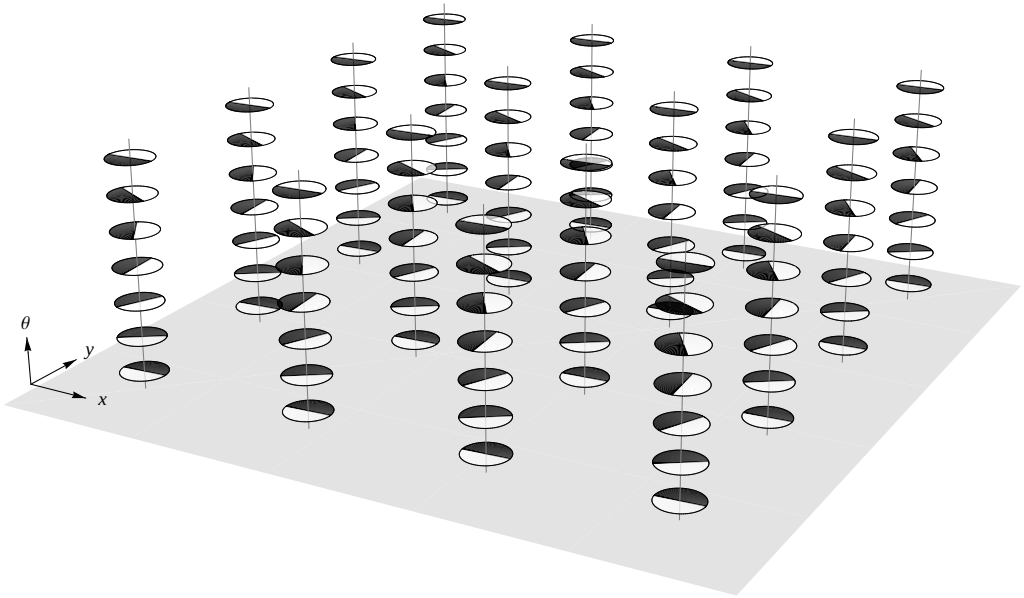


Figure 4.1: The visual cortex modelled as the group invariant under translations and rotations.

Being induced by the composition law, one can easily check that $+_R$ verifies the group operation axioms, where the inverse of a point $g_1 = (x_1, y_1, \theta)$ is induced by the rototranslation:

$$A^{-1}_{x_1,y_1,\theta} = R^{-1}_{\theta} \circ T^{-1}_{x_1,y_1},$$

and the identity element is given by the trivial point $e = (0, 0, 0)$. Then, the group generated by the operation $+_R$ in the space $\mathbb{R}^2 \times S^1$ is called the Rototranslation group or equivalently $SE(2)$.

A structured space with the symmetries described above allows for the cortex to be invariant to rotations and translations in the representation of a retinal image: the signals will be identical no matter what their position or orientation in the phenomenological space is.

4.4.2 The differential structure: horizontal plane and lifting

What distinguish a Lie group from a topological group is the existence of a differential structure. In the case of V1, Citti and Sarti in [14] proposed to endow the $\mathbb{R}^2 \times S^1$ with a sub-Riemannian structure. In the standard

Euclidean setting, the tangent space to $\mathbb{R}^2 \times S^1$ has dimension 3. They selected a bi-dimensional subset of the tangent space at each point, called the horizontal plane, as a model of the connectivity in V1. In the sequel we describe how to define the horizontal plane.

4.4.2.1 Horizontal plane

In the previous section we identify each Gabor filter with an 1-form (see 4.1) on the $\mathbb{R}^2$ plane. This form can be lifted to the cotangent space of $\mathbb{R}^2 \times S^1$ into the 1-form:

$$-\sin\theta dx + \cos\theta dy.$$

The horizontal tangent space will be identified by the kernel of this form, or in other words by the set of vectors orthogonal to the vector that we will denote $\vec{X_3}$:

$$\vec{X_3} = (-\sin\theta, \cos\theta, 0).$$

Its kernel is a bi-dimensional plane (see Fig. 4.2) called the horizontal plane and is generated by the first order operators $\mathbf{X}_1$ and $\mathbf{X}_2$:

$$\mathbf{X}_1 = \cos\theta\partial_x + \sin\theta\partial_y, \qquad \mathbf{X}_2 = \partial_\theta.$$

The choice of these 2 vectors from the whole kernel as generators of the horizontal plane will become clear in the next subsection. Indeed the tangent plane is by definition the plane of the vector fields tangent to the curves of the space. We will see that the only curves of the space are integral curves of the two vector fields X_1 and X_2, while there is no natural curve with a non vanishing component in the direction X_3. Then this has not to be considered a tangent direction.

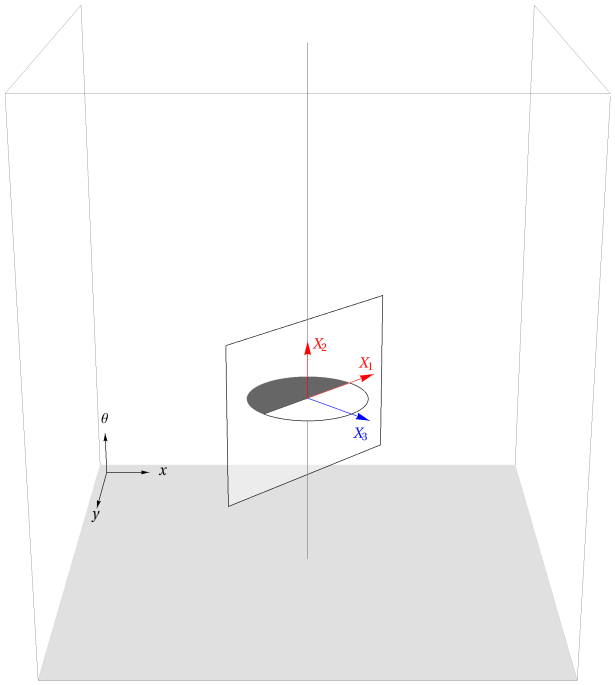


Figure 4.2: A schematic representation of a simple cell of V1 where the vectors $\vec{X}_i$ are indicated.

4.4.2.2 Lifting of curves

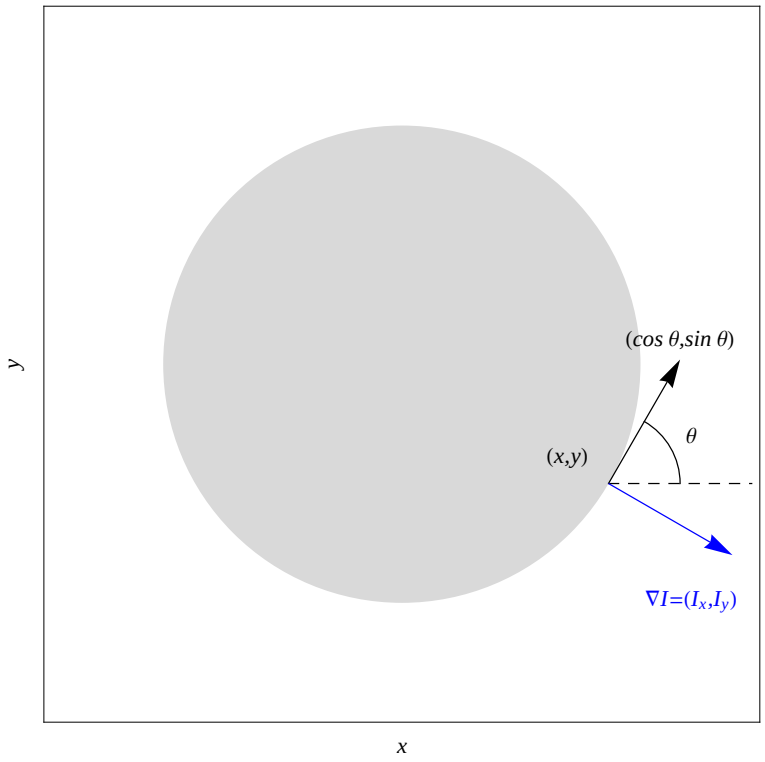


Figure 4.3: A contour in a 2D image can be modelled as a curve whose tangent is the vector $(\cos\theta, \sin\theta)$ as indicated in the figure.

A contour may be represented in the 2D plane as a regular curve

$$\gamma_{2D}(t) = (x(t), y(t)),$$

and we can assume that its tangent vector is non vanishing almost everywhere, so that it can be identified by an orientation $\theta(t) : D \subset R \to S^1$. This indicates that we are able to parametrize the curve by its arc-length

$$(\dot{x}(t), \dot{y}(t)) = (\cos(\theta(t)), \sin(\theta(t))). \tag{4.2}$$

The function θ takes values on the whole circle, in order to represent polarity of the contours: two contours with the same orientation but with opposite contrasts are represented by opposite angles on the unit circle. The action of the receptive profiles is to associate to every point $(x(t), y(t))$ the orientation

$\theta(t)$. In this way the two dimensional curve γ_{2D} is lifted to a new curve $\gamma(t)$ in the 3D space:

$$(x(t), y(t)) \to (x(t), y(t), \theta(t)). \tag{4.3}$$

We will call admissible curve a curve in $\mathbb{R}^2 \times S^1$ if it is the lifting of an edge (identified with a planar curve).

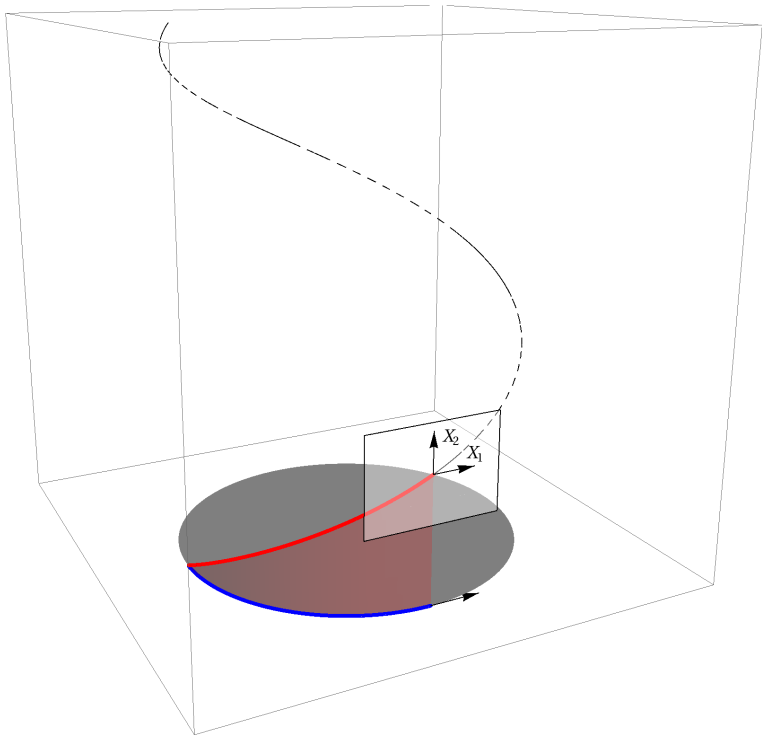


Figure 4.4: A contour represented by the blue curve $\gamma_{2D}(t)$ is lifted into the rototranslation group obtaining the red curve $\gamma(t)$. The tangent space of the rototranslation group is the span of the vectors $\vec{X_1}$ and $\vec{X_2}$.

In Fig.4.4 we illustrate the lifting process. By the parametrization we have chosen in formula (4.2) for the curve γ_{2D} (the blue curve in Fig. 4.4) we can immediately express the value of θ:

$$\theta = \arctan\left(\frac{\dot{y}}{\dot{x}}\right). \tag{4.4}$$

The lifting γ of the curve γ_{2D} is defined in (4.3), and depicted in Fig. 4.4 (the red curve). By definition it can be expressed by (x, y, θ), where

$$\dot{\gamma}(t) = (\dot{x}(t), \dot{y}(t), \dot{\theta}(t)) = (cos(\theta(t)), \sin(\theta(t)), \dot{\theta}(t)) = \vec{X_1}(t) + \dot{\theta}(t)\vec{X_2}(t).$$

It follows immediately that $\dot{\gamma}(t)$ has a non vanishing component in the direction $\vec{X_1}$ and a second component θ' in the direction of $\vec{X_2}$. In particular admissible curves are integral curves of two vector fields in a 3D space, and cannot have components in the orthogonal direction $\vec{X_3}$.

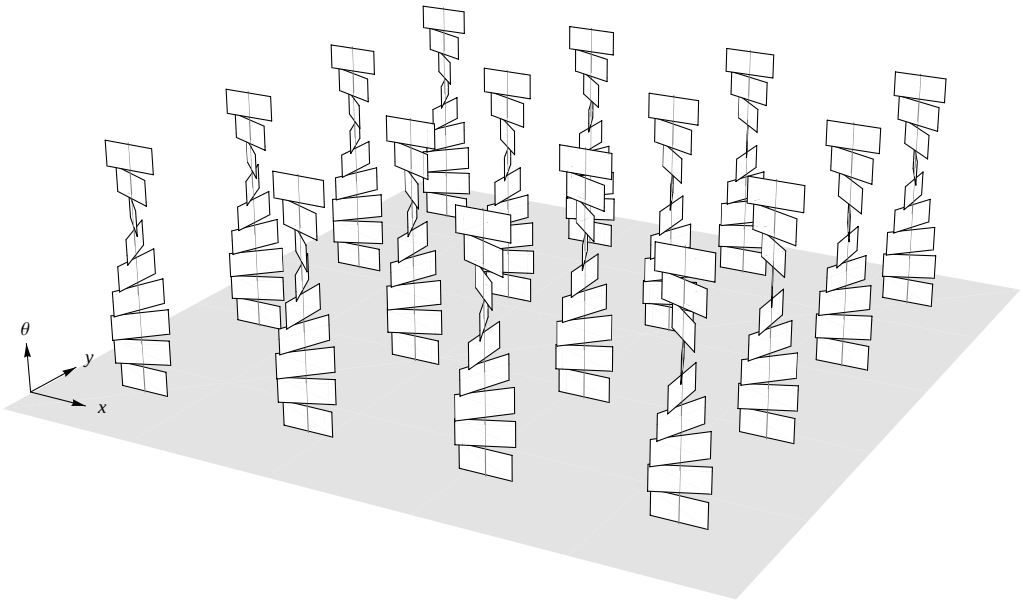


Figure 4.5: The horizontal tangent planes in each point of the rototranslation group determine the differential structure of the space.

As seen in the previous section, the horizontal plane is the span of the vectors $\vec{X_1}, \vec{X_2}$. An admissible curve in this group is an integral curve of the vector field $\vec{X_1}$, $\vec{X_2}$ and is defined as the solution for the following ordinary differential equation:

$$\dot{\gamma}(t) = \vec{X_1}(t) + k(t)\vec{X_2}(t). \tag{4.5}$$

Writing the same equation componentwise we get:

$$\begin{cases} \dot{x} = \cos\theta \\ \dot{y} = \sin\theta \\ \dot{\theta} = k(t). \end{cases} \tag{4.6}$$

It is well known that $k(t)$ is the curvature of $\gamma_{2D}(t)$. Writing the curve this way it become obvious that the shape of the curve is completely defined by the function k: its curvature or angular velocity.

4.4.3 The sub-Riemannian metric

4.4.3.1 The Lie algebra

We explicitly note that the vector fields $\vec{X_1}$, $\vec{X_2}$ and $\vec{X_3}$ are left invariant with respect to the group law of rotations and translations, so that they are the generators of the associated Lie algebra.

In the standard Euclidean setting, the tangent space to $\mathbb{R}^2 \times S^1$ has dimension three at each point. Here, only a two dimensional section of the tangent space is selected, the horizontal tangent space, generated by the vector fields $\vec{X_1}$ and $\vec{X_2}$.

4.4.3.2 The choice of the metric

Recall that a metric is simply the choice of the length of any vector of the tangent space; hence once our tangent space has been defined, we can immediately perform a choice of the metric. We will call the norm of the vector $\alpha_1\vec{X_1} + \alpha_2\vec{X_2}$

$$||\alpha_1\vec{X_1} + \alpha_2\vec{X_2}|| = \sqrt{\alpha_1^2 + \alpha_2^2}.$$

The metric induced by this choice is clearly Non-Euclidean, moreover it is not even Riemannian at any point, considering we do not prescribe the length of the vector X_3. A Reimannian approximation would be

$$||\alpha_1\vec{X_1} + \alpha_2\vec{X_2} + \epsilon\alpha_3\vec{X_3}|| = \sqrt{\alpha_1^2 + \alpha_2^2 + \epsilon\alpha_3^2}$$

and it is clear that we recover that previous, as ϵ goes to 0. This Riemannian metric can be represented as

$$diag(1, 1, 1/\epsilon)$$

in the basis $\{\vec{X_1}, \vec{X_2}, \vec{X_3}\}$, and its co-metric in the same basis is:

$$diag(1, 1, \epsilon).$$

Hence we can define a co-metric for the limit sub-Riemannian metric as

$$diag(1, 1, 0),$$

with respect to the same basis. Hence the space co-metric in the canonic basis will therefore become:

$$(g^{ij}) = \begin{pmatrix} \cos\theta & 0 \\ \sin\theta & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \cos^2\theta & \cos\theta\sin\theta & 0 \\ \cos\theta\sin\theta & \sin^2\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Given that the matrix g^{ij} is not invertible, it can not induce a Riemannian metric in the space. Spaces equipped with Sub-Riemannian metrics appear often when one of the dimensions is a state variable depending on the others. In this instance the state variable is θ.

4.4.4 Connectivity property and distance

Once equipped the tangent planes with an Euclidean metric, the length of an admissible curve (represented as in equation (4.6)) can be computed by integrating the tangent vector, as usual:

$$\lambda(\gamma)(t) = \int_0^t \|\gamma'(s)\| ds = \int_0^t \sqrt{1+k^2} ds. \tag{4.7}$$

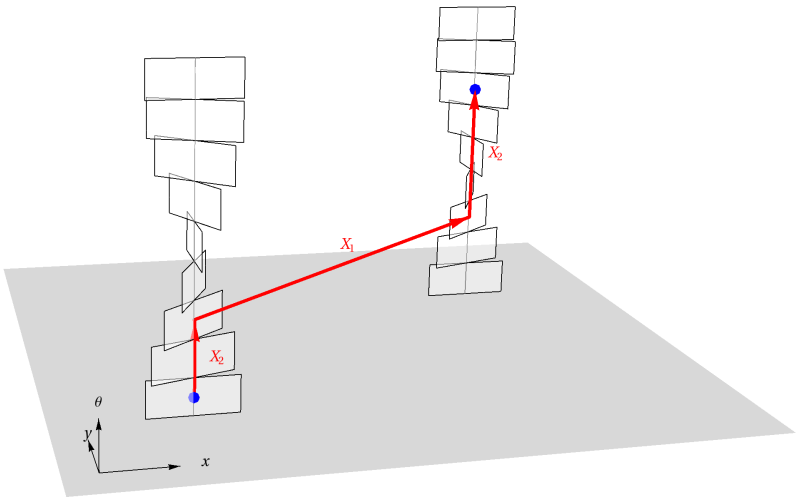


Figure 4.6: The figure shows a trivial way to reach any point of the $SE(2)$ moving along the integral curves of vectors belonging to the horizontal plane.

In order to define a distance in terms of the length, we need to answer the following question: *Is it possible to connect each couple of points of* $\mathbb{R}^2 \times S^1$ *using an admissible curve?*

This is not a simple question taking into account that at each point we have only the directions which are linear combinations of two vectors even if we are immersed in a three dimensional space. However, the answer is yes and it will become clear in the example shown in Fig 4.6. It is shown how trivial it is to arrive from an arbitrary point to another by just using admissible curves curves. For simplicity let us put the starting point at the origin $(0, 0, 0)$ and the end point (x, y, θ). We can move first in the direction of $\vec{X}_2$ until we arrive at $(0, 0, \alpha)$ with $\alpha = \arctan(x/y)$, so the vector $\vec{X}_1$ points to the fiber over the point (x, y). Once we arrive to the fiber we are at the point (x, y, α) and we can move again in the direction of $\vec{X}_2$ until we get to the point (x, y, θ).

The possibility of connecting each couple of points with an admissible curve, is called the connectivity condition. We have directly proved this condition, but we can deduce it from some general results strictly related to the non commutativity of vector fields. Indeed there is a differential relationship between X_1, X_2, and the vector field X_3, and this can be obtained as a commutator of the previous two:

$$X_3 = [X_1, X_2] = -\sin\theta\partial_x + \cos\theta\partial_y.$$

This is a particular case of the well known Hörmander condition:

Definition 10. *We say that the Hörmander condition is satisfied if X_1, X_2 and their commutators of any order span the Euclidean tangent space at every given point.*

In the present case this condition is clearly satisfied, if X_1, X_2 and $X_3 = [X_1, X_2]$ are linearly independent and span the tangent space to $\mathbb{R}^2 \times S^1$ at each point. Hence we can apply the Chow theorem, which ensures that if the Hörmander condition is satisfied, then the connectivity condition also holds (see [14] and the references therein).

This is exactly what we already see: any couple of points can be joined with an admissible curve, an integral curve of X_1 or X_2.

Consequently, it is possible to define a notion of distance between two points $p_0 = (x_0, y_0, \theta_0)$ and $p_1 = (x_1, y_1, \theta_1)$:

$$d(p_0, p_1) = \inf\{\lambda(\gamma) : \gamma \text{ is an admissible curve connecting } p_0 \text{ and } p_1\}. \quad (4.8)$$

In the Euclidean case this infimum is realized by a geodesic that is a segment. Here, the geodesics are locally curvilinear. The functional in (4.7) has the

same asymptotic behavior as the modified Elastica functional introduced by Nitzberg, Mumford and Shiota in [65]:

$$\phi(k) = \begin{cases} \nu + \alpha k^2 & \text{for } |k| < \beta/\alpha \\ \nu + \beta|k| & \text{for } |k| \geq \beta/\alpha \end{cases}$$

where α and β are two scalar weights. The above functional is quadratic when the curvature tends to zero and linear when it tends to infinity. The same functional is used in perceptual completion problems to complete occluded parts, to perform image inpainting or to retrieve subjective contours [14, 64, 55, 5, 84]. The minimizer of the functional allows for the kinks in the asymptotic behavior.

4.4.5 Integral curves and horizontal connections

The expression of the integral curves of the structure was given in by eq. (4.6). Choosing the curvature k constant over time, one can integrate the system of equations to obtain an explicit formula for the curve $\gamma(t)$ starting from the origin $(x_0, y_0, \theta_0) = (0, 0, 0)$:

$$\begin{cases} x = \frac{\sin(kt)}{k} \\ y = \frac{1-\cos(kt)}{k} \\ \theta = kt \end{cases} . \tag{4.9}$$

For each value of the parameter k a different curve with helicoidal path is obtained, since the projection over the plane is a circle with curvature k and it moves with constant speed in direction θ. On Fig.4.7 some integral curves in the 3D space of positions and orientations $\mathbb{R}^2 \times S^1$ for different values of k are shown. The projection on the space $\mathbb{R}^2$ is visualized in Fig.4.8.

It is clear that this projection is in good agreement with the association fields of Fields, Hayes and Hess [29], as well as with the pattern of lateral connectivity discovered by D. Fitzpatrick and his collaborators (see Section 3.4.5).

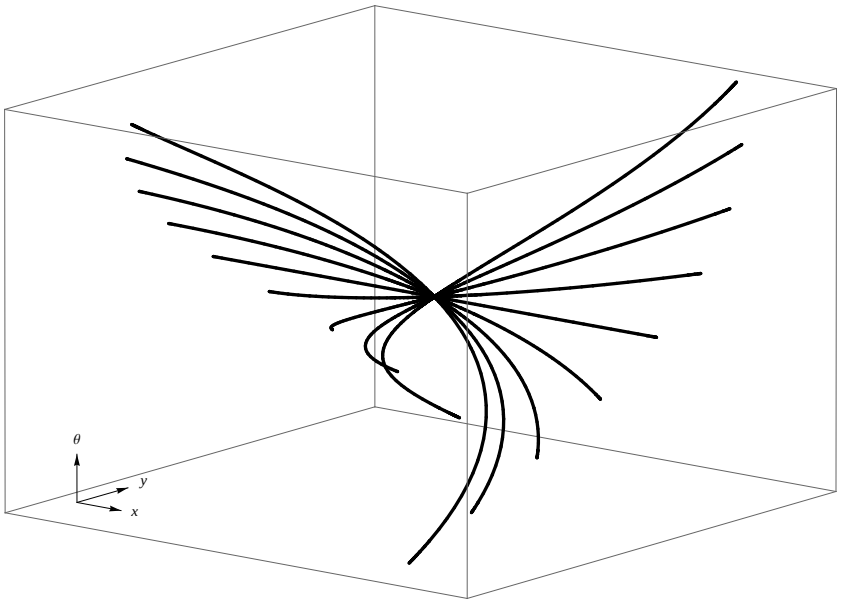


Figure 4.7: The fan of integral curves given by eq. (4.9) by varying the curvature k, visualized in the $\mathbb{R}^2 \times S^1$ space. The set of curves defines a surface with a singular point at the origin.

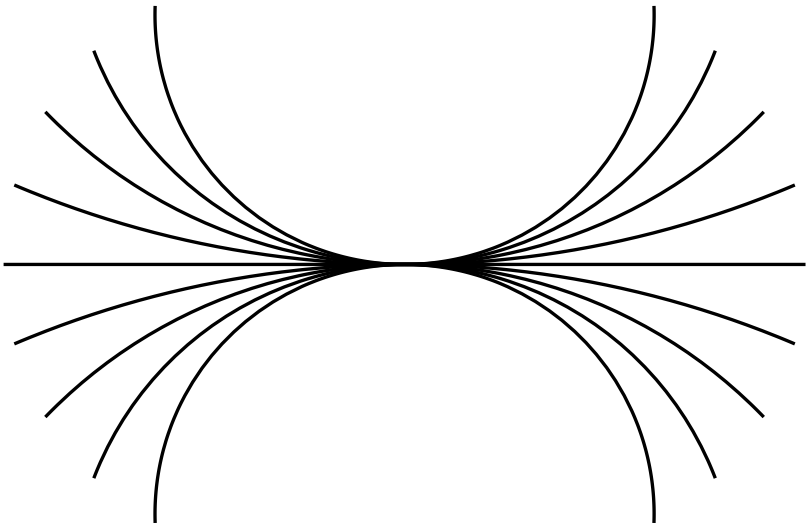


Figure 4.8: The projection of the integral curves with constant coefficients defined by eq. 4.9 modelling the association fields [29].

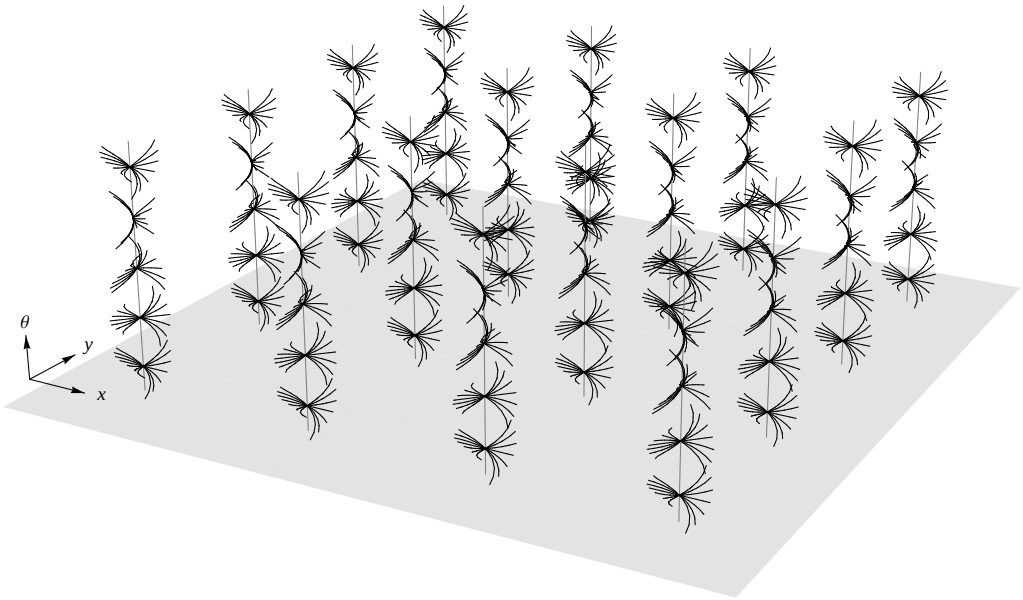


Figure 4.9: In each point of the cortical inspired space, the horizontal connectivity is locally modelled by the fan of integral curves.

4.5 Discussion

The SE(2) model of the cortex presented in this Chapter is 3D and continuous. Therefore, this model does not take into account the fact that the cortex is basically a 2D sheet formed by a discrete number of singular points, named pinwheels, which are surrounded by large areas which are only sensitive to roughly one orientation, as shown in the optical images of the previous chapter. A possible answer to this incongruence is well described in Chapter 4, Section 4.7 of the book of J. Petitot [74]. It essentially shows that, when considering the limit case in which the density of the pinwheels tends to infinity, each point becomes a pinwheel which may be interpreted as a fiber (a point where all orientations are present once). Then, the cortex becomes a full 3D space isomorphic to $\mathbb{R}^2 \times S^1$.

However it can be interesting to recall that the original fibration model introduced first by Hoffman does not take into account the existence of pinwheels. In facts it preceded the discovery of the pinwheels. The fundamental assumption was based on the strong redundancy present in the cortex when it codifies retinal positions and orientations. As already outlined by Hubel and Wiesel, there is a whole orientation hypercolumn of simple cells associated to each retinotopic position. Within the cells of each hypercolumn, their

receptive profiles are centered at the same retinal point even though the cells are arranged horizontally along the cortical sheet. When discretizing the 3D continuous space and collapsing it to 2D, the pinwheels should appear as a consequence of the fact that the orientation variable is periodic. This last point is just an hypothesis to be proved and even if it is a beautiful problem, it is beyond the scope of this thesis.

A novelty introduced by Citti and Sarti is the use of the sub-Riemannian metric and in my opinion is among the strengths of the framework. This kind of metrics often appear when one of the variables considered is a state variable ($\theta = \theta(\dot{x}, \dot{y})$ as stated in eq. 4.4). We have shown that the sub-Riemannian metric may be interpreted as a degenerated Riemannian metric where one of the components is forbidden.

Probably the main strength is the methodology presented in order to construct the cortical space. The same procedure followed in this chapter may be extended in order to consider other features encoded within the cortical architecture. One example of this extension, described in this thesis in Chapter 8, is based on [84] where the scale is taken into account.

In this chapter the model for the lateral connectivity is the fan of integral curves with constant coefficients (the helicoidal curves shown in Fig. 4.7). This pattern tell us how points belonging to different fibers (cells from different hypercolumns) are connected. The reader may observe how more structure was add progressively to the group. First, each point was identified with the parameters of simple cells (Fig. 4.1). The simple cells where then modelled as one-forms in the direction of X_3 which naturally define the horizontal tangent planes (Fig. 4.5). Finally long range connectivity was included (Fig. 4.9). The projection of the fan of integral curves on the retinal plane (Fig. 4.8) was proposed in [14] as a model of Field, Hayes and Hess association fields. We also described the relation between the cortical model and the elastica functional.

Chapter 5

Image completion in SE(2)

5.1 Introduction

Chapter 5 presents an implementation of the perceptual completion model proposed by Citti and Sarti in [14]. This work is reported in [79, 80]. In the Citti and Sarti model, an image is lifted onto a surface in the $SE(2)$ space. The completion was achieved by means of a propagation process modelled as a two step algorithm inspired by neural architectures. In the original work it was proved that the algorithm converges to a diffusion driven mean curvature flow in the sub-Riemannian settings so the mean curvature flow was proposed in order to provide completion. As a result of the curvature flow a minimal surface in the sub-Riemannian metric is achieved.

In this chapter we directly implement the two step algorithm of diffusion-concentration. The diffusion process, modelling cortical propagation, is restricted to the sub-Riemannian differential structure. The implementation of the completion algorithm is the first original contribution of this thesis.

As we will show, the minimal surfaces of the sub-Riemannian structure are ruled in geodesics which coincide with Mumford's elastica. Therefore we start the chapter reviewing existing PDE models for image completion inspired by this classical functional. We then begin with the model of completion, adapting that of Citti and Sarti. In Section 5.3 existing boundaries and level lines are lifted to the 3D space directly using differential instruments. All surfaces, also in absence of holes in the original image, have to be completed due to the possible amodal completion.

In Section 5.4 we describe the implementation of the two step diffusion-concentration algorithm. We study in depth the concentration step, and

introduce a renormalization technique to implement the multiple concentration feature which allows to handle occluded and occluding objects at the same time. In Section 5.5 we present the proposed numerical scheme, and in Section 5.6 we apply the algorithm to relevant cognitive cases and show some results.

5.2 Related work

Many computer vision techniques have been proposed to model perceptual completion, either heuristically based or biologically inspired. Here we review the PDE based models more directly related to the well known Elastica functional, and refer to the next chapter for probabilistic models of completion.

A modification of the Mumford-Shah functional, with the Elastica term, was first presented in [65]. Then, the Elastica functional had been extended to a completion of level lines in a missing region of an image in [55, 2]. However the Elastica term is of second order, leading to 4th order PDE. A strictly related technique had been introduced in [7, 5, 12, 98], where the high order functional is split into two second order differential equations. First the orientation information represented as an unitary vector field is propagated and then the color information is diffused along this vector field.

All these models work directly on the two dimensional image space. In order to make occluded and occluding objects present at the same time in the image, in [65] (and then again in [25]) a third dimension is introduced, and the objects present in the image are represented as a stack of sets, ordered by depth. In [85] the third added dimension is represented by the time, the algorithm first detects occluding objects, then those occluded. However these algorithms do not allow for self occluding objects.

To overcome the occlusion restriction, in [14] a completion model was proposed based on the functional architecture of the visual cortex, where the completion is fully performed in the rototranslation group $SE(2)$, allowing for the simultaneous reconstruction of occluding and occluded objects. The basic idea is the following: a two dimensional image is lifted onto a surface in the 3-dimensional sub-Riemannian space, an occlusion is considered as a hole in the surface, and the proposed model completes the missing part of the image with a minimal surface. Computing a minimal surface in the hole and re-projecting it over the image domain, the authors found the same level lines as those found by Morel and Masnou in [55]. The gray-level is propagated from the boundary to the interior using a Laplace-Beltrami operator on the computed surfaces.

Geometric properties of the Sub-Riemannian minimal surfaces have been proved by S. Pauls and R. Hladky in [38], which provided a quick method for disocclusions. However their approach is no able to handle simultaneous occluded and occluding objects.

5.3 Lifting the existing level lines in a 3D space

Let us start by recalling the Citti and Sarti model [14]. An image I can be represented as a bounded function defined on a domain $M \subset \mathbb{R}^2$, $I : M \to \mathbb{R}^+$. The points of M have coordinates (x, y).

As seen in the previous chapter, the output of the simple cells in response to a visual stimulus I is a function u defined on the 3D cortical space. This function u can be interpreted as the cortical activity. The maximal selection mechanism then detects, at every point (x, y) pertaining to a level line of I, the orientation $\theta(x, y)$ of that level line.

5.3.1 Orientation detection

At every point of the image we detected the tangent direction to the level lines $(I_y, -I_x)$, where I_x and I_y are the components of the image gradient. If θ is the angle between the tangent and the x-axis the tangent can be rewritten as $(\cos(\theta), \sin(\theta))$. Then:

$$\theta(x, y) = -\arctan\left(\frac{I_x}{I_y}\right), \quad \theta \in S^1.$$

And the image is then lifted to the three dimensional surface Σ:

$$\Sigma = \left\{(x, y, \theta) \in \mathbb{R}^2 \times S^1 \,:\, \theta(x, y) = -\arctan\left(I_x/I_y\right)\right\}.$$

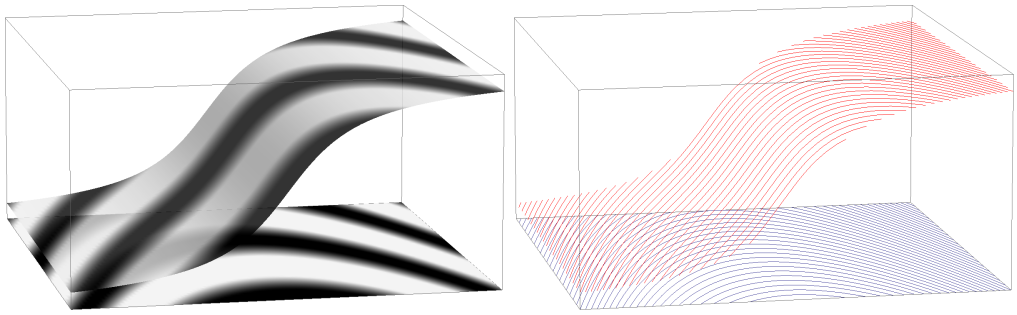


Figure 5.1: An image is lifted into the space of positions and orientations $\mathbb{R}^2 \times S^1$ (left). The resulting surface is foliated by the lifting of the image level lines (right).

This surface Σ is the lifting of every level line in the image. This point of view allows us to understand a remarkable property of Σ: it is a *ruled* surface (see [14, 38]). In fact, since two level lines of an image never cross, also the lifted level lines do not. Then, we say that the lifted surface is foliated by the lifted curves (see Fig. 5.1). We will call *rule* every curve foliating the surface Σ.

5.3.2 The lifted surface as a surface of maxima

Let us now represent the surface Σ as the surface of maxima of a suitable function $u : \mathbb{R}^2 \times S^1 \to [0, 1]$, representing the cortical activity. A possible choice of such a function is:

$$u(x, y, \theta) = f\left(\cos\left(\frac{\theta + \arctan\left(\frac{I_x}{I_y}\right)}{2}\right)^2\right), \tag{5.1}$$

where $f : [0, 1] \to [0, 1]$ is a monotonically increasing function such that $f(1) = 1$. The squared cosine function is chosen in order to have periodicity of u in the third coordinate considering it is an angle. Note we have imposed that the maximum value of u is 1 and that u is always positive. The function f is used to concentrate the function u around its maximum (the surface Σ). A possible choice of f is: $f(x) = x^\alpha$ such that $\alpha \geq 1$ or if we want u to be even more concentrated: $f(x) = e^x$.

From a modelling point of view the function u can be interpreted as an idealized version of the cortical activity, which attains its maximum on the surface Σ.

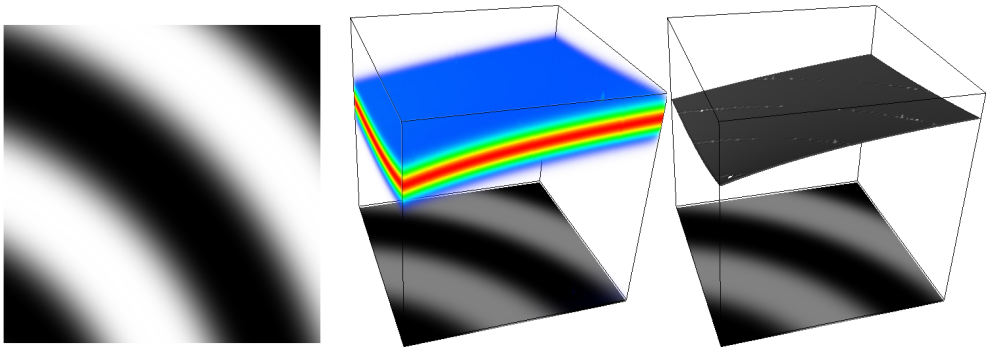


Figure 5.2: The lifted image can be viewed as a thick surface and the surface obtained with eq. 5.2. Shown on the right panel there is a test image. Shown on central panel there is the function u using the volume rendering technique. The red color indicates high values of u while blue indicates those smaller. Visualized on the left panel there is the lifted image Σ which corresponds to the surface of maximum values of u along the fiber over each point (x, y).

5.3.3 Differential geometry of the surface

The surface Σ can be represented as the zero level set of the partial derivative of u with respect to the variable θ:

$$\Sigma = \left\{ (x, y, \theta) \in \mathbb{R}^2 \times S^1 : \frac{\partial}{\partial \theta} u(x, y, \theta) = 0, \frac{\partial^2}{\partial \theta^2} u(x, y, \theta) < 0 \right\}. \tag{5.2}$$

The condition on $\frac{\partial^2}{\partial \theta^2} u$ is imposed in order to avoid the minimum of u being considered as part of the surface. The expression 5.2 allows us to recover Σ from u. Fig. 5.2 shows an example of the construction of the function u and the lifted surface.

Therefore, it is possible to define geometrical properties of Σ in terms of the function u_θ and its sub-Riemannian derivatives. The sub-Riemannian gradient $\nabla_{\text{SR}} u_\theta$ is orthogonal to the surface Σ and belongs to the horizontal tangent plane. Then, the surface has only one tangent vector in the sub-Riemannian structure which writes:

$$T_{\text{SR}} = \begin{pmatrix} -X_2 \\ X_1 \end{pmatrix} u_\theta.$$

This tangent vector is in the direction of the foliation of the surface. Correspondingly, the rules on the surface can be written as:

$$\gamma' = -X_2 u_\theta \vec{X_1} + X_1 u_\theta \vec{X_2}. \tag{5.3}$$

Analogously the diffusion on the surface is the diffusion along the rules. The foliation feature suggests a natural notion of area in the sub-Riemannian structure $\mathbb{R}^2 \times S^1$. Indeed the area of a lifted surface can be defined as the integral of the lengths of each rule. With this definition, a minimal surface with an assigned boundary condition is obtained requiring each rule to have minimal length.

5.4 The concentration-diffusion process

In [14] it is proposed that the cortical activity, modelled by the function u (in this work obtained by applying (5.1) to an input image I) is elaborated through the two cortical mechanisms of non maxima suppression, and diffusion along the long range connections applied in sequence. The mechanism consists of:

- the diffusion of existing information in the sub-Riemannian space with a suitable sub-Laplacian operator;
- the concentration of the diffused information along the third dimension θ.

In [14] a scheme of this algorithm is proposed and its convergence is proved. As a result of this convergence, the authors performed the completion by means of mean driven curvature flow, implemented through a level set method. Here, on the contrary, we directly implemented the diffusion and concentration process, which forced us to take an in depth understanding of the concentration step.

5.4.1 The sub-Riemannian operators

The function u represents the cortical activity, and is diffused along the horizontal connections described in Chapter 4 as integral curves of the vector fields X_1 and X_2. If we assume that these two operators have the same role in the diffusion, we can then describe the diffusion in terms of the operators X_1 and X_2 assigning to them the role of the Euclidean partial derivatives.

We define the sub-Riemannian gradient as:

$$\nabla_{\text{SR}} u = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} u$$

and the sub-Riemannian divergence:

$$\mathrm{div}_{\mathrm{SR}}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \nabla_{\mathrm{SR}} \cdot \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = X_1 u_1 + X_2 u_2$$

The notation SR (sub-Riemannian) is used in order to avoid confusions with the classical operators. Finally, the so called sub-Laplacian operator, which is the analogous of the classical Laplacian but in this structure is defined as:

$$\begin{array}{rcl} \Delta_{\mathrm{SR}} u & = & \mathrm{div}_{\mathrm{SR}}(\nabla_{\mathrm{SR}} u) \\ & = & (X_{11} + X_{22})u \\ & = & \left(\cos^2(\theta)\frac{\partial^2}{\partial x^2} + \sin^2(\theta)\frac{\partial^2}{\partial y^2} + 2\cos(\theta)\sin(\theta)\frac{\partial^2}{\partial x \partial y} + \frac{\partial^2}{\partial \theta^2}\right) u. \end{array} \tag{5.4}$$

Then, the sub-Riemannian diffusion equation can be written as:

$$u_t = \Delta_{\mathrm{SR}} u. \tag{5.5}$$

Despite of the fact the sub-Laplacian operator is built just with two directional derivatives in a three dimensional space, the diffusion process reaches each point due to the non-commutativity of the space.

All the above operators belongs to the horizontal tangent plane, which was identified as the generator of the differential structure of the cortical inspired space.

5.4.2 Sub-Riemannian Diffusion

Let D be the domain of the occlusion and therefore the set $D \times S^1$ contains all the fibers above D. The first step of the diffusion driven algorithm consists of diffusing using the Sub-Laplacian operator (5.4) for a short time with the fixed boundary conditions in the boundary of $D \times S^1$:

$$\begin{cases} \frac{\partial}{\partial t} u = \begin{cases} \Delta_{\mathrm{SR}} u & \text{if } (x, y, \theta) \in D \times S^1 \\ \frac{\partial^2}{\partial \theta^2} u & \text{if } (x, y, \theta) \in (\mathbb{R}^2 \setminus D) \times S^1 \end{cases} & , t \in [0, h] \\ u(0) = u_0 \end{cases} \tag{5.6}$$

In the occluded region we diffuse using the sub-Laplacian operator. This operator propagates data in the direction of the vectors X_1 and X_2. The diffusion in the direction of X_1 alone would expand the information taken from the boundary into the occlusion just in a straight line parallel to the (x, y) plane. By adding the diffusion into the direction of X_2, we allow for the propagation on the curvilinear paths on $\mathbb{R}^2 \times S^1$, even if we make the

surface thicker represented by u as a side effect. Outside $D\times S^1$ we use the equation $\frac{\partial}{\partial t}u = \frac{\partial^2}{\partial \theta^2}u$ in order keep the same thickness of the surface as in the interior of $D\times S^1$. Note that if we just use this equation for a short time the maximum of u does not move and therefore the surface Σ does not change. For the problem of disocclution it is necessary only to consider values of u near the boundary of $D\times S^1$. These values only will be propagated inside $D\times S^1$. Nevertheless, for improvement of visualization we will consider a larger domain outside $D\times S^1$.

The initial function u_0 for the diffusion algorithm is given by the output of the concentration step in each iteration. In the initial diffusion step, where no concentration was done yet, $u_0 = 0$.

5.4.3 Concentration

After the diffusion, we have to take into account the role of the concentration mechanism. Hence we have to concentrate the function u over the surface, i.e. make the thicker version of the surface thinner.

After diffusing u for a period of the time h, the second step of the diffusion driven motion is to perform a concentration over the maximum of u and to denote $\bar{u}$ as the new function which implicitly defines the concentrated surface:

$$\bar{u}(x,y,\theta) = \left(\frac{u(x,y,\theta)}{u_{\max}(x,y)}\right)^{\gamma}, \ \gamma > 1 \tag{5.7}$$

where:

$$u_{\max}(x,y) = \max_{\theta\in S^1}\{u(x,y,\theta)\}. \tag{5.8}$$

The concentration process we present here is completely original. Indeed we introduced a renormalization step. We renormalize the function u in such a way that the maximum over each fiber is 1, and the surfaces does not vanishes asymptotically. The concentration, obtained elevating the function u to a suitable power greater than one, preserves the value of the maximum and reduces all the other values of u. Thus this mechanism concentrates the function around its maximum (see Fig. 5.3).

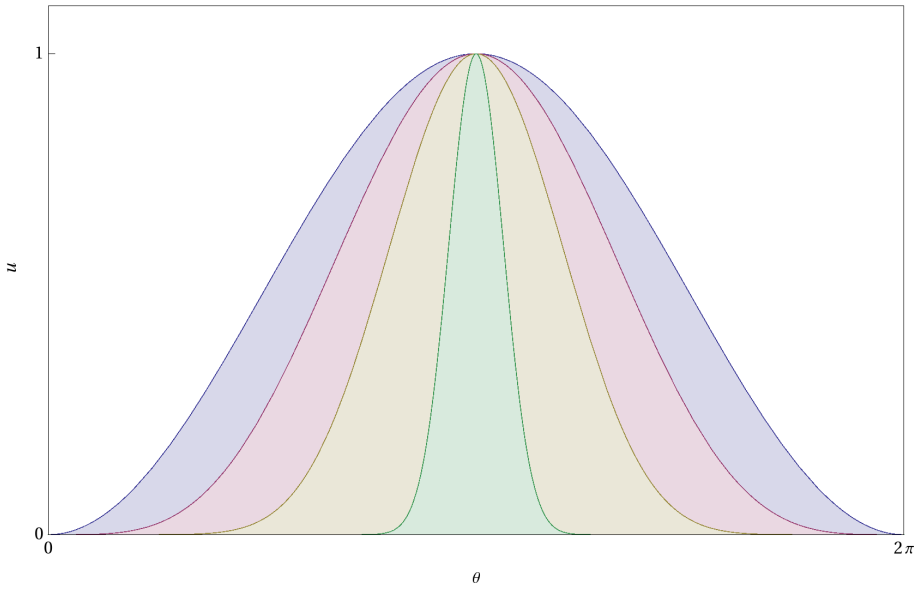


Figure 5.3: Example of the concentration process of a single fiber.

5.4.3.1 Multiple concentration

The three dimensionality of the space allows the coexistence of occluded and occluding objects at the same time. In terms of the function u it means that we expect to have more than one local maximum in each fiber. However, the equation (5.8) described before, allows for only one maximum per fiber. We will modify the concentration in order to avoid this limitation. In particular we propose a new renormalization criteria.

We first detect the local maxima on a fiber over the point (x, y) as the set $\{\theta \in S^1, \frac{\partial}{\partial\theta}u(x, y, \theta) = 0, \frac{\partial^2}{\partial\theta^2}u(x, y, \theta) < 0\}$. We call them $\theta_1, \ldots, \theta_n$ with $\theta_i < \theta_{i+1}$. Then we construct a piecewise linear function u_{norm} (Fig. 5.4) connecting every local maximum detected and which is periodic in the variable θ:

$$u_{\text{norm}}(x, y, \theta) = u(x, y, \theta_j) + (\theta - \theta_j)\frac{u(x, y, \theta_{j+1}) - u(x, y, \theta_j)}{\theta_{j+1} - \theta_j} \tag{5.9}$$

with $\theta \in [\theta_j, \theta_{j+1}]$.

We use eq (5.9) to re-normalize every single column of u as follows:

$$\bar{u}(x, y, \theta) = \left(\frac{u(x, y, \theta)}{u_{\text{norm}}(x, y, \theta)}\right)^{\gamma}, \ \gamma > 1.$$

After renormalization the function $\bar{u}$ preserves the local maxima of u and attains the value 1 at each of these points.

As we mentioned before, this modification allows for more than one maximum on each fiber.

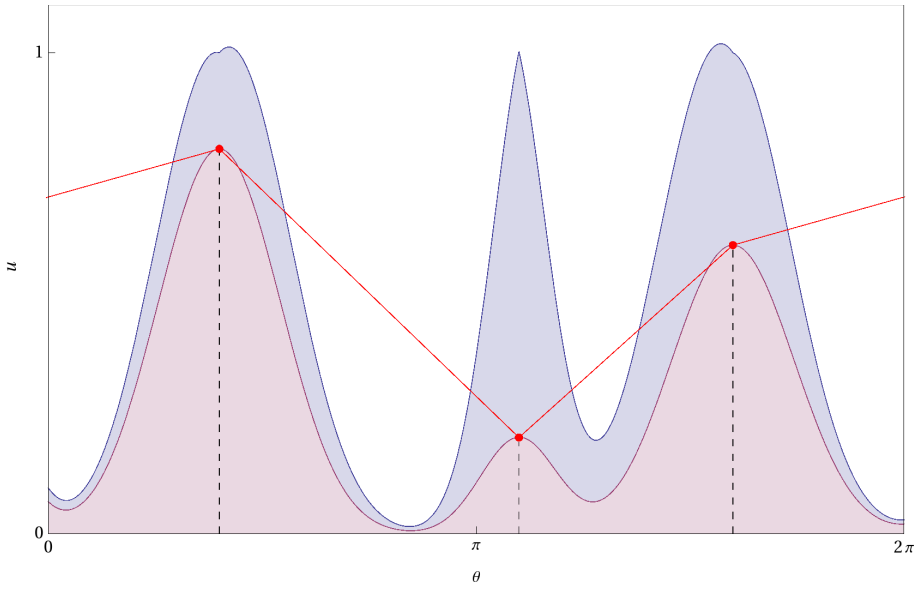


Figure 5.4: Example of the improved re-normalization process of a single fiber.

Hence applying this improved concentration technique and the sub-Riemannian diffusion iteratively, we compute minimal surfaces in $\mathbb{R}^2 \times S^1$ which are a union of graphs of the variable (x, y), which may partially overlap. It corresponds to the completion of both occluding and occluded object.

5.4.4 Laplace-Beltrami diffusion

At this point we have accomplished the missing information of the lifted surface with a minimal surface in the sub-Riemannian space. The lifting and completion processes take into account only the direction of the level lines of the image, as geometric information. Then the intensity information of the image is totally missed.

Let us define a function v by extending the values of the image I on the 3D space:

$$v(x, y, \theta) = \begin{cases} I(x, y) & (x, y, \theta) \in (\mathbb{R}^2 \backslash D) \times S^1 \\ 0 & (x, y, \theta) \in D \times S^1 \end{cases}$$

We will use the Laplace-Beltrami diffusion algorithm in the sub-Riemannian setting introduced in [14], to propagate the function v along the rules of the minimal surface. Since the rules of the surface, defined in (5.3) only depend on $\nabla_{\text{SR}} u_\theta$, the Laplace Beltrami operator is a linear operator in the variable v whose coefficients depend on $\nabla_{\text{SR}} u_\theta$:

$$\frac{\partial}{\partial t} v = \frac{|X_2 u_\theta|^2 X_1^2 v + |X_1 u_\theta|^2 X_2^2 v}{X_1^2 u_\theta + X_2^2 u_\theta} - \frac{X_1 u_\theta X_2 u_\theta X_{12} v - X_1 u_\theta X_2 u_\theta X_{21} v}{X_1^2 u_\theta + X_2^2 u_\theta}. \tag{5.10}$$

The Laplace-Beltrami operator is the analogous of the Laplace operator restricted to surfaces or manifolds. In this case, the Laplace-Beltrami operates on the surface of maxima under the sub-Riemmanian metric. The idea is that the diffusion of the intensity is only performed along the rules of the surface, which if projected should be the level lines of the image in the occluded region. The height of the rules is their orientation. We are then diffusing along the image level lines. This idea is not new and is similar to the inpainting algorithm proposed by Bertalmio et Al. in [5].

The expression of the operator in (5.10) seems complicated, but its deduction in this case is relatively straightforward. The sub-Riemannian Laplace-Beltrami is a second derivative along the direction $X_1 + kX_2$. As the surface is minimal, it is ruled by geodesics with constant curvature $k = \frac{X_1 u_\theta}{X_2 u_\theta}$ (this property was proved in [38]). Then it follows that the Laplace Beltrami is (up to normalization):

$$\left(X_1 - \frac{X_1 u_\theta}{X_2 u_\theta} X_2 \right) \left(X_1 - \frac{X_1 u_\theta}{X_2 u_\theta} X_2 \right)$$

Expanding this last expression and using the fact $\frac{X_1 u_\theta}{X_2 u_\theta}$ is constant along the rules, we get the right hand side of (5.10).

5.5 Numerical scheme

For the diffusion we use a finite difference scheme. Let us consider a rectangular grid in space-time (x, y, θ, t). The grid consists of a set of points $(x_l, y_m, \theta_q, t_n) = (l\Delta x, m\Delta y, q\Delta\theta, n\Delta t)$.

Following the standard notation, we denote the value of the function u at a grid point by u^n_{lmq}. We use forward differences in order to approximate the

time derivative:

$$D_t u = \frac{u^{n+1}_{lmq} - u^n_{lmq}}{\Delta t}$$

and centered differences for the spatial ones:

$$D_x u^n_{lmq} = \frac{u^n_{(l+1)mq} - u^n_{(l-1)mq}}{2\Delta x}$$

$$D_{xx} u^n_{lmq} = \frac{u^n_{(l+1)mq} - 2u^n_{lmq} + u^n_{(l-1)mq}}{(\Delta x)^2}.$$

The second directional derivatives are approximated with:

$$\begin{aligned} D_{11} u^n_{lmq} &= \cos(\theta_q)^2 D_{xx} u^n_{lmq} + \sin(\theta_q)^2 D_{yy} u^n_{lmq} \\ &\quad +2\cos(\theta_q)\sin(\theta_q) D_{xy} u^n_{lmq} \\ D_{22} u^n_{lmq} &= D_{\theta\theta} u^n_{lmq} \end{aligned}$$

We impose the Neumann boundary conditions on x and y and the periodic boundary conditions on the third direction θ. The time step Δt is upper bounded by the usual Courant-Friedrich-Levy condition that ensures the stability of the evolution [11]. The completion algorithm was fully implemented in C++ using the ITK library. For visualization purposes we used the VTK library.

5.6 Experiments and results

5.6.1 Macula cieca example

In this experiment we consider the completion of a figure which has been partially occluded. This example mimics the missing information due to the presence of the macula cieca (blind spot). The blind spot is the place in the visual field which corresponds to the lack of photoreceptors on the retina, due to the optic nerve passing through it. Since there are no cells to detect light, part of the visual field is not perceived. The human visual system modally completes it, so the blind spot is not perceived.

As described in the previous section the occluded image is lifted to a surface with a hole in the three dimensional space and an initial surface is defined in the missing part with a classical Euclidean diffusion equation. Then the surface is evolved by applying iteratively equations (5.6) and (5.8) until a steady state is achieved.

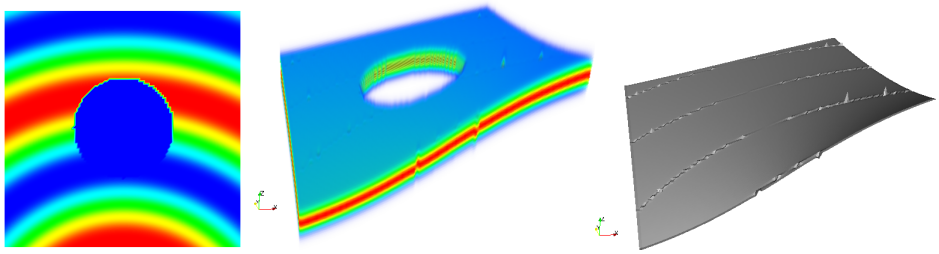


Figure 5.5: Macula cieca example: original image (left), Initially lifted surface (center) and the minimal surface computed (right).

The image dimensions are 100×100 pixels, and we use 100 values to discretize the variable θ. For the preprocessing step 100 iterations of the Euclidean heat equation were made using a time step of $\Delta t = 0.1$. The steady state was reached after 20 iterations with a concentration power in (5.7) of $\gamma = 2$ and 20 steps with $\Delta t = 0.1$ of the sub-Riemannian heat equation (5.6).

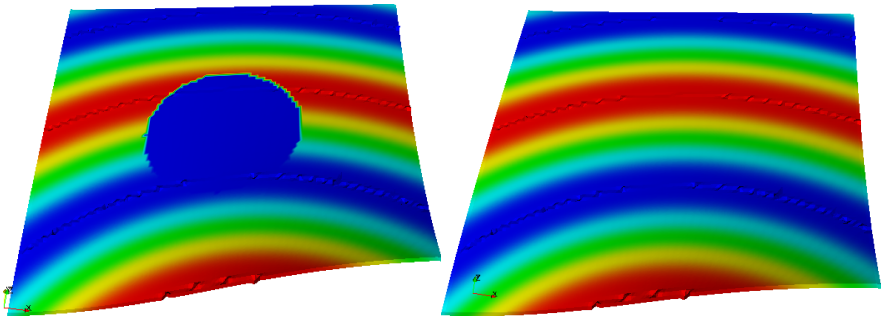


Figure 5.6: Gray level diffusion in the macula cieca example.

5.6.2 Occlusion example

In Figure 5.7 an occlusion problem is considered. The initial image (top) shows an underlying object partially occluded by a vertical stripe. The human visual system simultaneously segments the occluding object and amodally completes the occluded one, capturing both at the same time as perceived units. In the numerical experiment first the image is lifted in the sub-Riemannian space and the missing information is completed. The result shows that the partially occluded object has been completed and the occluding one has been segmented. Both objects are present at the same time in the three dimensional space.

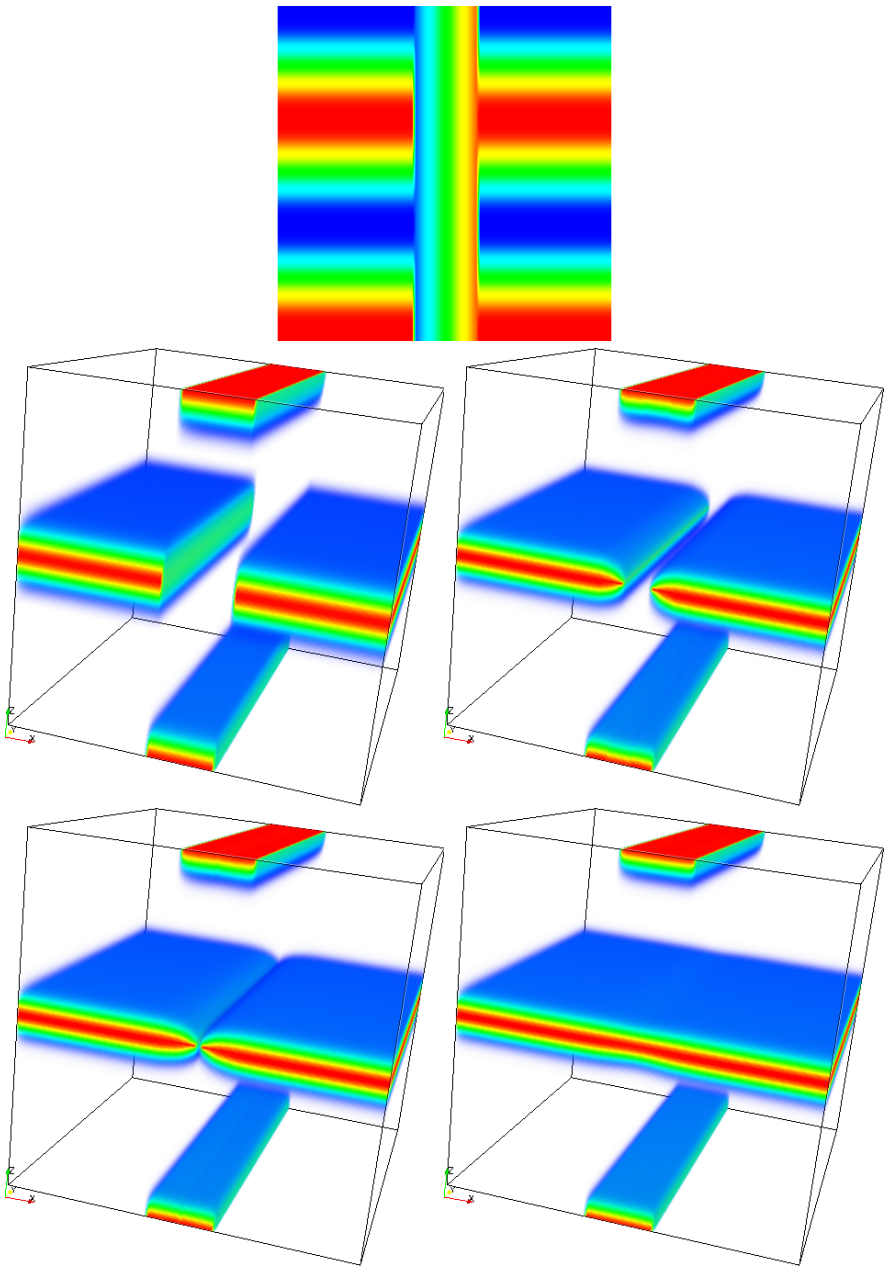


Figure 5.7: Occlusion example: mean curvature evolution with 2 simultaneous surfaces. The upper panel shows the test image, a vertical bar is occluding a background made of an horizontal wave. Shown in the bottom panel is the function u in the initial state (immediately after the lifting), and after successive iterations of the diffusion-concentration algorithm.

For this example the dimensions were again $100 \times 100 \times 100$ pixels. No preprocessing step is needed. The steady state was reached after 10 iterations with a concentration power of $\gamma = 2$ in the equation 5.7 and 10 steps with $\Delta t = 0.1$ of the sub-Riemannian diffusion step.

5.6.3 Other examples

Shown in Figure 5.8 are three examples of completion performed with the diffusion concentration method. In the first case (top row), the missing area is expected to be completed with a surface ruled by straight lines. In the second experiment, the final surface should be ruled with curved lines and the result is the expected. These two examples are similar to those presented by Hladky and Pauls in [38] where the authors perform the completion of the missing part with minimizing independently the length of every level line in the missing region. Finally, on the bottom row, the completion was performed on a real image. The algorithm only work under controlled conditions where there is a gradient of illumination on the objects and there are no textures.

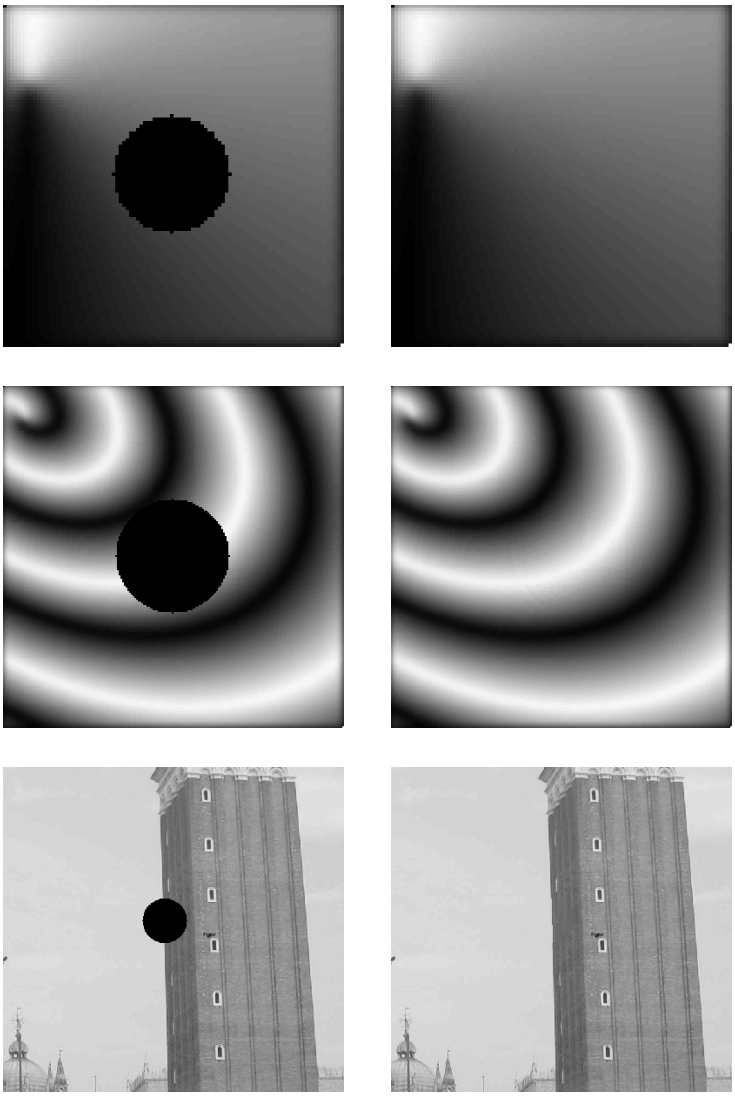


Figure 5.8: Shown in every row are three examples of perceptual completion obtained by means of the diffusion concentration algorithm. In the example in the first row each level line in the occluded region is expected to be straight. In the second row, the level lines in the occluded region are expected to be curved. The final row shows the result of applying the completion method on a real image.

5.7 Discussion

The objective of this work was to test the plausibility of the diffusion concentration scheme in the cortical inspired space. We focused on the propagation of the orientation information and tested if the resulting minimal surface is coherent with the rest of the image. We outlined that the cortically inspired diffusion-concentration algorithm is better than the simple sub-Riemannian heat operator (5.5) to accomplish the completion task. Different sub-Riemannian completion models in literature adopt the simple sub-Riemannian heat operator. In the opposite, in the present study the diffusion was immediately followed by a concentration step allowing to keep the activity information concentrated around the surface of maxima. In [20, 21] the sub-Riemannian heat operator operator was used for filtering elongated textures which is a different problem than the perceptual completion. In the following chapter a new operator will be introduced which could be better suited for modeling the pattern of lateral connectivity.

Let us observe that with respect to the original work of Citti and Sarti many parts of the algorithm have been simplified, for example, the lifting of the image to the $\mathbb{R}^2 \times S^1$ is achieved by computing the orientation of the gradient. However only non-textured, gray scale images can be used. Moreover, completion is performed only in the direction induced by the contours or the level lines of the images. The inside of the objects was not taken into consideration. The purpose was facing the problem of perceptual completion of contours and not the filling in. The perceptual completion is much more complex than only contour completion, and many features need to be taken into consideration, not only the local orientation.

From the numerical point of view the concentration diffusion scheme presents some open problems. The convergence of the diffusion concentration to the mean curvature flow of the surface of maximum was formally proved on [14], but the convergence of the numerical method used is not proved. However, in the examples presented, the resulting surfaces performed correctly the completion.

Chapter 6

Boundary propagation in SE(2)

6.1 Introduction

In this chapter a probabilistic approach to the perceptual completion problem is carefully reviewed. It is also provided the link to the cortical model of V1 presented in the previous chapters. The objective is to show how these models can be re-interpreted with a group theoretical approach. The models presented here are all based on an advection-diffusion equation in the space $\mathbb{R}^2 \times S^1$, i.e. a Fokker-Planck equation, which is used in different ways to address propagation and enhancement of contours. We will describe how this Fokker-Planck operator is related to the SE(2) geometry of the cortical space presented in Chapter 4. Then we introduce a new time-independent fundamental solution of the Fokker-Planck operator which encodes a more reliable connectivity pattern in the group in comparison to the integral curves and the sub-Riemannian heat kernel of the previous chapter.

The models presented in this chapter were mainly developed over the last 20 years, but all constituted a formalization of the concept of good continuation of boundaries elaborated by the Gestalt Theory. Many models rely on the seminal work of David Mumford [62], which considered a contour as a path described by a particle moving according to certain deterministic laws but suffering a random drift. This allowed us to set the contour problem on a stochastic differential equation framework. Finally, this kernel is compared with the Tensor Voting [58], a well known heuristic technique for perceptual completion proposed by G. Medioni.

More precisely, the chapter is structured as follows:

- Section 6.2.1 contains theoretical basis of stochastic calculus. The

equations presented are illustrated with a very simple example, an isotropic Brownian motion in the plane.

- Section 6.3 presents the Mumford direction process and the associated Fokker-Planck equation.
- Section 6.4 is devoted to reviewing previous works based on the Fokker-Planck equation for image processing and boundary completion.
- Section 6.5 presents the Fokker-Planck equation as the natural advection-diffusion equation in the SE(2) group introduced in Chapter 4.
- Section 6.6 presents two different alternatives for calculating the fundamental solution, using an analytical approximation of the group differential space and numerically simulating the equation with a finite element method technique.
- Section 6.7 compares our model with the tensor voting method.

6.2 Langevin and Fokker-Planck equations

6.2.1 Theoretical background

At the beginning of the nineteenth century, the Scottish botanist Robert Brown observed that pollen grains suspended in liquid performed an irregular motion. The motion was later explained by the random collisions with the molecules making up the liquid. To describe the motion mathematically it is natural to use the concept of a stochastic process $W(t)$, interpreted as the position at time t of a pollen grain. This process is called Brownian motion or equivalently the Wiener process.

Similar stochastic behaviors were observed in a great number of applications, and in our case we will use them to model contours in images. For simplicity alone, let us come back to the original example and interpret a contour as the trajectory described by a particle moving in a fluid according to certain deterministic law (for example under the effects of a damping force), but suffering unpredictable random fluctuations due to collisions with the microscopic molecules of the fluid. The equation modelling this situation is known as the Langevin equation. Because of these fluctuations we do not know the exact position of the particle, but instead we can compute the probability to find the particle in a particular region. Such information is given by the Fokker-Planck equation, a deterministic partial differential

equation whose variable is the probability density function of the particle location.

In the sequel we will give the general form of the Langevin equation and its associated Fokker-Planck equation respectively. In [67, 78] readers can find a detailed explanation and deduction of both equations. A good introductory but rigorous text on stochastic differential equations is Lawrence C. Evanss lecture notes [26].

6.2.1.1 Langevin equation

The stochastic process introduced before is well described by the Langevin equation. The m-dimensional Langeving equation has the general form:

$$\begin{cases} d\xi(t) &= H(\xi(t),t)dt + B(\xi(t),t)dW(t) \\ \xi(0) &= x_0, \end{cases} \tag{6.1}$$

where:

$$\xi(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{pmatrix} \tag{6.2}$$

are the stochastic variables (the position of the point in the moving particle example),

$$H(\xi(t),t) = \begin{pmatrix} h_1(\xi(t),t) \\ \vdots \\ h_n(\xi(t),t) \end{pmatrix} \tag{6.3}$$

is the deterministic drift and

$$B(\xi(t),t)dW(t) = \begin{pmatrix} b_{11}(\xi(t),t) & \dots & b_{1m}(\xi(t),t) \\ \vdots & & \vdots \\ b_{n1}(\xi(t),t) & \dots & b_{nm}(\xi(t),t) \end{pmatrix} dW(t) \tag{6.4}$$

is the stochastic or noise term. $W(t)$ is a n-dimensional Brownian motion.

Observe that the stochastic differential equation (6.1) is a very general equation for a Brownian motion and can be interpreted as the motion equation of a particle under a deterministic law given by the vector H altered by a certain random noise given by $BdW(t)$.

6.2.1.2 Fokker-Planck equation

The underlying probability distribution of the stochastic process defined by the Langevin equation is given by the Fokker-Planck equation, also known as Kolmogorov forward equation. Let p be the transition probability from the state $\xi(t_0)$ at the initial time t_0. The Fokker-Planck equation associated with the general Langevin equation (6.1) is written as:

$$\frac{\partial}{\partial t}p = L_{fp}p \tag{6.5}$$

where

$$L_{fp} = -\sum_{i=1}^{n}\frac{\partial}{\partial x_i}A_i + \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2}{\partial x_i \partial x_j}g^{ij} \tag{6.6}$$

is the general form of the Fokker-Planck operator. The coefficients A_i are equally called drift, advection or transport coefficients as they correspond to a first derivative in the evolution equation and are defined as:

$$A_i = h_i + \sum_{k=1}^{n}\sum_{j=1}^{m} b_{kj}\frac{\partial}{\partial x_k}b_{ij}. \tag{6.7}$$

On the other hand, the g^{ij} coefficients correspond to second order terms and therefore are called diffusion coefficients, defined as

$$g^{ij} = \sum_{k=1}^{m} b_{ik}b_{jk} = (BB^T)_{ij} \tag{6.8}$$

In the case of linear constant coefficients, the operator in (6.6) can be written in matricial notation as follows:

$$L_{FP} = -\nabla^T \cdot A + \nabla^T \cdot G\nabla \tag{6.9}$$

where

$$\nabla = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{pmatrix} \quad \text{and} \quad G = \begin{pmatrix} g^{11} & \dots & g^{1m} \\ \vdots & & \vdots \\ g^{n1} & \dots & g^{nm} \end{pmatrix}.$$

We explicitly note that, if the diffusion term G is positive definite, then it expresses the inverse of a metric of the space. In case in which B is rectangular, or has no maximum rank, its columns describe a basis of a subspace of the tangent space, where G is positive definite; namely the horizontal tangent space.

6.2.2 Example: the isotropic Brownian motion

The simplest Brownian motion in the plane is a random isotropic process such that:

$$\begin{cases} dx = \sigma dW_1(t) \\ dy = \sigma dW_2(t) \end{cases} \tag{6.10}$$

where W_1 and W_2 are independent standard Brownian motions and $\sigma > 0$ is the "standard deviation in the derivatives".

In the notation used in eq. (6.1) the isotropic process is described by:

$$\xi(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix},\ H = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ and } B = \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix}.$$

We can discretize the process to simulate certain random paths. The results are shown in Fig 6.1. These images are valuable in order to visualize the probability distribution underlying the stochastic process as they can be understood as Monte Carlo experiments. Shown on the top row are the arrival points of 100 realizations of the process, always using the same initial starting point at the center. Each figure correspond to a different value of t which increases its value from left to right. In the figure in the bottom-right corner another Monte Carlo experiment is performed and 3000 paths are plotted together. The gray value of each pixel is proportional to the density of paths under that location. This is an estimate of a normalized integral of the probability p for all positive times.

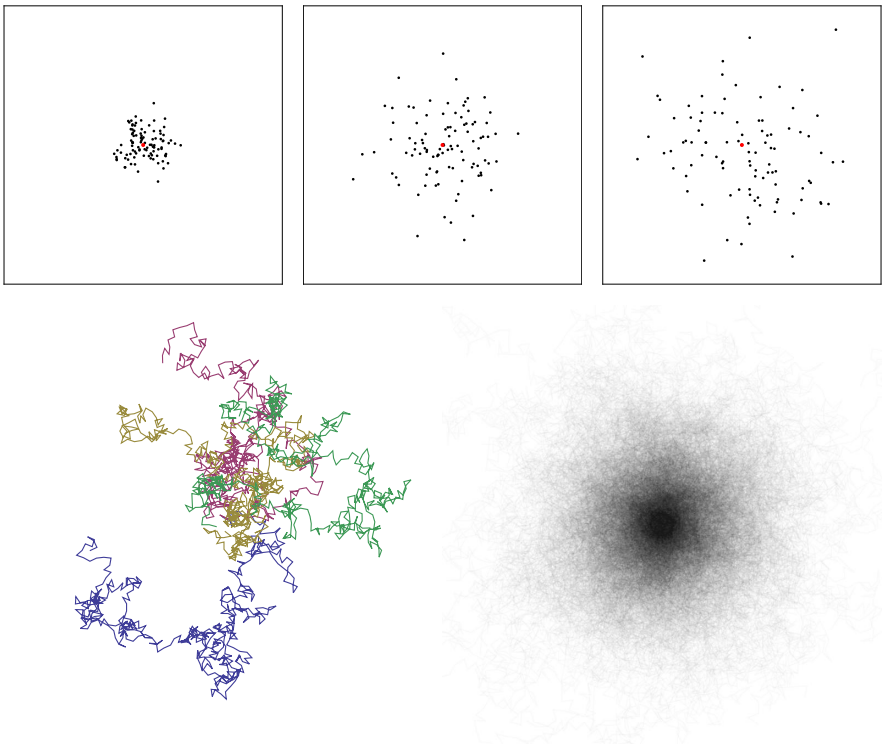


Figure 6.1: Sample path of the isotropic process. Indicated on the top row are the arrival points of 100 simulated paths for three different values of t. From left to right the value of t increases. Shown on the bottom right panel is the result of plotting 3000 paths together on a lattice. The darkness of each pixel is proportional to the density of paths at the point. This picture can also be understood as an approximation of an accumulation for all times of pictures like that on the top row.

As explained before, one can compute explicitly an evolution equation for the probability of transition $p(x, y, t)$. This is the Fokker-Planck equation associated with the stochastic process (6.10). It can be computed according to eq. (6.5) and in being a simple diffusion equation:

$$\frac{\partial p}{\partial t} = \frac{\sigma^2}{2} \Delta p \tag{6.11}$$

where Δ is the 2D Laplacian operator:

$$\Delta = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right)$$

and the initial condition is $p(x, y, 0^+) = \delta(x)\delta(y)$.

For $\sigma^2/2 = 1$, the partial differential equation (6.11) is the well known heat equation whose fundamental solution $\Gamma(x, y, t)$, i.e. the solution of:

$$\begin{cases} \left(\frac{\partial}{\partial t} - \Delta\right) p(x, y, t) = 0, \\ p(x, y, 0^+) = \delta(x)\delta(y) \end{cases}$$

writes:

$$\Gamma(x, y, t) = \frac{e^{-\frac{x^2+y^2}{4t}}}{\sqrt{4\pi t}}.$$

An isotropic Brownian motion in the plane is not a good model for a contour. For example, it is evident that it does not encode the good continuity cue.

6.3 David Mumford's approach

The Field medal winner D. Mumford [62] developed in the early nineties a Bayesian approach to the computer vision problem of reconstructing occluded edges. He proposed modelling edges in natural images as continuous and almost everywhere differentiable planar curves so that, when occluded in part, they would tend to reappear with approximately the same tangent line. The simplest way to model this behavior, is to allow the curvature (function of the arc length) to be white noise, so the direction function becomes a one dimensional Brownian motion. This leads to define contours by means of the following stochastic process:

$$\begin{cases} dx = \cos\theta dt \\ dy = \sin\theta dt \\ d\theta = \sigma dW(t) \end{cases}. \tag{6.12}$$

These equations can be interpreted as the motion of a particle moving with constant speed in a randomly changing direction. The effect is that particles tend to travel in straight lines, but over time, they drift to the left or right by an amount proportional to σ^2. When $\sigma^2 = 0$, the motion is completely deterministic and particles never deviate from straight paths. When $\sigma^2 \to \infty$, the motion is completely random in the plane.

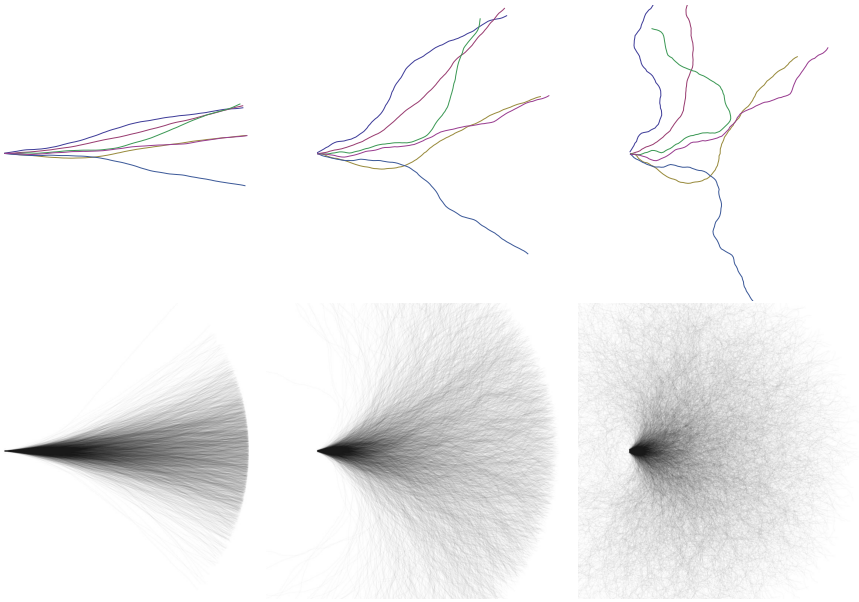


Figure 6.2: The top row represents six random paths resulting from simulating the process (6.12) for three different values of σ. From left to right the value of σ increase. Visualized in the bottom row are 3000 random paths accumulated for the same values of σ in the upper row.

Fig. 6.2 shows some simulated random paths generated according to (6.12), for different values of σ. It becomes empirically evident that these paths are much better suited as contour continuation models than the two-dimensional isotropic random path of the previous section. The position of points for different values of t are shown in Fig. 6.3. In this instance it is necessary to indicate not only the starting point but also the initial direction (indicated by the red arrow). Let us observe that the front is strongly polarized, by contrast to that of Fig. 6.1.

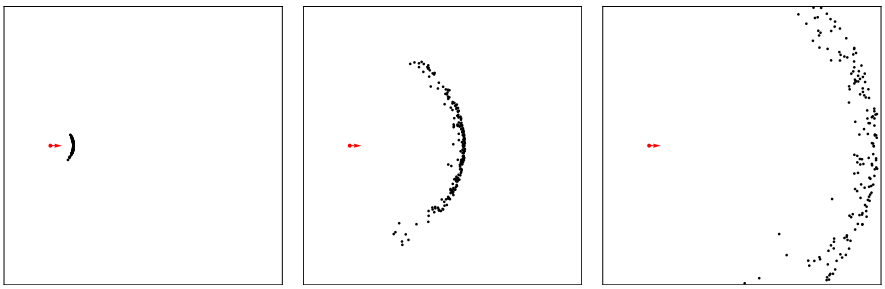


Figure 6.3: The position points for different times t of 100 simulated particles of the stochastic process (6.12).

Note that in this case, we need three variables to model the random process:

$$\xi(t) = \begin{pmatrix} x(t) \\ y(t) \\ \theta(t) \end{pmatrix}.$$

The first two variables $(x, y) \in \mathbb{R}^2$ indicate the position while the third $\theta \in S^1$ is the orientation. Following the same methodology of the previous section we get:

$$H = \begin{pmatrix} \cos\theta \\ \sin\theta \\ 0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{6.13}$$

The Fokker-Planck equation associated to the Mumford stochastic process (6.12) can be computed using (6.6):

$$\partial_t p = -\cos(\theta)\frac{\partial}{\partial x}p - \sin(\theta)\frac{\partial}{\partial y}p + \frac{\sigma^2}{2}\frac{\partial^2}{\partial \theta^2}p. \tag{6.14}$$

Within this formulation, the Fokker-Planck operator consists of an advection term in the direction of the vector $(\cos(\theta), \sin(\theta))$, the direction tangent to the path, and a diffusion term on the orientation variable θ. The Fokker-Planck equation is a deterministic partial differential equation modelling the evolution in the time of $p(x, y, \theta, t)$.

In contrast to the heat equation, there is not a simple explicit formula for the fundamental solution of eq. 6.14. Recently, in [93] and in [22] certain explicit expressions in the Fourier space have been given in terms of Mathieu functions.

6.4 Related work in Computer Vision

A probabilistic framework for propagation of contours and their representation on the space of positions-orientations has been largely used in computer vision application. Within this section we will describe several of them which are closely related to the work presented in this thesis as they are explicitly based on Mumford's ideas. We will describe them in the following: first the approach of the stochastic completion fields (Section 6.4.1), second we will introduce the curve indicator random fields model (Section 6.4.2) and finally the invertible orientation scores technique (Section 6.4.3).

6.4.1 Stochastic completion fields

The approach of the stochastic completion fields, introduced by Williams, Jacobs and co-workers in a number of articles [95, 89, 96, 99] is a method for computing the likelihood that a completion joining two contour fragments passes through any given position and orientation within the image's plane. This likelihood is a measurement of the saliency of a possible contour. The method is based on the same assumption made by Mumford, i.e. that contour continuation can be modelled by the process (6.12).

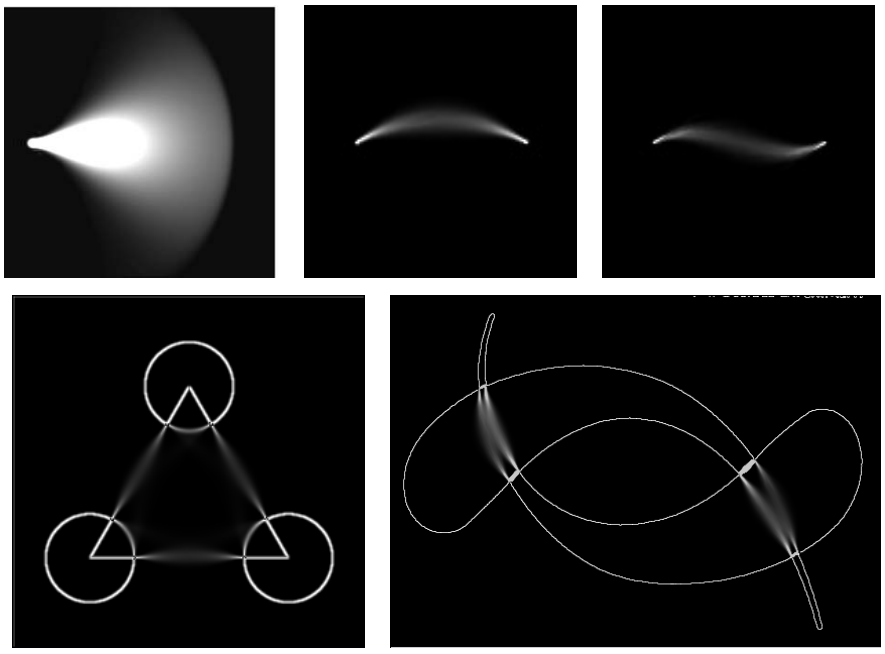


Figure 6.4: The stochastic completion fields. Top row left: a single kernel which is an approximated Fokker-Planck fundamental solution (suitably integrated in time). Top row center and right: two examples of the stochastic completion field between different configurations of the source and the sink. Bottom row: the method applied to two Kanizsa examples of illusory contours, the triangle (left) and the fish (right) (images are reproduced from [96]).

Let us consider two points (which are states of the stochastic process) $P = (x_p, y_p, \theta_p)$ and $Q = (x_q, y_q, \theta_q)$ representing the endpoints of an occluded edge. The stochastic completion field is the multiplication of two fields, the *source* field and the *sink* field, associated to the state P and the state Q respectively. The source field represents the probability that the particle starting at P will reach the state ξ, the sink field represents the probability that a particle

beginning at Q will reach ξ and their product (point to point multiplication) represent their joint probability. Both the source and the sink fields were identified with the fundamental solution of the Fokker-Planck equation (6.14) suitably integrated in time. In order to perform the completion, the initial conditions are set manually, i.e. the position and orientation of the sink and the source are set ad hoc. Therefore, the method was used mainly as a model of illusory contours in synthetic images. The completion fields were computed using different approximations of the Fokker-Planck fundamental solution. In [95] the authors used a Monte Carlo simulation on a lattice (similar to our experiment in Fig. 6.2). In [89] the authors deduced an analytical expression for an approximated fundamental solution (see Section 6.6.1) while in [96] they constructed this approximated solution by the numerical integration of the corresponding Fokker-Planck equation. In [99] a translation-rotation invariant finite element method (FEM) was developed to approximate the fundamental solution.

6.4.2 Curve indicator random field

The curve indicator random field (CIRF) is a framework for curve enhancement, completion and segmentation proposed by J. August and S. Zucker in [3, 4] which is partially inspired by the stochastic completion fields approach. Again the main idea is to use the positions orientations space and Mumford's Fokker-Planck equation for propagating the direction information. Nevertheless, there exists important differences between both works. In the application presented in [4], the CIRF is not used for illusory boundary completion but as an algorithm for improving the saliency of the contours present in the image, as a preprocessing step for edge segmentation. Instead of manually setting the source and sink points, they build a three dimensional function $m \in \mathbb{L}_2(\mathbb{R}^2 \times S^1)$ measuring the local distributions of orientation for each point of the image (x, y) [44]. This function m is convolved with an approximation of the fundamental solution for the Fokker-Planck equation for the direction process, suitably integrated in time. Note that in contrast to the stochastic completion fields, there is no product of distributions of sink and source, but instead an undifferentiated summation of fundamental solutions.

Another improvement of the CIRF is the optional inclusion of the curvature $k = \dot{\theta}$ as a second state variable so that the space on which the image is represented is the $\mathbb{R}^2 \times S^1 \times \mathbb{R}$. This modification allows better filtering performance when the curvature of the structures present in the image is high. The construction of the underlying stochastic process is presented in analogy

to Mumford's presentation of the stochastic process. A justification based on biological evidence for the consideration of the curvature is given in [6].

Figure 6.5: The curve indicator random field filtering a noisy Euler spiral. The original image is on the left; the result after filtering using the curve indicator random field based on Mumfords direction process (center) and the curvature-based process (right). Images are reproduced from [3].

6.4.3 Invertible orientation scores

The techniques described in the sequel were introduced in a number of scientific publications by Haar Romeny, Duits and co-workers [19, 22, 20, 21]. Their work is focused on the design of an *invertible orientation score* which is a function representing an image in the space $SE(2)$. The idea is the construction of a transformation $\mathcal{W} : \mathbb{R}^2 \to SE(2)$ which allows for the lifting of the image onto an orientation distribution or orientation score. The function $\mathcal{W}$ should be invertible so that the image can be reconstructed from the orientation score.

The construction of an orientation score is associated with an anisotropic kernel φ_θ which measures the local orientation at each location. Each filter φ_θ is obtained by rotating a mother kernel φ by an angle θ and each θ-plane of the orientation score is constructed by filtering the image with the corresponding kernel. The kernel φ should fulfill certain requirements so that the invertible transformation can be numerically implemented:

- only a finite number of samples of the orientation space are necessary to construct an invertible orientation.
- the reconstruction is achieved by integrating the orientation score over all orientations.
- the kernels should be compactly supported.

The transformation function constructed with the above restrictions is rotation-translation invariant, the same as the reconstruction.

Similar representations of the image in the space of positions-orientations have been largely used in computer vision but the main novelty of Duits and his colleagues work is the emphasis on the invertible transformation.

Once the orientation score is constructed, the authors distinguish between the use of the left-invariant processes of (6.18) and (6.17). The pure sub-Riemannian diffusion (6.18) process is used for contour enhancement (signifying in this context the de-noising of elongated structures) while the Mumford direction process (6.17) based on the advection diffusion (suitably integrated in time) is used for contour completion of missing boundaries.

6.5 Left Invariant Fokker-Planck equation in $SE(2)$

In the previous section we represented contours as planar curves which are realizations of the direction process proposed by Mumford. The contours were identified with the paths described by moving particles. The stochastic process is defined by setting the direction function so that it is a Brownian motion. The curves can be therefore represented in the three dimensional space of position and orientations:

$$\gamma(x(t), y(t), \theta(t)) \in \mathbb{R}^2 \times S^1.$$

In Chapter 4, we have equipped the space $\mathbb{R}^2 \times S^1$ with the structure of the rototranslation group $SE(2)$. Planar curves were *lifted* into the group $SE(2)$ by means of the map:

$$(x, y) \to (x, y, \theta)$$

were θ is the direction function. A subspace of the Lie algebra was considered in order to give the space a differential structure. This is called the horizontal tangent plane and is the space spanned by the two left invariant vector fields $X_1 = (\cos\theta, \sin\theta, 0)$ and $X_2 = (0, 0, 1)$. We will see that the Fokker-Planck operator of the Mumford direction process is a natural operator in the differential structure of the group, and introduce a new time independent Fokker-Planck equation.

Let us first write the Langevin equation (6.12) on the basis of the Lie Algebra, and the vector fields X_1, X_2. This can be done with a special choice of vector H:

$$H = \begin{pmatrix} \cos\theta & 0 \\ \sin\theta & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}.$$

Analogously, the matrix B has to be left invariant, so that it can be defined in terms of the coefficients of X_1 and X_2 as

$$B = \begin{pmatrix} \cos(\theta) & 0 \\ \sin(\theta) & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \tilde{b}_{11} & \tilde{b}_{12} \\ \tilde{b}_{21} & \tilde{b}_{22} \end{pmatrix}$$

With this choice of matrices the Langevin equation becomes

$$\gamma' = h_1 X_1 + h_2 X_2 + (\tilde{b}_{11}\Gamma_1 + \tilde{b}_{12}\Gamma_2)X_1 + (\tilde{b}_{21}\Gamma_1 + \tilde{b}_{22}\Gamma_2)X_2.$$

Then, the corresponding Fokker-Planck operator becomes:

$$FP_{SR} = -H \cdot \nabla + \nabla^T B B^T \nabla, \tag{6.15}$$

and the Fokker-Planck equation is:

$$\partial_t = FP_{SR}.$$

Remembering that the vector

$$\nabla_{\text{SR}} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \nabla$$

is the sub-Riemannian gradient, we see that

$$H\nabla = (h_1, h_2)\nabla_{\text{SR}} = A\nabla_{\text{SR}}$$

if $A = (h_1, h_2)$ and $B^T\nabla = \tilde{B}^T\nabla_{\text{SR}}$, then the Fokker-Planck operator rewrites as:

$$FP_{\text{SR}} = -A \cdot \nabla_{\text{SR}} + \nabla_{\text{SR}} \cdot \tilde{B}\tilde{B}^T\nabla_{\text{SR}}. \tag{6.16}$$

This general formulation enables us to write as a special case the Fokker-Planck equation for the Mumford stochastic process, as the two matrices in (6.16) are left invariant and are of the previous type. It amounts to choose in the Langevin equation a deterministic coefficient in the direction of X_1 (in other words $h_1 = 1$, $h_2 = 0$), and white noise in direction X_2 ($B = diag(0, 1)$)[1]:

$$\partial_t = -X_1 + \frac{\sigma^2}{2} X_{22}. \tag{6.17}$$

On the contrary, if in the Langevin equation we choose $H = 0$, $B = diag(1, 1)$, assigning the same weight to the direction X_1 and X_2, we obtain the sub-Riemannian diffusion equation used on Chapter 5

$$\partial_t = \Delta_{\text{SR}} = X_{11} + X_{22}. \tag{6.18}$$

[1]The notation X_{ij} means $X_i(X_j)$.

6.5.1 Propagation in the structure of the cortex

In view of the fact that we have interpreted the Fokker Plank equation (6.16) as a propagation and a diffusion in the $SE(2)$ group, which models the structure of the cortex, we will interpret this equation as the natural propagation of the signal in the geometrical structure of the cortex. By now we have proposed two possible propagators: the diffusion equation (6.18), and Fokker-Planck equation (6.16). As we said, the first one assigns the same role to the directions X_1 and X_2, which is coherent with the sub-Riemannian Lie group structure only. This operator was used in the previous chapter as the diffusion operator in the diffusion-concentration algorithm for perceptual completion; the diffusion along was not enough to propagate information. Indeed, we have endowed our cortical structure with a co-tangent structure. This amounts to assign to the direction θ, which is associated to a cotangent variable, a role completely different (from a differential point of view) with respect to the role of the variables x, y. In some sense the variable θ, which expresses the orientation, takes the place of a direction with an homogeneity different from the homogeneity of directions x, y. Hence the natural propagation in the contact structure, which respects the difference between the directions X_1 and X_2 assigns a stochastic meaning in the direction X_2 and a deterministic meaning in the direction X_1. The result is therefore expressed by a Fokker-Planck equation (6.17).

6.5.2 A forward-backward time independent Fokker-Planck equation

Most of the models reviewed in Section 6.3 consider a kernel for propagating information which encodes the probability of having an oriented contour at a certain position conditioned by the fact that there was an oriented contour at a particular reference point. In the previous models this time independent probability was obtained integrating over time the fundamental solution of the Fokker-Planck equation (6.14). Here on the contrary we propose directly using the fundamental solution of the time independent Fokker-Planck operator:

$$-X_1 p(x,y,\theta) + \frac{\sigma^2}{2} X_{22} p(x,y,\theta) = \frac{1}{2}\delta(x,y,\theta). \tag{6.19}$$

On the other hand each point of a contour can be described by a particle that can go either backwards or forwards with identical probabilities in the direction of the contour. The model for the probability density propagation is then the sum of two Fokker-Planck Green functions, one which corresponds to

a particle moving forwards, and the other corresponding to a particle moving backwards:

$$X_1 p(x,y,\theta) + \frac{\sigma^2}{2} X_{22} p(x,y,\theta) = \frac{1}{2}\delta(x,y,\theta). \tag{6.20}$$

In the two equations above, we have written the Fokker-Planck operator using the notation of the previous section calling the differential operators $X_1 = \cos(\theta)\partial_x + \sin(\theta)\partial_y$ and $X_2 = \partial_\theta$.

Section 6.6 is devoted to the computation of the time independent Fokker-Planck fundamental solution. First, let us describe this kernel in order to get some intuition. Visualized in Fig. 6.6 is the comparison between the forward-backward Fokker-Planck kernel and that corresponding to the sub-Riemannian diffusion. The images show arbitrary iso-surfaces of the kernels (colored in red that of the Fokker-Planck and in green that of the diffusion). The kernels are intended as connectivity patterns in the structure of the cortex. They should be a stochastic version of the integral curves of the vector fields $X_1 + kX_2$, which were introduced in Chapter 4 as a deterministic model of the cortical connectivity. This is why the kernels are plotted together with the fan of integral curves (the black curves). From the picture one can clearly observe their relationship with the Fokker-Planck fundamental solution. The kernel seems a thicker version of the fan. The value of the probability decays slowly along the integral curves and quickly in the direction normal to the surface ruled by the integral curves. There is also an increment of the decays proportional to the curvature of each integral curves. On the other hand, the heat kernel has different properties. There is a strong diffusion along the fiber at the origin. This may explain why if we want to propagate information with this operator (as proposed in Chapter 5) a step of concentration is instantly needed. Otherwise the direction information is lost after some diffusion time, and there is no longer selectivity of the direction along fibers. This is not observed in the cortex: simple cells strongly respond to one orientation when there is a contour (real or subjective).

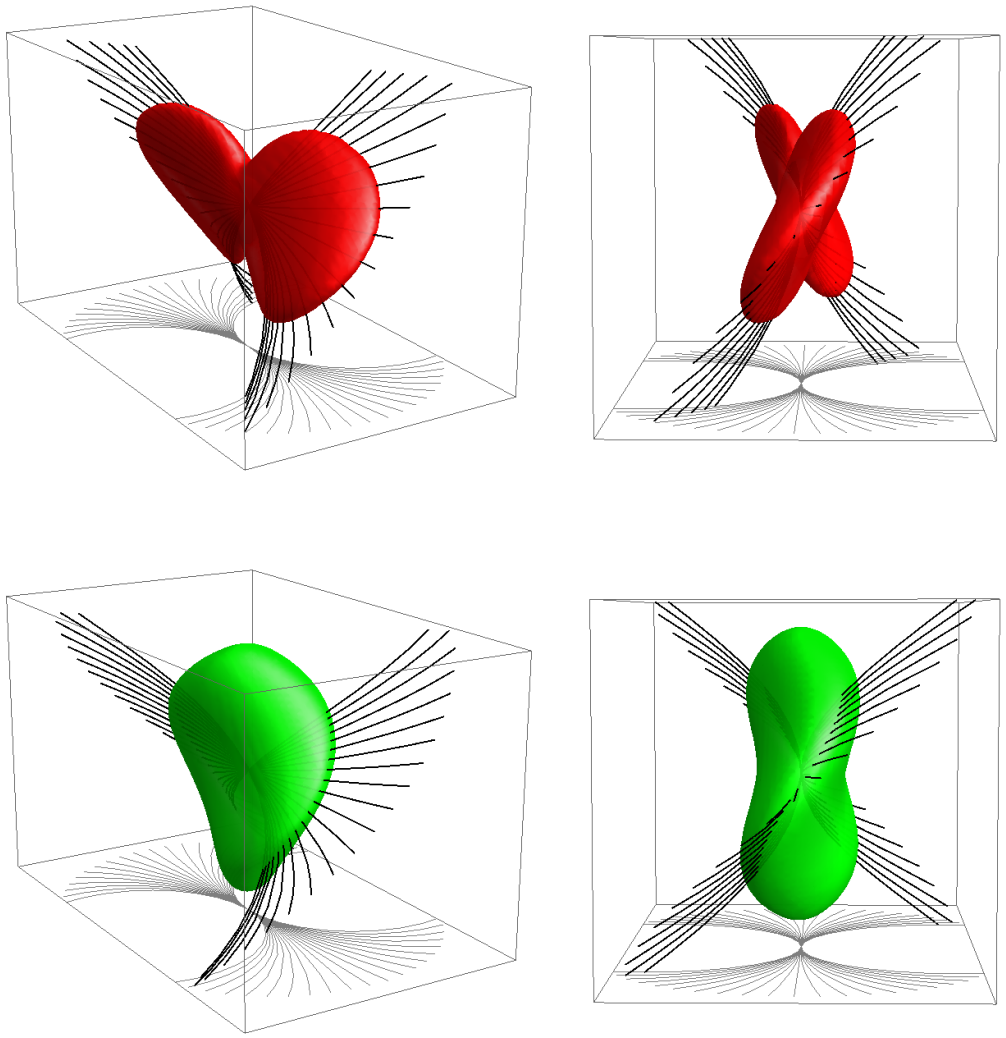


Figure 6.6: Comparison between the sub-Riemanninan Fokker-Planck (red) and heat (green) kernels. The surface visualized correspond to arbitrary iso-surfaces of each kernel. Shown in black are the integral curves of the vector fields $X_1 + kX_2$ introduced in Chapter 4 as a model for the association fields [29]. The kernels are intended in this chapter as models of the cortical propagation of the orientation information in V1.

Figure 6.7 illustrates the additional structure we are proposing as a model of the visual cortex functional architecture. Each fan of integral curves which in Fig. 4.9 represented the space connectivity, was replaced by the Fokker-Planck

kernel.

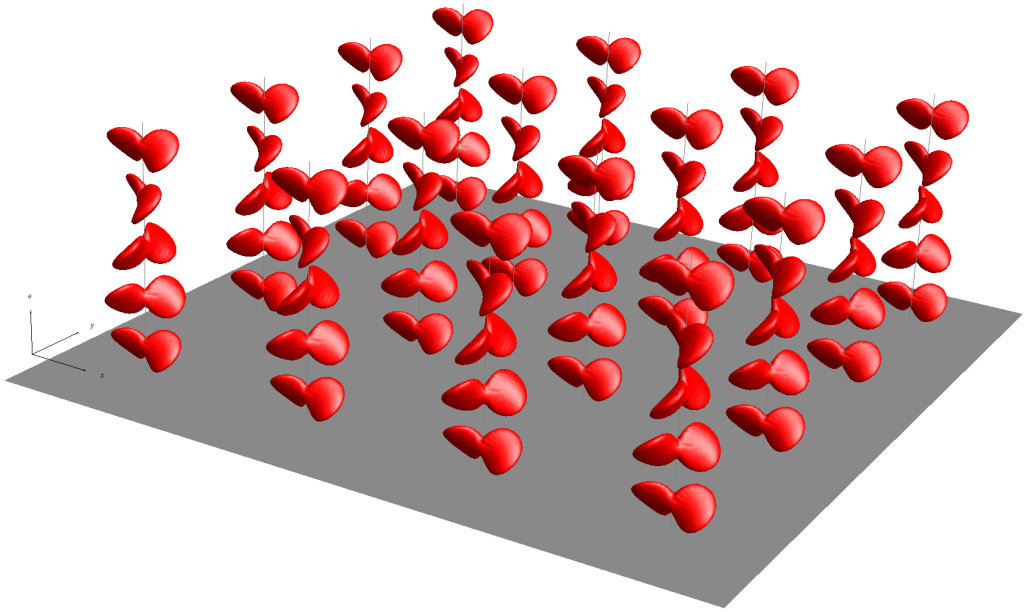


Figure 6.7: This image illustrates the idealized model of the V1 cortical architecture. It is obtained substituting in Fig 4.9 the fan of integral curves by the time independent Fokker-Planck fundamental solution.

6.6 The Fokker-Planck fundamental solution

This section discusses how to compute the time independent Fokker-Planck fundamental solution. There is not a simple explicit expression useful from an engineering point of view. Therefore either we perform certain analytical approximations or we use a numerical method. The first option is discussed in Subsection 6.6.1 while the second is addressed in Subsection 6.6.2.

6.6.1 Analytical computation of the fundamental solution

The fundamental solution of FP_{SR} cannot be analytically computed in terms of simple explicit functions ([22]), so we will compute an approximation of it, by considering a first order approximation of the operator algebra of $SE(2)$. The differential operators X_i can be approximated by their first order Taylor

expansion term $\tilde{X}_i$ respectively. Therefore if θ is small then $\cos(\theta) \approx 1$, $\sin(\theta) \approx \theta$ and we therefore get:

$$\begin{cases} \tilde{X}_1 = \frac{\partial}{\partial x} + \theta \frac{\partial}{\partial y} \\ \tilde{X}_2 = \frac{\partial}{\partial \theta} \end{cases} . \tag{6.21}$$

The operators $\tilde{X}_i$ are no longer generators of the Lie Algebra of the $SE(2)$ group. The Lie algebras span a vector field formed by polynomials of the canonical base of the tangent space (in this instance ∂_x, ∂_y and ∂_θ). They are called first order Heisenberg Lie algebras and they are associated to their respective Heisenberg group. The first Heisenberg Lie algebras are nilpotent since all their second order commutators are zero while the Lie algebra of the $SE(2)$ has infinite commutators. However, the $\tilde{X}_i$ are good approximations since $|X_1 - \tilde{X}_1| \to 0$ when $\theta \to 0$, $X_2 = \tilde{X}_2$ and the dimension of the spaces spanned by the X_i and their commutators and by the $\tilde{X}_i$ and their commutators are the same. Therefore, the Fokker-Planck operator from eq. (6.17) in the $SE(2)$ can be also approximated by the Fokker-Planck operator in a Heisenberg group:

$$-X_1 + \frac{\sigma^2}{2} X_{22} \approx -\tilde{X}_1 + \frac{\sigma^2}{2} \tilde{X}_{22} = -\frac{\partial}{\partial x} - \theta \frac{\partial}{\partial_y} + \frac{\sigma^2}{2} \frac{\partial^2}{\partial \theta^2}. \tag{6.22}$$

The fundamental solution $\Gamma_H(x, y, \theta)$ of the operator from eq. 6.22 has been explicitly computed by Hormander in [40], using the Fourier transform:

$$\Gamma_H(x, y, \theta) = \frac{\alpha}{x^2} e^{\frac{-(x\theta - \frac{3}{2}y)^2 - \frac{3}{4}y^2}{x^3}}. \tag{6.23}$$

Figure 6.8 shows θ-constant planes of the function $\Gamma_H(x, y, \theta)$. Each panel show some level sets colored according to its value. As all the gray-scale images in this chapter, black indicates the maximum value and white the minimum.

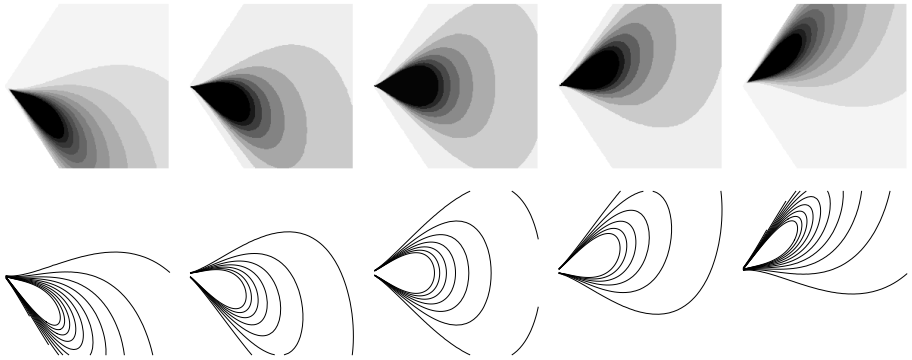


Figure 6.8: Fundamental solution of the Fokker-Planck equation in the Heisenberg group, eq (6.23). Visualized from left to right are the iso-contours (filled at the top, empty an the bottom) as the value of the angle θ increases.

The function $\Gamma_H(x, y, \theta)$ can be integrated over the third variable in order to obtain the projection over the plane xy of the fundamental solution:

$$\Gamma_H^*(x, y) = \frac{\alpha}{x^{\frac{3}{2}}} e^{\frac{-\frac{3}{4}y^2}{x^3}}. \tag{6.24}$$

This function is plotted in Fig. 6.9.

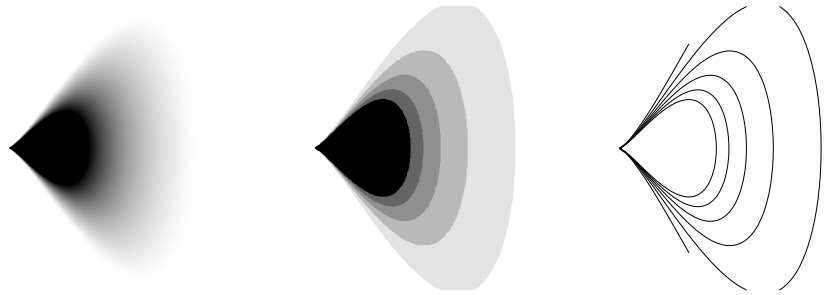


Figure 6.9: From left to right: the density map, the filled and the empty level sets of the function 6.24, which represents the projection of the fundamental solution on the image plane.

For small values of x, y and θ both $\Gamma_H(x, y, \theta)$ and $\Gamma_H^*(x, y)$ are good approximations of the Fokker-Planck fundamental solution. The main drawback is that the approximated kernel is no longer left-invariant with respect to the group law. This can be trivially checked observing that the third coordinate is no longer periodic. This is why in this thesis we use the numerical computation described in the sequel.

6.6.2 Numerical computation of the fundamental solution

We have numerically computed the fundamental solution of the equations (6.19) and (6.20), using COMSOL Multiphysics v3.5, a Finite Element Method solver.

In the following we list some of the details of the numerical simulation:

- The boundary conditions are Neumann in the spatial coordinates and periodic in the directional coordinate.
- The distribution $\delta(x, y, \theta)$ is numerically approximated by a Gaussian normalized function:

$$\delta_\epsilon(x_i, y_i, \theta_i) = \left(\frac{1}{\epsilon\sqrt{\pi}}\right)^3 e^{-\frac{1}{\epsilon^2}(x_i^2+y_i^2+\theta_i^2)}. \tag{6.25}$$

  where (x_i, y_i, θ_i) are mesh points.
- The solution is computed on an adaptive mesh.
- The corresponding linear system solver uses a direct method based on the library UMFPACK.

Figure 6.10 shows some cut-planes for different values of θ of the numerically computed fundamental solution (only one branch). The colormap is the same as that of Fig. 6.8.

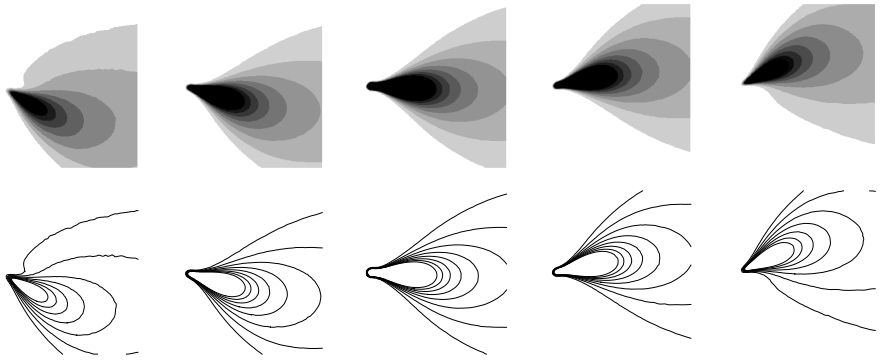


Figure 6.10: Fundamental solution of the Fokker-Planck equation calculated with COMSOL. Visualized from left to right are the iso-contours (filled at the top, empty an the bottom) as the value of the angle θ increases.

Plotted in Figure 6.11 is the projection on the xy-plane of the fundamental solution computed with COMSOL.

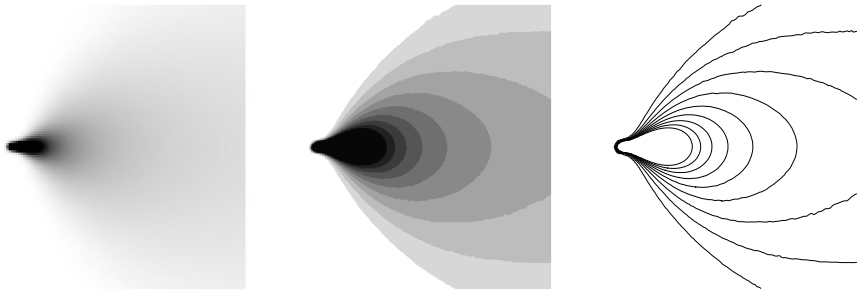


Figure 6.11: From left to right: the density map, the filled and the empty level sets of the projection of the COMSOL calculated Fokker-Planck fundamental solution on the image plane.

Fig. 6.6 shows an iso-surface visualization of the computed kernel. The descriptions of the kernel and some of it uses were given in the previous section. In the next chapter the Fokker-Planck fundamental solution is used to model certain statistics in natural image databases. The introduction of this kernel is the main contribution of this chapter, in the following section it is compared with the Tensor Voting method.

6.7 Comparison with G. Medioni's Tensor Voting method

In this section we compare the propagation model introduced previously in this chapter with a well known technique of perceptual completion, Gerard Medioni's Tensor Voting [58].

The Tensor Voting had a lot of success during the last ten years in the computer vision community. It tries to propagate the contour information using a mixture of the Gestalt rules of good continuation and proximity. It operates on a space of second order positive definite tensors which represent the local orientation and its uncertainty. The method was developed heuristically but the authors claim that has connections with the biological perceptual mechanism. This section is devoted to reveal the connection between the Tensor Voting and our model of the visual cortex. Tensors are intended as second order approximations of the hypercolumns. The stick voting field (which in the tensor voting is the kernel used to propagate the direction

information) will be compared to our models of lateral connectivity, the integral curves of $X_1 + kX_2$ and the Fokker-Planck fundamental solution.

6.7.1 From orientation distribution to tensors

In image processing, a typical representation of the local orientation at a location is provided in the form of two-dimensional non negative definite tensors. Each tensor indicates the saliency of a perceptual structure at that location.

A two-dimensional tensor T in this context is a function $\mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ given by:

$$T(u, v) = v^T M_T w,$$

where

$$M_T = \begin{pmatrix} T_{11} & T_{12} \\ T_{12} & T_{22} \end{pmatrix}$$

is a symmetric non-negative definite matrix and v, $w \in \mathbb{R}^2$ are two dimensional vectors. Since the matrix M_T fully describes the tensor, both the tensor and the matrix are identified ($T = M_T$).

A tensor is also equivalent to an ellipse where the axis of the ellipse are the eigenvectors of M_T and their aspect ratio is the ratio of the eigenvalues. The major axis is in the direction of a potential curve passing through the location. The shape of the ellipse indicates the certainty of the preferred orientation. In the limit, a degenerate ellipse with only one non-zero eigenvalue represents a perfectly oriented point. These ellipses are called *stick* tensors. On the other hand, a *ball* tensor is an ellipse with equal eigenvalues and it represents a location without a preferred orientation. The size of the tensor encodes the saliency of the information. Larger tensors convey more salient information than those smaller.

Let λ_1, λ_2 and e_1, e_2 be the eigenvalues and the corresponding eigenvectors associated with the tensor T such that $\lambda_1 \geq \lambda_2$. Then T may be written as the following linear combination:

$$T = \lambda_1 e_1 e_1^T + \lambda_2 e_2 e_2^T = (\lambda_1 - \lambda_2) e_1 e_1^T + \lambda_2 (e_1 e_1^T + e_2 e_2^T). \tag{6.26}$$

Note that none of the coefficients are able to be negative as the tensor is non negative definite and $\lambda_1 \geq \lambda_2$.

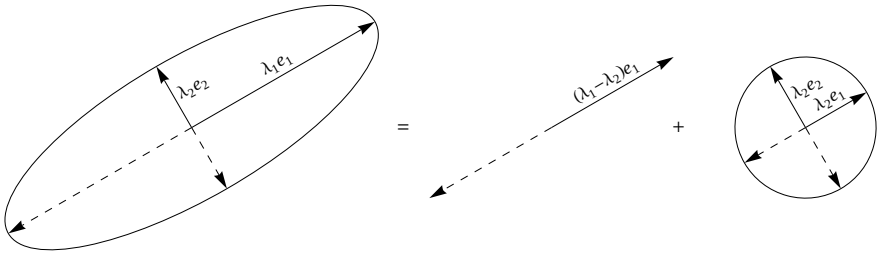


Figure 6.12: A generic 2D non negative definite tensor may be written as a linear combination of a stick tensor plus a ball tensor.

Equation 6.26 implies that T can be written always as a linear combination of a *stick* tensor $S = (e_1 e_1^T)$ plus a *ball* tensor $B = (e_1 e_1^T + e_2 e_2^T)$ (see Fig. 6.12). The stick tensor indicates the local orientation of the point (the curve saliency) while the ball component the uncertainty of that direction. If $(\lambda_1 - \lambda_2) \gg \lambda_2$, the *stick* component dominates indicating a strong curve saliency. If $(\lambda_1 - \lambda_2) \simeq 0$ there is a high uncertainty in the direction. We can distinguish two scenarios here, a small saliency (small eigenvalues) indicates that the point belongs to an homogeneous area. Otherwise, if the saliency is high, the uncertainty is a consequence of two or more local orientations. For example the point is at the intersection of two curves.

There are many ways of obtaining a field of tensors from an image, for example one may compute the so called structure tensor [48], which is computed applying two Gaussian derivatives.

The two dimensional tensors introduced before may be understood as the collapse of an orientation distribution function. As previously explained, the tensors contain more information than just the image gradient. They are essentially 3D objects. The 3D is clear if we consider the matrix representation as it is defined by T_{11}, T_{22} and T_{12} ($T_{12} = T_{21}$). A more illustrative representation of the 3 degrees of freedom is given by the 2 eigenvalues and the orientation of larger eigenvector.

Let the function $u \in \mathbb{L}_2(S^1)$ be an orientation distribution function. There are many ways to construct a tensor T_u associated to u. In this section we want to compare our model with the tensor voting, so the natural choice is integrating stick tensors associated to each possible direction and weighting them with u:

$$T_u = \frac{1}{2\pi} \int_0^{2\pi} u(\theta) S_\theta d\theta, \tag{6.27}$$

where S_θ is the unitary stick tensor associated to the orientation θ:

$$S_\theta = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} (\cos\theta, \sin\theta) = \begin{pmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{pmatrix}$$

Equation 6.27 had been used several times in the literature to represent distributions of orientation as tensors, principally in the area of material sciences (see [23] for a review). An elegant interpretation of this formula in terms of the first harmonics of the function u was given in [90].

6.7.2 The stick voting field in $SE(2)$.

The Tensor Voting method [58] is an heuristic technique for the perceptual completion problem. It proposes propagating the curve saliency information from input points to their neighbors. Input points are encoded as tensors and they propagate through tensor fields. The receiver points accumulate the tensors induced from the inputs (*votes*). The final saliency of a receiver is the sum of all the votes from the inputs. Each point could be both an input point and a receiver. The voting process can be evaluated over all the points in an image (dense tensor voting) or only on some of the points of interest (sparse tensor voting). Taking into consideration that a generic tensor can be decomposed into stick and ball parts according to eq. 6.26 and the voting fields should be invariant under rotations and translations, it is enough to specify the field generated by a stick tensor and by a ball tensor. In this thesis we are interested in the propagation of the curve saliency information, therefore, only the stick tensor field is described.

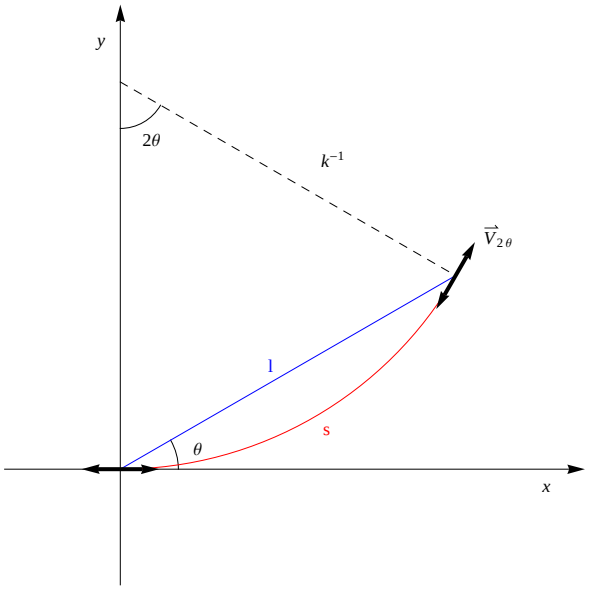


Figure 6.13: The stick voting field and its parameters.

Let O be an input point on which we know there is a curve so the input information is encoded by a stick tensor, and $P = (x, y)$ a neighboring point. For simplicity the coordinates (x, y) are taken relative to the point O and the orientation of the curve (the orientation of the eigenvector associated to the non-zero eigenvalue of the tensor) is considered tangent to the x-axis. The fundamental heuristic assumption in determining the propagation field is: in the absence of any other information, the arc of the osculating circle from O to P (the circle connecting O and P, tangent to the orientation of O) is the smoothest path. In the case of alignment between O and P, the circle degenerates to a straight line. The method is inspired in the good continuation Gestalt principle, indeed it is in accordance with the association fields. The second assumption is that the influence of the input attenuates with the distance along the osculating curve and with the curvature of the curve. The above consideration was used to define the voting field S_O generated by a stick tensor at O. It is a tensor field of stick tensors only, each one tangent to the osculating circles from O and whose saliency decays exponentially with the arc length s of the osculating circle at (x, y) and with the curvature k. No vote is performed if the receiver is at an angle larger than 45°. Then, the stick voting field writes:

$$S_O(x, y) = e^{-\frac{s^2+k^2}{\sigma^2}} \vec{V}_{2\theta} \vec{V}_{2\theta}^T \tag{6.28}$$

were:

$$s = \frac{\theta l}{\sin(\theta)}, \quad k = \frac{2\sin(\theta)}{l}, \quad \theta = \arctan\frac{y}{x}, \quad l = \sqrt{x^2+y^2}$$

and

$$\vec{V}_{2\theta} = \begin{pmatrix}\cos(2\theta)\\ \sin(2\theta)\end{pmatrix}$$

Taking into account that the 2D projection of the integral curves of the vector fields X_1 and X_2 are circles, and that the length of the curves (see eq. (4.7)) are $\|\gamma'(s)\|^2 = s^2 + k^2$, we can then express 6.28 as:

$$S_O(x,y) = e^{-\frac{1}{\sigma^2}\int_0^1 \|\gamma'(t)\|dt}\Pi(X_1)\Pi(X_1)^T, \tag{6.29}$$

where $\Pi(X_1)$ is the 2D projection of the vector field X_1. Note that (6.29) outlines the relationship between the tensor voting and our model. By construction the stick voting field is always tangent to the integral curves of $X_1 + kX_2$. The decay function of the stick field is proportional to the length of the osculating circle from the origin combined with the curvature of this path. This is coherent with the choice of the $SE(2)$ metric proposed in this thesis. Visualized in Fig. 6.14 is the decay function $|S_O(x,y)|$. The figure should be compared with Fig. 6.11 which is the projection on the image plane of the Fokker-Planck kernel.

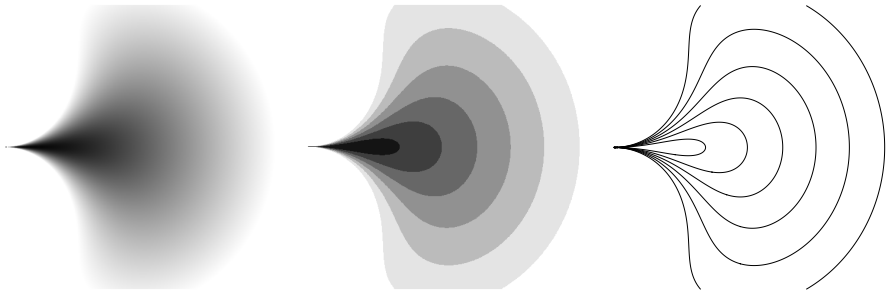


Figure 6.14: The tensor voting energy decay function is the modulus of the stick field of eq. (6.28). Visualized from left to right are the density image (the gray level proportional to the energy), some filled iso-contours and the corresponding empty iso-contours.

The tensor field associated to $S_O(x,y)$ (which is essentially a vector field), is depicted in Fig. 6.28. This field should be compared with Fig. 6.16 which displays the tensor representation of the Fokker-Planck fundamental solution numerically computed. This tensor field is the result of applying (6.27) to the fundamental solution.

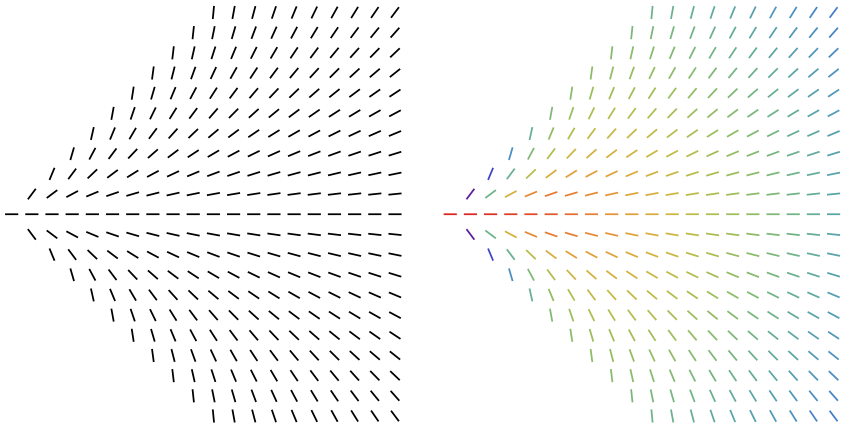


Figure 6.15: Stick voting field. Visualized in the figure above is one branch of the voting field which is essentially a vector field. Shown on the left are the orientations only while on the right the orientations are colored according to the decay function (6.28).

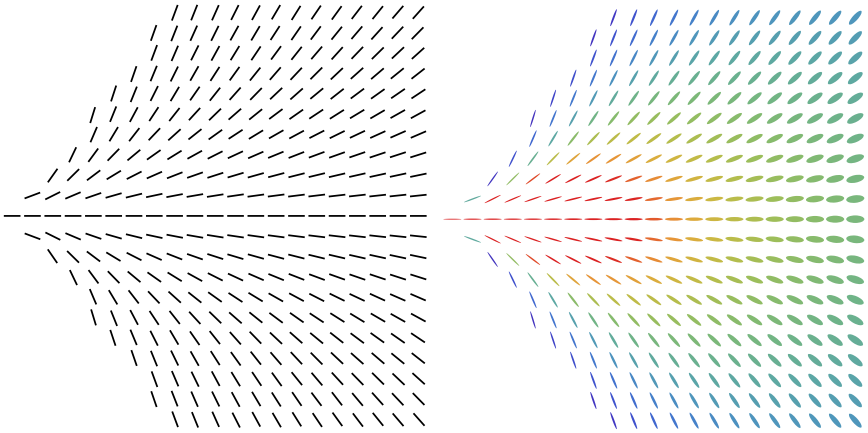


Figure 6.16: The 2D projection of the Fokker-Planck fundamental solution numerically computed. Left: The orientation of the maximum value over each (x, y) point. Right: The tensor representation calculated according to (6.27).

From these experiments, we may establish the connection between the tensor voting and our $SE(2)$ model of the cortex. The stick field propagates information as a vector field tangent to the retinal projection of the integral curves of $X_1 + kX_2$. However it operates on a space of tensors. Then, when information arriving from different sources is combined, the tensors encode no more than 3 parameters which may be interpreted as a 3D approximation

of a distribution of orientations. On the other hand, the $SE(2)$ model stores the contribution received from the sources within a function defined on S^1 associated to each point.

6.8 Discussion

In this chapter we introduced a stochastic framework in the $SE(2)$ model of the cortex. Then, we reviewed a set of well-known computer vision methods for boundary propagation (all of them inspired by the seminal work of D. Mumford [62]) which may be interpreted within the sub-Riemannian framework. All these approaches are based in the Fokker-Planck operator (6.17). We proved that this operator respects the $SE(2)$ symmetries and is coherent with the sub-Riemannian structure. A comparison between the Fokker-Planck and the sub-Riemannian heat operator (used for diffusion of the cortical activity in the previous chapter) was made. We believe that the Fokker-Planck operator is better suited as a stochastic model of the lateral connectivity and as a propagation kernel of the orientation information. However the discussion about which operator performs best is still an open problem. See for example the recent preprint [10] in which the authors proposed the sub-Riemannian heat kernel for perceptual completion.

Subsequently we have shown that the Fokker-Planck kernel fits very well the edge co-occurrence distribution of natural images. This statistic is closely related to Mumford's stochastic process which models the reappearance of occluded contours.

Last section was devoted to the interpretation of a famous perceptual completion algorithm, the tensor voting, in terms of the $SE(2)$ geometry and it was proved that there exists a strong connection between them. The most interesting comparison is perhaps the relationship between fibers (as distributions of orientations) and second order tensors. This comparison opened the interesting question about what geometrical structure is encoded by the visual cortex. Three candidates are in order:

- a section of the fiber bundle, encoding an orientation vector field as a function of the position in the retinal plane;
- a second order tensor field (like in [58] or in [13]) as a function of the position encoding more information about orientation distributions;

- a distribution of probability defined on the full fiber bundle as a function of the of position and orientation θ and corresponding to a infinite dimensional tensor field on the retinal plane.

Even if there is not a unique answer to the problem of geometrical encoding, from our analysis it emerges that there is no apparent reason for which the cortex should not use the full fiber bundle structure to encode geometrical information. This structure is really richer than the ones proposed in the literature based on vector coding and tensor voting. This aspect is strongly outlined in every part of this thesis where it is stressed the importance to perform image analysis in the full Lie group instead of on the image plane. In Chapter 5, our interest has been focused in defining multiple local maxima on fibers; in Chapter 6 the Fokker-Planck kernel in the full $\mathbb{R}^2 \times S^1$ has been introduced; in Chapter 7 it will be outlined that statistical distributions of co-occurrences naturally lives in the full $\mathbb{R}^2 \times S^1$.

Chapter 7

Natural image statistics in SE(2)

7.1 Introduction

The objective of this chapter is to compute and model the distribution of orientated edges in natural scenes within the rototranslation group $SE(2)$. This work was published in [81].

In order to estimate the probability of co-occurrence of oriented edges, we improved a standard approach [51, 86, 24, 70, 3] which computes co-occurrence histograms in large natural image databases. The experimental setup consists of first detecting edges and their orientations, and then calculating a cross-correlation histogram. The edge detection mechanism is inspired by the primary visual cortex architecture. Emulating the action of simple cells in V1, a bank of oriented edge detection filters is applied to each image in order to measure the local orientation. If the response over a point (x, y) is strong enough, the point is classified as a contour with orientation θ (the angle of the filter with a maximum output). Therefore, oriented edges are represented as points $(x, y, \theta) \in \mathbb{R}^2 \times S^1$. Then, a cross-correlation between the detected edges is computed. Assuming a translation invariance a 4D histogram is obtained in $\mathbb{R}^2 \times S^2$ where the coordinates are the coordinates of the relative position and the respective angles. Finally, taking into consideration invariance of rotations an histogram in $\mathbb{R}^2 \times S^1$ is constructed. The first two coordinates are the coordinates of the relative position with respect to the primary edge while the orientation difference is the third coordinate.

The probability in the group $SE(2)$ is modelled using both a deterministic and a stochastic model. In the first instance, we show that the connections

between different edges computed as integral curves of the histogram are in agreement with the association fields [29] and with the integral curves of the cortical-inspired contact structure of the rototranslation group introduced in Chapter 4. In the stochastic model, the connectivity between the points in the group becomes a probability of co-occurrence, regulated by the fundamental solution of the stochastic advection-equation in the group, i.e. the time independent Fokker-Planck equation introduced in Chapter 6. The model equations, while discretized and simulated, are able to predict the qualitative and quantitative behavior of the probability density of edges co-occurrence in natural images and a parametric identification allows us to estimate the variance of this process.

We start this chapter by reviewing the basic concepts of the natural image statistics and some previous work. Then, in Section 7.3 we introduce a novel methodology in computing the edge co-occurrence probability in natural images and visualize the results. In Section 7.4 we describe the deterministic model and in Section 7.5 we describe the stochastic model for the edge co-occurrences. Comparisons between experimental data and numerical simulation and parametric identification of the variance of the stochastic process are also provided.

7.2 Background

7.2.1 Natural image statistics

There is no doubt that there is a strong connection between the design of our sensory system and the statistical properties of the environment in which we live. This connection is implemented through a mixture of adaptations which are genetically determined (hardwired) and adaptations which adjust to the environment during our lifetime (plastic) [31]. This link is clear in the design of our sensory organs. For example, the physical properties of light or sound were a clear driving force behind the evolution of our eyes and ears.

From another point of view, as illustrated in Chapter 2, there is not a deterministic connection between our phenomenological perception and the physical world around us. Most perceptual capabilities depend on the combination of many sources of stimulus and their organization. Each of these sources is only probabilistically related to our perception, and the final phenomenological representation is a combination of all these sources. Moreover, the physiological measurement of the physical signals is contaminated by noise and distortions of all kinds. Consequently, it is appropriate to use statistical

terms in order to characterize the natural stimuli. In particular, if we restrict our attention to the visual system, the stimuli are the natural images or more explicitly images of natural scenes. These images are projections of the environment in which we live.

We can see a wide variety of images when observing the world around us. This may lead us to the idea that natural images are quite random. Natural images, however, are not white noise so there must be certain structures which we are able to use to characterize them. Based on this conjecture, there have been enormous amounts of work during the last 20 years trying to reveal the underlying structure of the visual stimulus. The application field of this research is very wide, including efficient image compression, image analysis, learning and modelling of the perception and the biological visual system. Two reviews on the statistical modelling of natural images and their relationship with human vision can be found in [87, 31].

Many kinds of statistics can be calculated, the most simple one can think of, is to describe the distribution of the image values at a single position assuming there is translation invariance, i.e. the natural images are still natural when they are translated. Therefore, these statistics, called first order statistics, are the same for each point in the image, making them very easy to calculate since the data can be accumulated at one point. In the case of gray scale images for example, the distribution of light intensity values describes the range one can find in a natural environment.

Natural images have the property that separate positions are not independent. This property cannot be captured by first order statistics. The mutual dependence between two points is described by second order statistics. The most popular statistics of this type is the autocorrelation function, measuring the correlation between the intensity at two points of an image, and the power spectrum: the Fourier transform of the autocorrelation function. A classical result is that in natural images the autocorrelation function behaves as $1/f^2$ where f is the spatial frequency [27]. Unlike white noise where its spectrum is uniformly distributed, here it is concentrated at low frequencies. A second very important property is its scale invariance.

Even though many properties of interest in natural images can be explained by means of second order statistics, high order statistics (encoding the mutual relationship between three or more points) are needed to describe other characteristics like the distribution of local orientations, edges, corners, textures and many other features. Because it is very difficult to evaluate these statistics directly, the images are often transformed in order to convert these high order statistics to first or second order.

7.2.2 Past work on edge distributions

Many experiments have been conducted to characterize the statistical distribution of the position and orientation of gradients in natural images. Experiments aimed at estimating the statistical distribution of co-occurrence of position-orientation couples from large databases of natural images. In those experiments two types of structures can be identified with different statistical properties. The first is a co-circular, scale invariant for small angle differences and in accordance with the Gestalt rule of good continuation. It has been interpreted as due to edges belonging to the same contour. The second is a parallel structure (parallel edges are statistically significant) mainly due to the presence of textures. In this Chapter we will address only the co-circular structure from both a deterministic and a stochastic point of view. The parallel structure will be analyzed with similar instruments in Chapter 9.

In the sequel we review some of the previous experiments on natural image statistics which inspired our work.

Krüger

The first work on statistics of co-occurrence of edges is due to Krüger [51]. He computed a 4D co-occurrence histogram in a database of 98 outdoor natural images (512×512 px) and considered 4 different orientations using Gabor filters (a squared sum of the response from both the odd and even kernel is used as a measurement of the edge energy [61]). Some of his results are illustrated in Fig. 7.1. He observed an elongated structure correlating collinear edges. In order to prove that this structure is independent from the shape of the filters, he tested the methodology on images of white noise and the structure disappeared. Nevertheless, the low angle resolution of the histogram did not allowed him to note the co-circularity pattern reported in subsequent works. A second observation is the presence of a less important but still significative parallel structure which can be seen in the level curves corresponding to low values which looks like a wide cross.

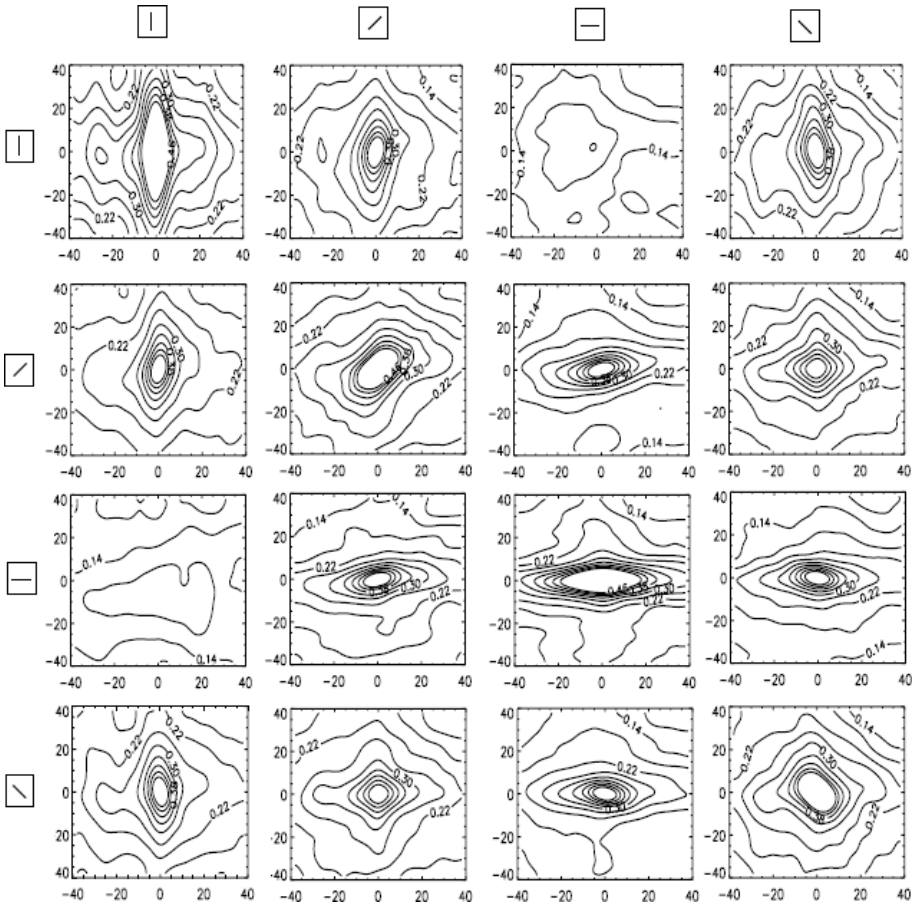


Figure 7.1: Level sets of the co-occurrence histogram reported by Kruger [51].

The two following works presenting correlation reports were published almost contemporaneously at the beginning of the decade and are probably the most cited papers in the field.

Geisler et al.

The work of Geisler et al. [32] was done using 20 representative natural images (512×512 px), filtered with Gabor kernels at 18 different orientations between 0 and π (every 10 degrees). A drawback of this work is that they manually labelled all the edges belonging to the same contour, making the approach non-automatic. This very in depth post-processing allowed them to get rid of the parallel structure. Rotation invariance is assumed by considering a 3D system of coordinates (d, ϕ, θ) formed by the distance d between the two elements, their orientation difference ϕ and the direction relative to the second

edge θ (see Fig 7.2 left). The resulting histogram is shown in Fig. 7.2 right where a superposition of color-coded oriented bars is used. The color of each bar is proportional to the probability (after a Bayesian normalization). The position and orientation are in accordance to the configuration. Using this visualization technique one can get an idea of the whole 3D histogram, and moreover, a graphical illustration of the variance in the relative orientation of each single relative position.

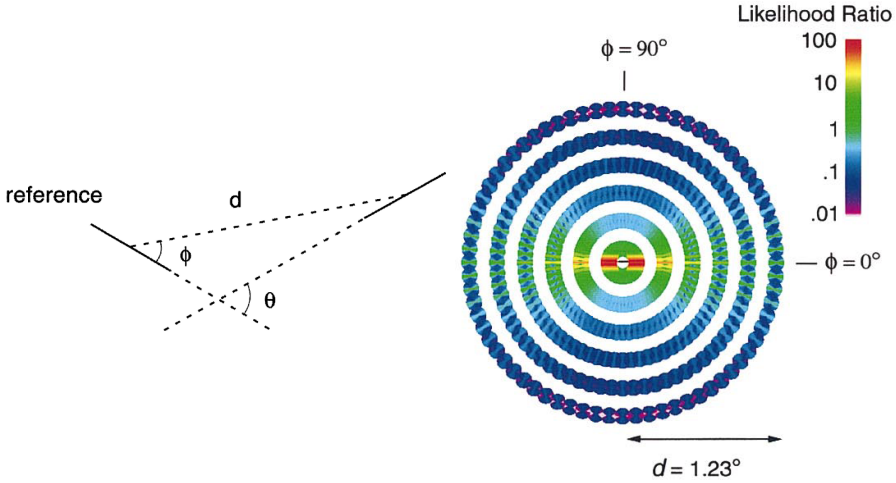


Figure 7.2: Geisler et al. experiment [32]. The coordinate system is visualized on the left and on the resulting histogram on the right.

Sigman et al.

Without any doubt, our main source of inspiration is the work from Sigman et al. [86]. They were the first to compute the co-occurrence probability in a very large database using high performance computing (the reported histograms are reproduced in Fig. 7.3). For implementation details we refer to Section 7.3, as we followed their approach. We used the same natural image database and performed the filtering using the very efficient steerable architecture with the parameters reported by them. The main difference is that we used an odd filter instead of an energy edge function. This is because we were interested in taking into account the polarity of contours as we will detail later. In [86] it was formally tested for the first time a co-circularity rule underlying the structure of natural images. The parallel structure was observed as a statistically significant structure.

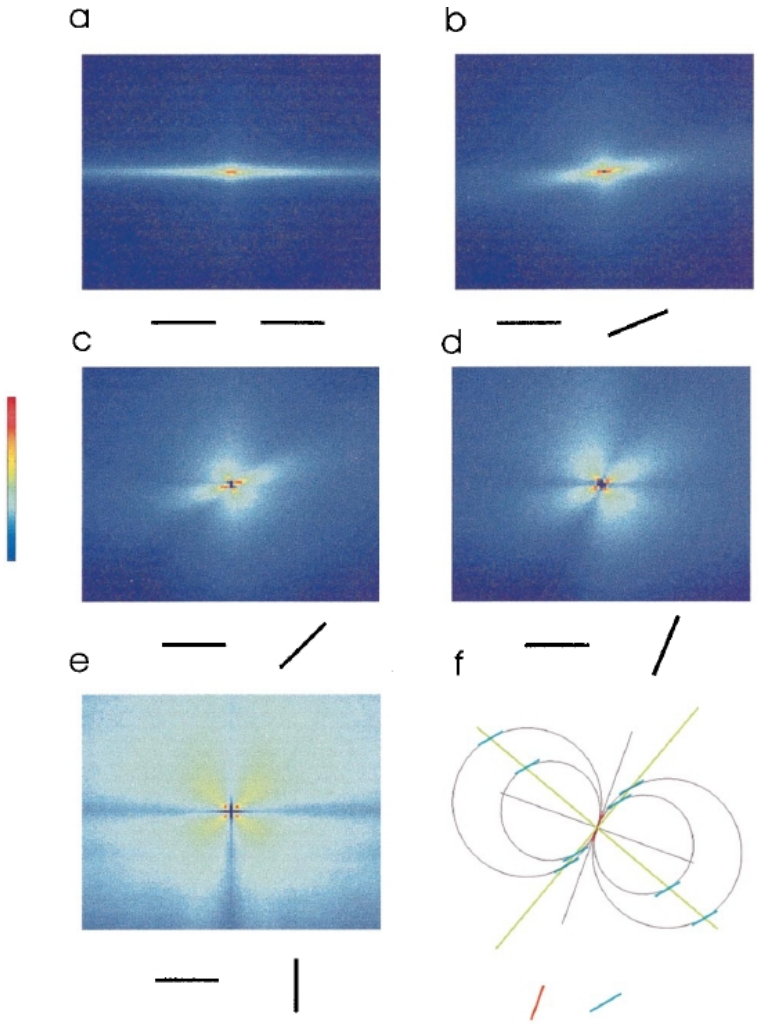


Figure 7.3: The histograms reported by Sigman et al. [86]. Their shape is interpreted in accordance with the co-circularity rule.

Elder and Golberg

In [24], using a method similar to that of Geisler et al. (using human interaction for labelling edges), Elder and Golberg formally show that the statistics of co-occurrence is invariant with respect to scale and that a co-circularity constraint underlies this distribution. A second and interesting improvement was to extend the mechanism to test the statistical power of gestalt cues other than that of good continuation (similarity and proximity). Similar results with a fully human segmented database were presented by Martin et al. [53].

Jonas and Zucker

In [3, 4], Jonas and Zucker used the correlations statistics in an image processing application. They computed the histograms for individual images showing the differences according to the type of structures present in each one (see Fig 7.4). These histograms were used to adapt or "tune" the CIRF algorithm described in Section 6.4. A characteristic of this work is the use of a very fancy edge detector filter bank [44] and the rotation invariance hypothesis which they called the "statistical homogeneity assumption".

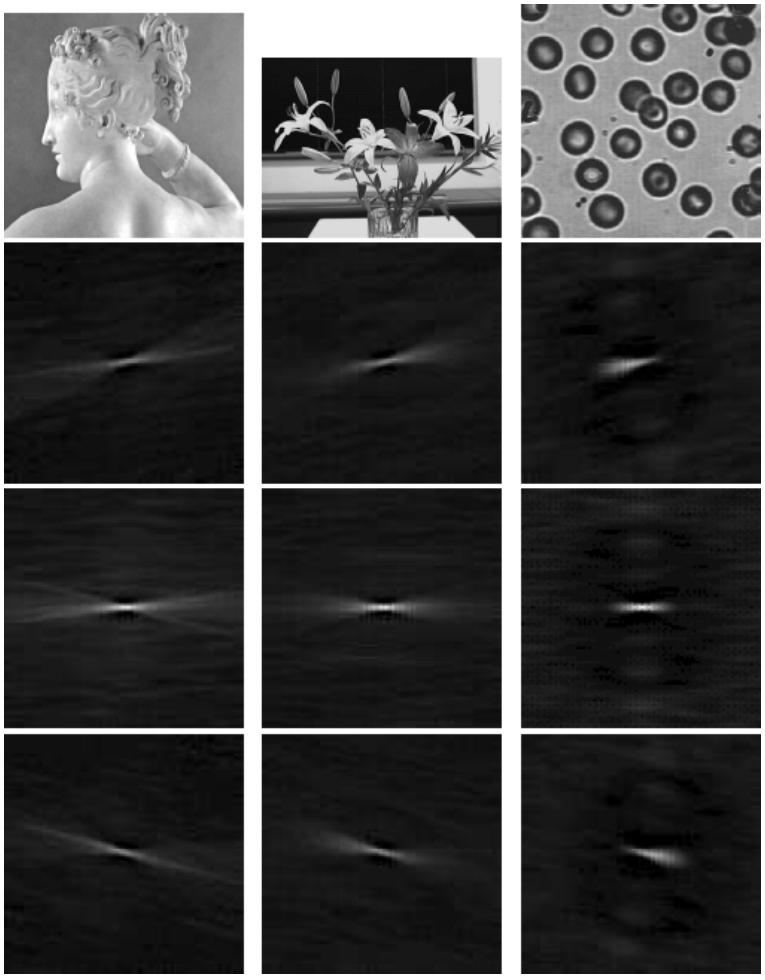


Figure 7.4: Examples of some histograms reported in [3]. Shown in each column at the top is the original image and below the corresponding histograms for 3 different orientations (image reproduced from [3]).

7.3 Natural image edge co-occurrence statistics

In this section we describeet al the technique for computing the co-occurrence histogram of edge orientations in a database of natural scenes. As indicated before, many co-occurrence histograms have been reported over the last few years [51, 86, 24, 70, 3, 32], but we adapted the approach of Sigman et al. [86] given that we considers it to be the most reliable method. We chose the same image database and the same parameters for the numerical implementation. Nevertheless our method has some differences. We decided to use only an odd-symmetric kernel for the edge detection mechanism instead of the edge energy function [61]. This choice allowed us to consider the polarity of the contrast of contours, distinguishing between image gradients from a darker zone to a lighter zone to the opposite. Therefore the considered set of filters is indexed by the whole circle $\theta \in [0, 2\pi)$. The kernels used were directional derivatives of a Gaussian filter (DoG) implemented using the steerable filter architecture which permits us to achieve a very efficient computational implementation [30]. The second improvement is the rotation invariance hypothesis which was reported and tested in some of the previous histogram reports [24, 3] but was not considered in [86].

Our method can be summarized in four steps:

1. Filter each image of the database with a set of oriented edge detection kernels.

2. For each image perform non-maximal suppression with a threshold in order to obtain a list of points (x, y, θ) corresponding to edges with their respective orientations.

3. Count how many times two detected edges with relative position $(\Delta x, \Delta y)$ have orientations (θ_c, θ_p) and store the data in a 4D histogram in $\mathbb{R}^2 \times S^2$.

4. Project the data of the 4D histogram to a 3D histogram where the third coordinate is the relative orientation.

7.3.1 The Image Database

The image database was obtained from the website: *http:// hlab.phys.rug.nl / imlib / index.html* [91, 92] and it has been used many times in literature

to compute natural image statistics [88, 86]. It consists of 4000 high quality gray scale digital images, 1536×1024 pixel and 12 bits in depth. Analyzing images captured by digital still cameras have been the most common method of measuring natural scene properties. In Fig. 7.5 some representative images from the database are shown. The images are all outdoor photographs and were taken in strict natural scenarios as in human modified environment (rural and urban).

Figure 7.5: Examples of some natural images taken from the image database [91] and used for computing the contour statistics.

7.3.2 Steerable filtering

We used the steerable filter architecture from [30] in order to process the image database because of its computational efficiency. In our calculations, as in many other applications, we needed to find the response of a filter at many orientations, i.e. each kernel is the rotation of a mother kernel. Instead of performing one filtering for each possible orientation, one can apply fewer filters corresponding to fewer angles and interpolate between the responses. With the correct filter set and the correct interpolation rule, it is possible to determine the response of the filter at arbitrary orientations without

explicitly applying it. Therefore, the basic idea for reducing computations is to synthesize filters of arbitrary orientations from linear combinations of basis filters.

We decided to use first order derivative of Gaussian (DoG) filters as orientation detectors. Shown in Fig. 7.6 are oriented DoG filters for 8 different orientations. As seen in Chapter 3 and 4 these filters are directional derivative operators and are often used to model simple cells of V1 which are also intended as local orientation extractors.

The computation of the steerable basis in this instance is very simple. Let G^θ be a first derivative of a Gaussian function in the direction of θ (we set the variance of the Gaussian function $\sigma = 1$ for simplicity):

$$G^\theta = -2(x\cos\theta + y\sin\theta)e^{-\frac{x^2+y^2}{2}}. \tag{7.1}$$

This function can be trivially re-written as:

$$G^\theta = \cos(\theta)G^0 + \sin(\theta)G^{\frac{\pi}{2}}$$

where G^0 and $G^{\frac{\pi}{2}}$ are the first x-derivative and y-derivative of a Gaussian function, respectively:

$$G^0 = -2xe^{-\frac{x^2+y^2}{2}},\ \ G^{\frac{\pi}{2}} = -2ye^{\frac{x^2+y^2}{2}}.$$

This means we can obtain a filtered image corresponding to any oriented kernel by taking linear combinations of the images filtered with G^0 and $G^{\frac{\pi}{2}}$. Therefore, our steerable basis is $\{G^0, G^{\frac{\pi}{2}}\}$, meaning that only two 2D convolutions are needed to generate the whole set of filtered images.

Finally we observe that both G^0 and $G^{\frac{\pi}{2}}$ are separable kernels:

$$G_1^\theta = \left(-2xe^{-\frac{x^2}{2}}\right)\left(e^{-\frac{y^2}{2}}\right),\ \ G_1^{\frac{\pi}{2}} = \left(-2ye^{-\frac{y^2}{2}}\right)\left(e^{-\frac{x^2}{2}}\right), \tag{7.2}$$

thus reducing the 2D convolutions to the concatenation of two 1D filters.

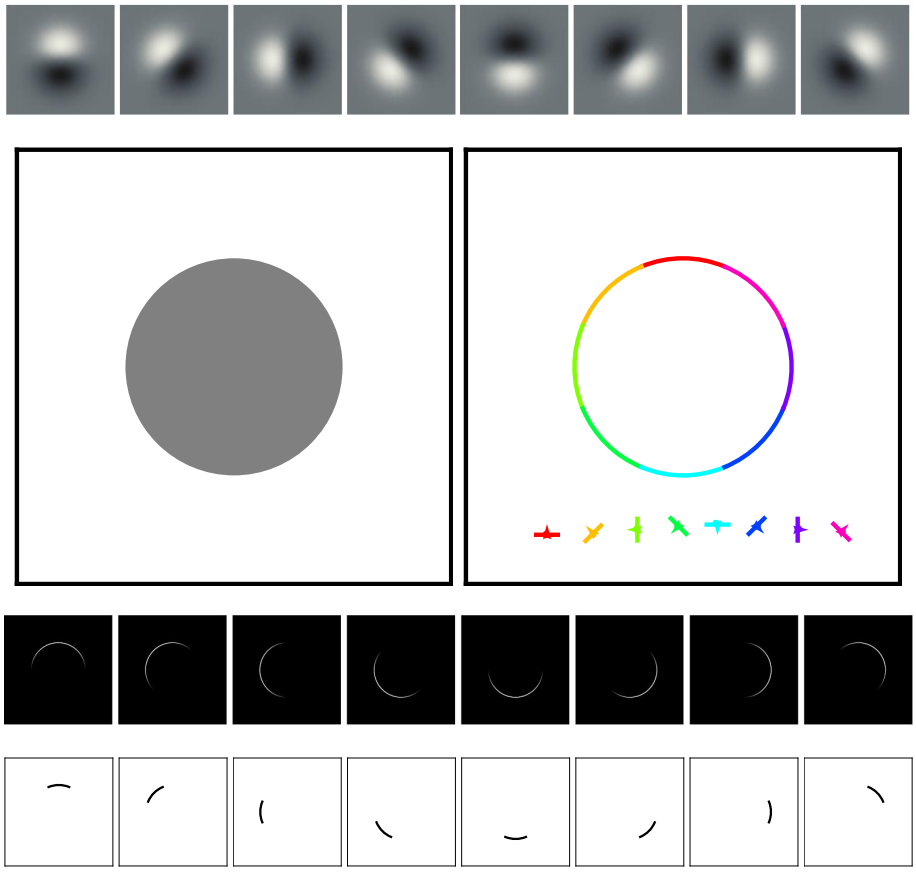


Figure 7.6: The edge detection mechanism is inspired from the architecture of the primary visual cortex since we try to reproduce the hyper-columnar architecture using the bank of DoG filters (top row). In this figure only 8 different orientations were considered. Each filter (top) imitates the impulse response of the simple cells of V1. The test image is a circle. The output of the filtering are the gray level images. Last row is the result after non maximal suppression. A circular color code scheme is used to represent the orientations detected in the central colored image.

In our implementation, the variance of the DoG filter is set to $\sigma = 1$ and a support of 7 pixels is chosen following [86]. The orientations were discretized at 32 different values $(-15\frac{\pi}{16}, -14\frac{\pi}{16}, \ldots, -\frac{\pi}{16}, 0, \frac{\pi}{16}, \ldots, 15\frac{\pi}{16}, \pi)$.

For each image the filtering process produced a stack of filtered images, each one obtained by a convolution with the kernel of orientation θ. We performed non-maximal suppression consisting of selecting for each pixel the maximum output of the filters as θ changes. Then we constructed a list of triplets

containing the pixels (x, y) where the maximum exceeded a fixed threshold, and the corresponding orientation θ where the maximum is achieved.

7.3.3 Cross-correlation assuming translation invariance

For each image we have computed a list of points (x_i, y_i, θ_i), which represents the set of positions of the contour points and their orientations, and we can now estimate the statistics of co-occurrences. Given any pair of points (x_i, y_i, θ_i) and (x_j, y_j, θ_j) in the list we say that we have a co-occurrence of two points with relative positions $\Delta x = x_j - x_i$, $\Delta y = y_j - y_i$ and with orientations θ_i and θ_j. Hence every co-occurrence is represented by a quadruplet $(\Delta x, \Delta y, \theta_i, \theta_j)$, which we store in a four-dimensional histogram. In other words, we count how many times the quadruplets occur. In this procedure we only take into account couples of oriented points satisfying $|\Delta x|, |\Delta y| < d$. The translation invariance is reflected in the fact that we are considering relative positions of the edges.

The same procedure is repeated for all considered images and the result is accumulated in a 4D histogram.

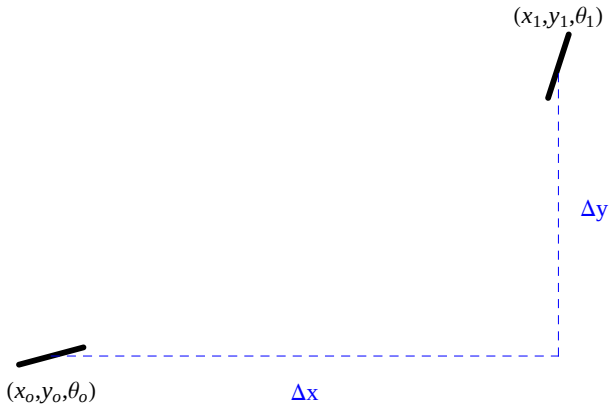


Figure 7.7: Geometric configuration of two edges in an image. The black bars represent the first edge at position (x_o, y_o) with orientation θ_o and the second at (x_1, y_1) with orientation θ_1. The co-occurrence histogram is defined with respect to the $(\Delta x, \Delta y, \theta_0, \theta_1)$.

Within this work we chose $d = 32$ pixels so that the total size of the histogram is $65 \times 65 \times 32 \times 32$.

7.3.4 Cross-correlation assuming rotation invariance

The rotation invariance implies that we consider relative orientations and not absolute ones. This is achieved by projecting the points $(\Delta x, \Delta y, \theta_i, \theta_j)$ of the 4D histogram on the rotated point:

$$(\eta, \xi, \Delta\theta) = (R_{\Delta\theta}(\Delta x, \Delta y), \theta_j - \theta_i) \tag{7.3}$$

where $R_{\Delta\theta}$ is the matrix of rotation of the vector $(\Delta x, \Delta y)$ by an angle $\Delta\theta = \theta_j - \theta_i$.

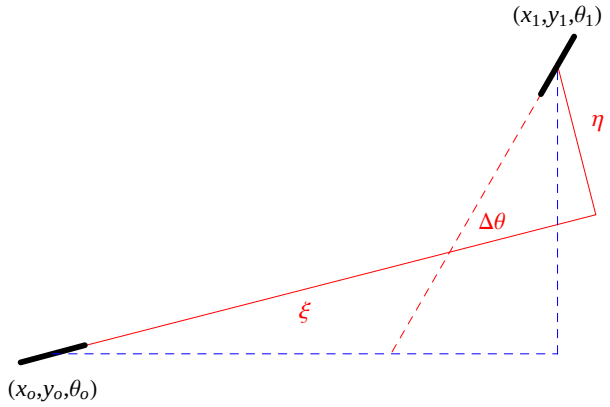


Figure 7.8: The change of coordinates allowed us to consider relative orientations so that the final histogram lives in the space $\mathbb{R}^2 \times S^1$.

Finally, we normalized the full histogram over the total number of occurrences to get the probability of observing an edge element at every possible relative position and orientation difference from a given (reference) edge element. This normalized histogram estimating the probability density function of co-occurrences of a particular geometrical configuration of contours is denoted from now on as $H(\eta, \xi, \Delta\theta)$.

With the parameters reported, the total size of H is $65 \times 65 \times 32$. The total computation time for the histogram was around 10 hours. The computation time for each image depends on the quantity of contours found on it. The mean value registered for the 4000 images in the database was approximately 10 seconds. The histograms were calculated on a Core 2 Duo 2.1GHz laptop computer and the programs where implemented in C language. Visualization of the histogram is provided in Section 7.3.5.

7.3.5 Results

In this section visualization of the computed histograms is provided. Different visualization techniques are used, each of them is important as it allows to extract different conclusions.

The histogram is essentially a 4D matrix (3D after the rotation-translation invariant reduction). Let us consider firstly the 4D case. The first 2 coordinates are the relative positions (measured in pixels) of a certain configuration of oriented edges, which is indicated by the last 2 coordinates. If the angular coordinates are fixed, we obtain a 2D image depending on the spatial variables alone. Theses 2D matrices could be displayed as histograms or simply as images where the color is proportional to the amount of co-occurrences.

In Fig. 7.9 we present an histogram (top) and an image representation (left-bottom). In both images the angles are fixed $\theta_c = \theta_p = 0$ so the histogram corresponds to two horizontal edges. In both images the x-axis is Δx and the y-axis is Δy. The colormap is the visual spectrum of light, small wavelengths (purple, blue) which correspond to low values of co-occurrences while high wavelengths (orange, red) to high probability of co-occurrence. The colormap is equalized so that the full range of colors is used. This means that the purple corresponds to the lowest value of co-occurrence while the red to the highest. The maximum value obviously is achieved at the entrance $\Delta x = \Delta y = 0$ where there is always co-occurrence. As expected, the observed structure is not isotropic (see Fig. 7.9 right-bottom). It is observed a strong correlation along the x-axis which decreases as the distance to the central point increases. This is reasonable as it corresponds to the coaxial configuration. On the contrary, on the direction orthogonal to the direction of the central edge the amount of co-occurrences quickly decays. Since the visualization in the Figures 7.9 (top) and (left-bottom) carry exactly the same informations, we will always use the last one in the sequel.

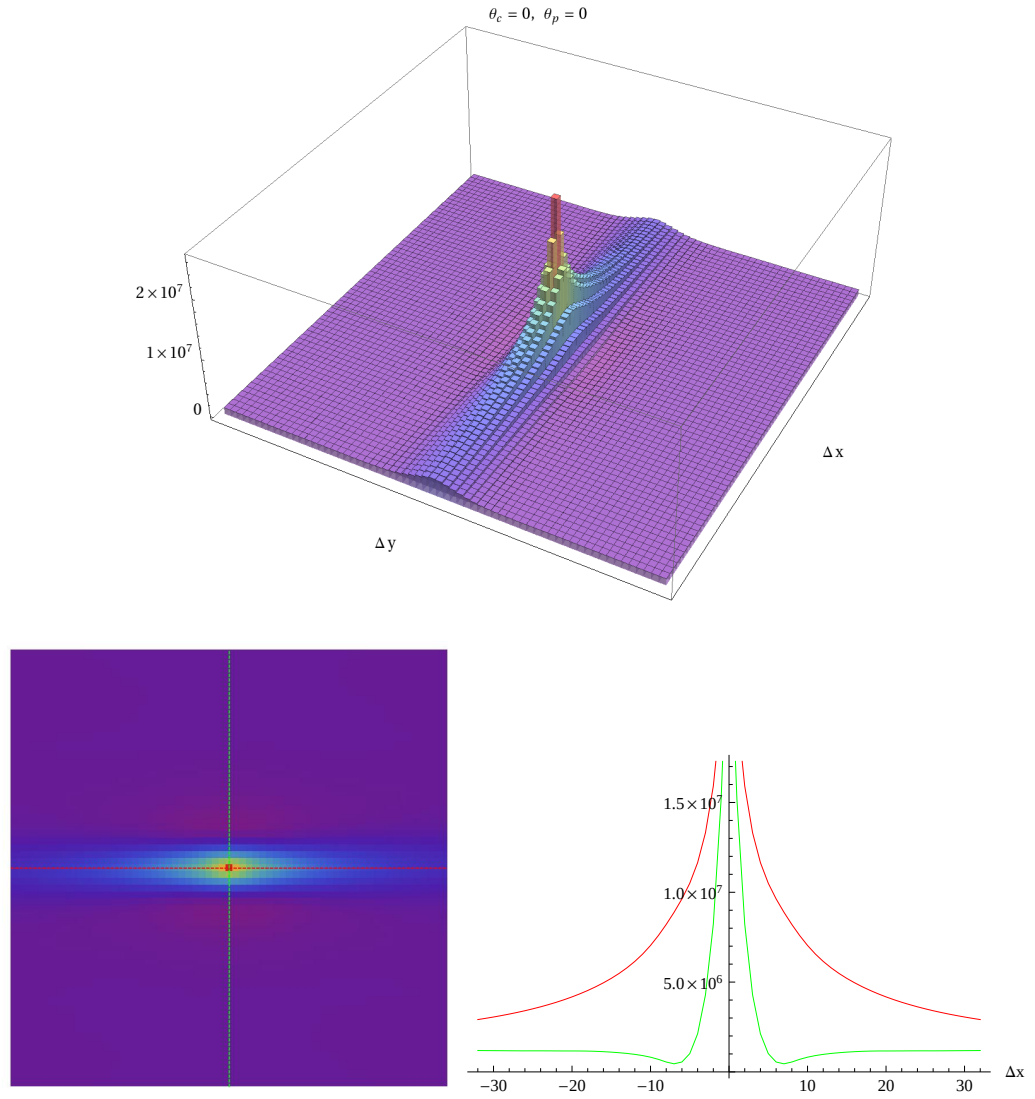


Figure 7.9: Shown in the figure is the co-occurrence histogram for the case in which both the reference edge and the secondary edge are horizontal. (Top) The height of each block is proportional to the amount of co-occurrences of the configuration $\theta_c = \theta_p = 0$. This visualization gives the idea of the histogram, nevertheless is not really helpful since the amount and dimension of data to be displayed is too large. (Left-Bottom) In this thesis the histograms will be mostly visualized as images where the color of the pixel is proportional to the number of co-occurrences, the x-axis represents Δx and the y-axis represents Δy. The colormap is equalized so that the full range of colors is used. (Right) The graphs represent the values of the histogram sampled along the red and green lines shown in the figure on the left. The amount of co-occurrences is indicated in the y-axis. A slow decay is observed along the horizontal path (the co-axial configuration, tangential to the edge) while a fast decay is observed along the orthogonal path (the trans-axial configuration).

Visualizing 4d data is quite challenging. Displayed in Fig. 7.10 is the 4d histogram with a technique similar to that used in [51] (see Fig. 7.1). Each panel contains an histogram corresponding to each possible configuration. The histograms are distributed on a grid. The orientation of the central edge changes along the rows while the orientation of the secondary edge changes along the columns, as indicated in the figure by the half white, half black circles. The set of orientations was sub-sampled from 32 to 16 for the space restriction in the figure. Each panel is equalized individually so that the full range of colors is used. For example, the image plotted in Fig. 7.9 which corresponds to the couple of angles $(\theta_c, \theta_p) = (0, 0)$ is represented in Fig. 7.10 as the first image from the left in the first row.

Analyzing Fig. 7.10, one can empirically appreciate that the rotation invariance hypothesis is verified. Each column is approximately identical to its precedent, up to a rotation of the interior structure and a shifting of the panel. Hence this implies that the histogram only depend on the angle difference, not on the angles independently.

This remark supports the rotation invariance hypothesis made in the previous section. In Fig. 7.11 the histograms are visualized after performing the dimensionality reduction. Note the similarity of this image with the first row of the 4D histogram plotted in Fig. 7.10 (or with any of the rows or columns up to a rotation and a shifting). The same histograms are plotted zoomed in Fig. 7.12 and in Fig. 7.12. The coordinates are the relative position and the relative orientation. Each panel corresponds to different value of $\Delta\theta$ which is indicated in each figure. The reference edge is considered always horizontal. The orientation of the secondary edge is indicated on right-bottom of each histogram. The images of the histograms are circular as a result of the rotation. Fig. 7.12 are displays the histograms for values of $\Delta\theta$ from 0 to π while in Fig. 7.13 are displayed those where $\Delta\theta$ takes values from π to 2π. It may be observed that the whole set of histograms is similar to the first row (or the first column) of the 4D grid in Fig. 7.10, or to any of the rows (columns) after performing an appropriate rotation. The shape of the histogram will be clear within the following sections where we will provide analysis and modelling of them.

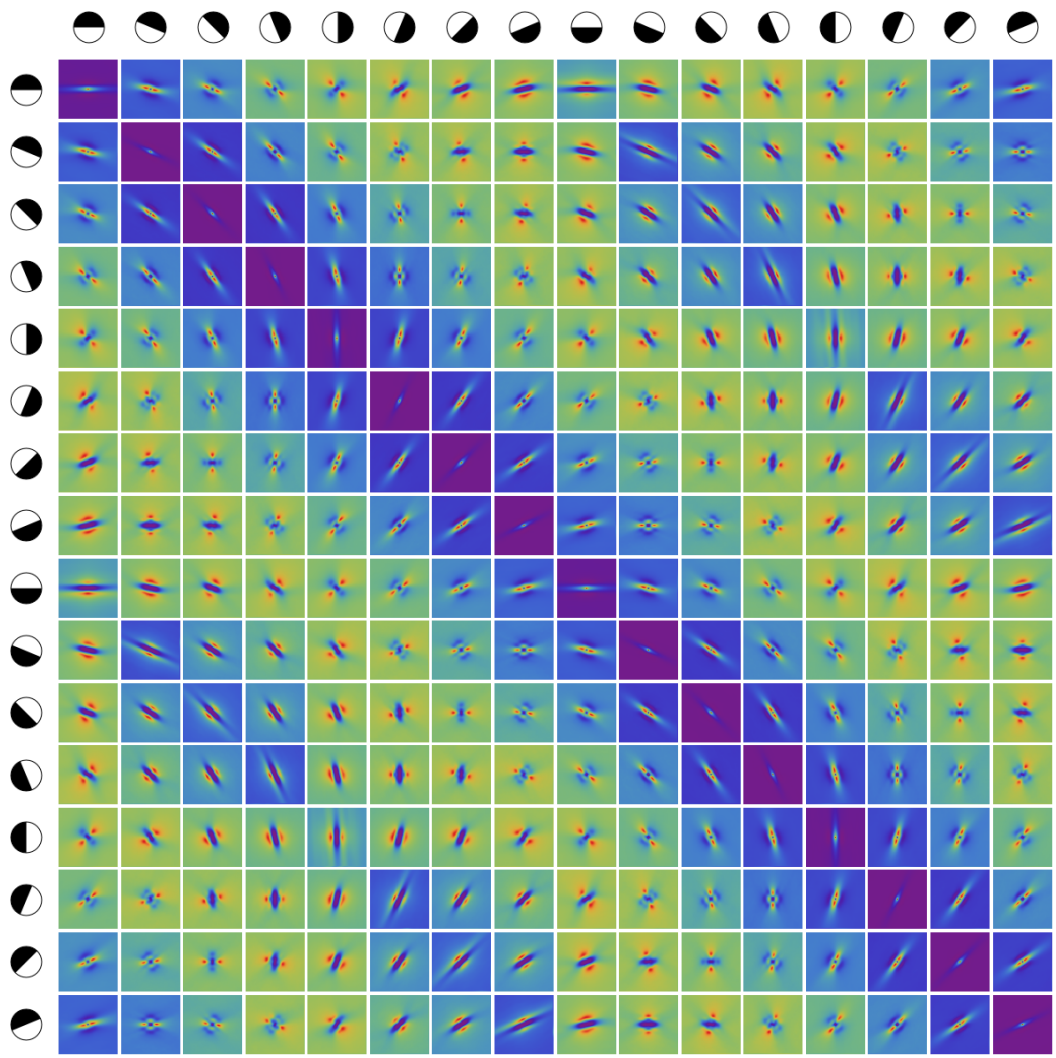


Figure 7.10: A visualization of the 4D histogram. Each panel shows an histogram corresponding to a different configuration of the two edges. The orientation of the reference edge is indicated at the top of each column while that of secondary at the beginning of each row. The coordinates of this figure are then: Δx and Δy in each panel, θ_c and θ_p in the grid of panels. For space restrictions, only 16 different orientations are plot (the full histogram has 32). From this image one can empirically verify the co-circularity hypothesis. Every row (or column) is approximately equal to that adjacent among a rotation of each panel by the same angle and a shifting of one place of the stack of panels.

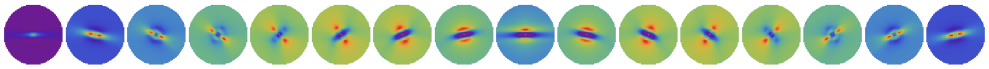


Figure 7.11: Visualization of the cross-correlation 3D histogram, obtained after preforming the dimensionality reduction. Only 16 of 32 histograms are plotted, so they can be compared with those on Fig. 7.10. All the 32 histograms are visualized in large in the 2 following figures (Fig. 7.12 and Fig. 7.13).

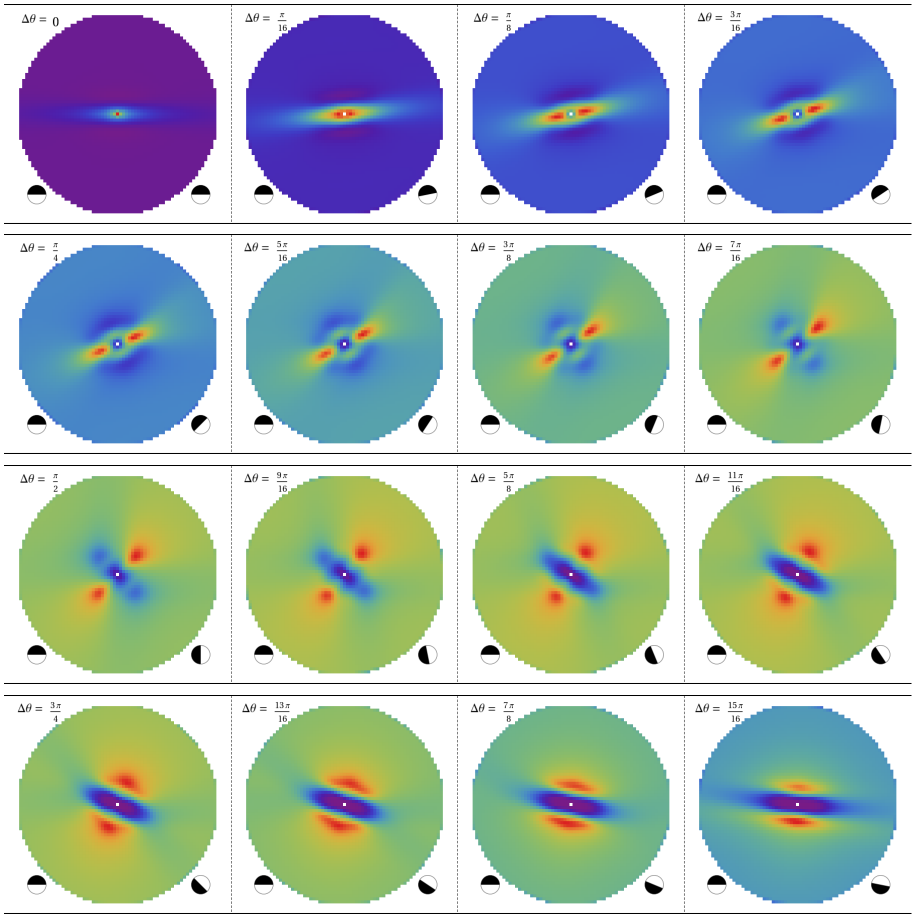


Figure 7.12: Visualization of the cross-correlation 3D histogram. Shown in each figure are the histograms for different values of $\Delta\theta$. The orientation of the reference edge is always horizontal while the orientation of the reference edge is indicated in the right-bottom corner of each panel. In this figure $\Delta\theta$ takes values from 0 to π. The histograms are circular because of the rotation step. This figure continues in Fig. 7.13 where the histograms corresponding to the remaining values of $\Delta\theta$ are shown.

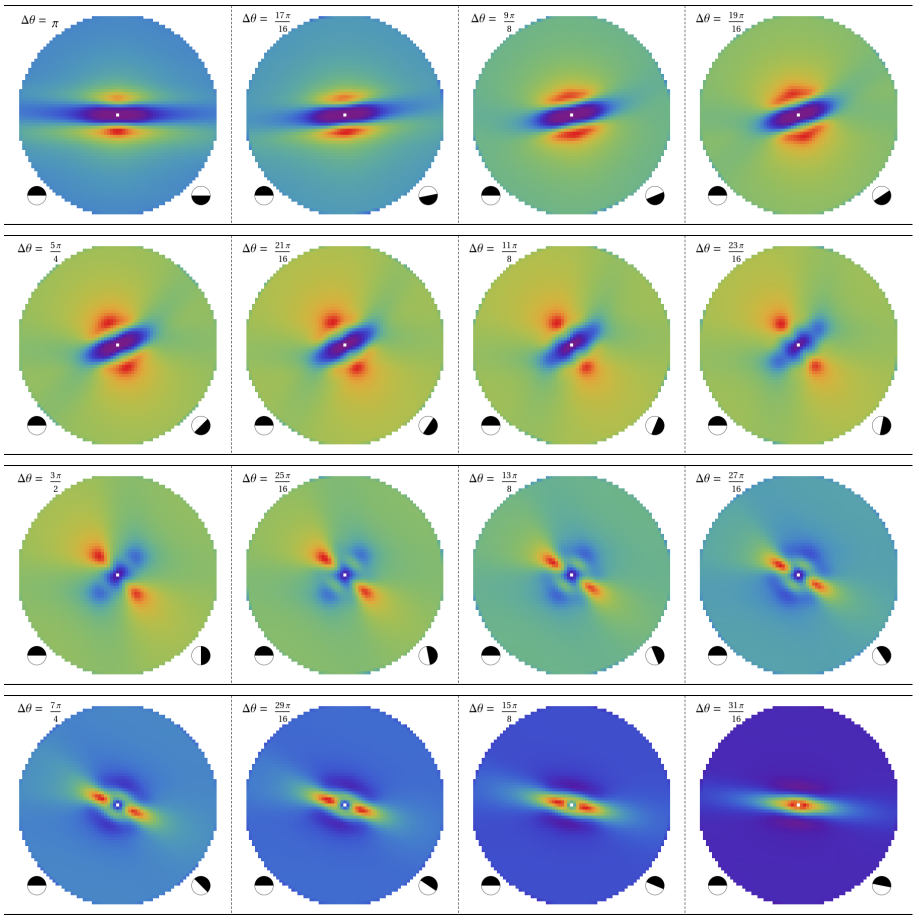


Figure 7.13: Continuation of Fig. 7.12. Shown in this figure are the histograms corresponding to values of $\Delta\theta$ from π to 2π.

7.4 Association fields obtained from the histogram

One of the most important and studied patterns in natural image statistics of contour configurations is the co-circular pattern, introduced in this field for the first time by Parent and Zucker in [71]. This property has an interesting psychophysical counterpart, i.e. the association field, (see [29]) explaining the good continuation law of Gestalt psychology. The association filed is shown in Fig. 2.9.

From the neurophysiological point of view, it is largely accepted that the neural correlate of these association fields are the long range inter columnar

connections [11, 75, 6]. These connections exist between simple cells of V1 which share almost the same orientation, and are responsible for contour integration and completion of subjective contours. Many perceptual completion models inspired by this cortical architecture explain the association field as co-circular connections. The association field has been modelled as integral curves of the rototranslation group in [14] (see Chapter 4).

The co-circular property of the edge statistics was first observed in [86]. Within our work we strongly focus on curves, and the co-circularity pattern is shown by means of a family of curves reproducing the association fields. Indeed we will deduce a 2D vector field from the histogram, compute its integral curves and compare them with the previously detected family of curves, again through integral curves.

Shown in Fig.7.14 is a diagram similar to that used in [86] to interpret the distribution of edges for each possible configuration. This figure helps understanding the shape of the histograms shown in Figures 7.12 and 7.13. Each black bar corresponds to an edge which supposes that there was an horizontal edge in the center. If the hypothesis of co-circularity is verified, edges with the same orientation are distributed around a straight line whose angle direction is half the edge orientation angle. This may be empirically checked in the histograms.

The histogram is a discrete approximation of the probability density function for the co-occurrence of edges. At each point (x, y, θ) it expresses the probability of having a contour conditioned to the fact that there exists a horizontal contour at the reference point. As in [86], we select for each point (x, y) the most probable orientation θ_m, such that

$$H(x, y, \theta_m(x, y)) = \max\{H(x, y, \theta), \theta \in S^1\}. \tag{7.4}$$

This orientation determines the direction of the unitary vector field:

$$\vec{V}(x, y) = (\cos(\theta_m(x, y)), \sin(\theta_m(x, y))) \tag{7.5}$$

depicted in Fig. 7.16. Its integral curves can be computed solving the differential equation:

$$\begin{cases} \dot{\gamma}(t) = -V(x, y) \\ \gamma(0) = (x_0, y_0) \end{cases} \tag{7.6}$$

and these are visualized in Fig.7.16.

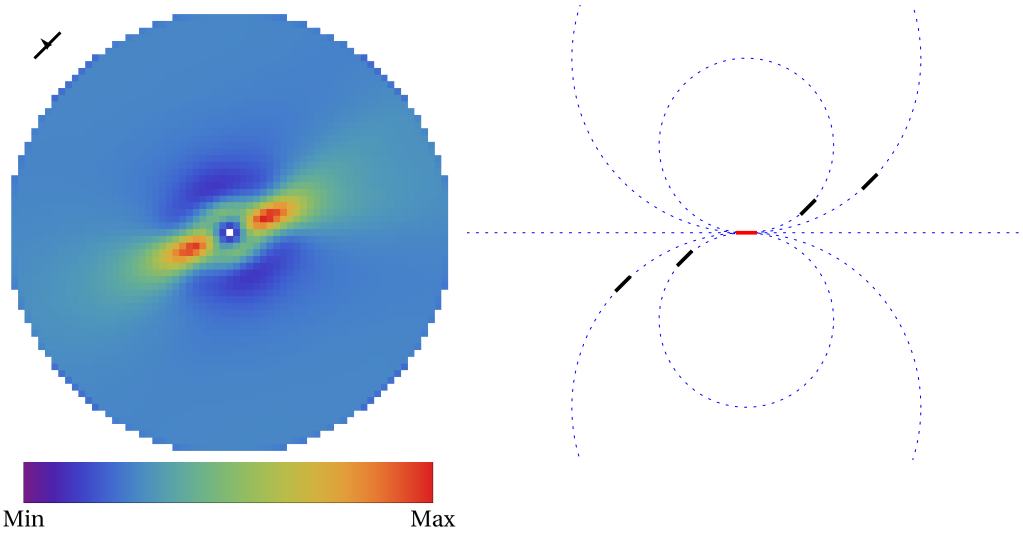


Figure 7.14: This figure gives an intuitive idea of why the co-circularity pattern is encoded in the histograms. Shown on the left is the histogram corresponding to an arbitrary value of $\Delta\theta$, which is indicated by the black bar on the left-top corner. As in the previous images, the color is proportional to the number of co-occurrences in each position. The colormap based on the light spectrum is indicated on the bottom. The image on the right represents the co-circular configuration. The blue dashed curves are circumferences tangent to the horizontal red bar. A relative edge will be co-circular with respect to the horizontal reference edge if it is tangent to one of the dashed curves. Therefore, if the orientation of the reference edge is α, the co-circular rule is respected if the edges are placed along the line with orientation $\frac{\alpha}{2}$ with respect to the x-axis. A coherent behavior in the histogram is observed.

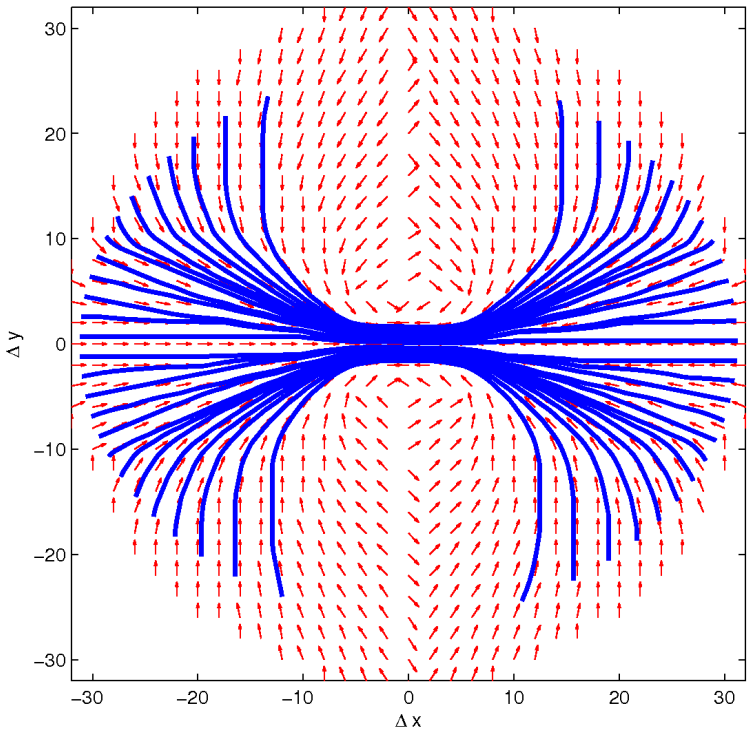


Figure 7.15: Visualized in the figure is the vector field of unitary vectors oriented by the maximal probability superimposed with its integral curves. The resulting curves are in agreement with the association field of Field, Hayes and Hess (see Fig. 2.9) and with the integral curves proposed in [14] (see Fig. 4.8).

As it is clear from the results, the integral curves of the vector field $\vec{V}(x, y)$ (Fig. 7.16) optimally approximate the 2D projection of the integral curves of the vector fields X_1, X_2 (Fig. 4.8), and both show a co-circular pattern modelling the association fields of Fig.2.9.

One can easily compute a quantitative comparison between the co-circularity rule and the most probable orientation for each position given by the histograms (the function $\theta_m(x, y)$). Co-circularity is verified if $\theta_m \approx -2 \arctan \frac{x}{y}$. Plotted in Fig. 7.16 in blue are the 2-tuples $\left(-2 \arctan \frac{x}{y}, \theta_m\right)$ for each possible relative position (x, y). All the points should be distributed around the red curve which corresponds to $\theta_m(x, y) = -2 \arctan \frac{x}{y}$ (strict co-circularity). The function θ_m only takes discrete values, therefore the points are distributed

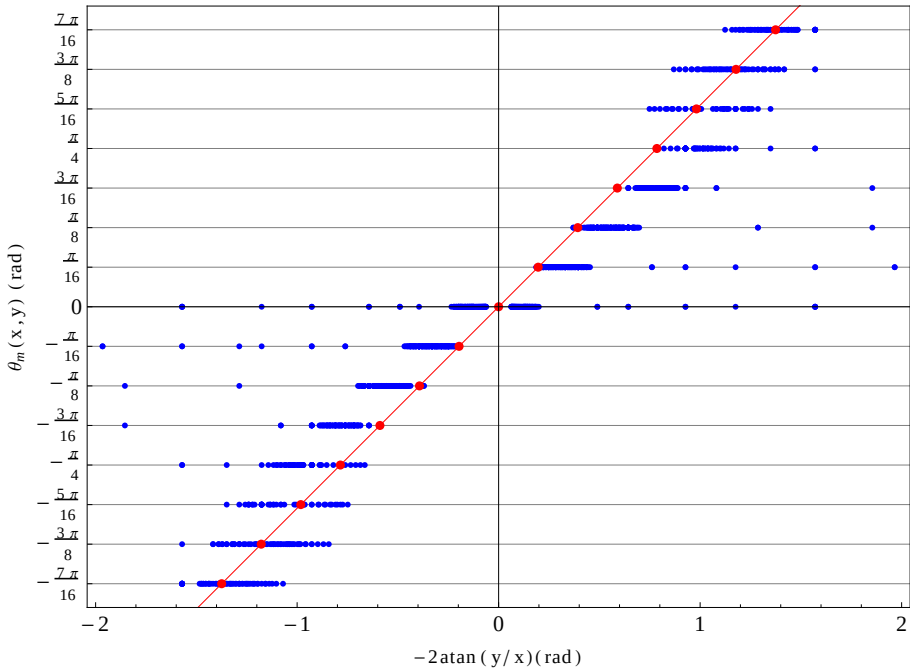


Figure 7.16: A quantitative measurement of the co-circularity rule encoded in the histograms. The blue points in the graph are the 2-tuples $\left(-2\arctan\frac{x}{y}, \theta_m(x,y)\right)$ for each possible position (x,y). The function θ_m obtained from the histogram only takes discrete values. The red curve indicates the condition of strict co-circularity $\theta_m(x,y) = -2\arctan\frac{x}{y}$. The mean square error E_θ defined in eq. 7.7 is approximately 0.15 rad and it is on the order of the quantization step ($\frac{\pi}{16} \approx 0.2$ rad).

along horizontal lines. The mean square error is $E_\theta = .15$ rad where:

$$E_\theta = \frac{1}{n}\sqrt{\sum_{x,y}\left(\theta_m(x,y) - 2\arctan\frac{x}{y}\right)^2} \tag{7.7}$$

n is the total number of points. The error may seem high (0.15 rad $\approx 8^o$) but it is on the order of the quantization step $\frac{\pi}{16} \approx .2$ rad$\approx 11^o$. This is a strong prove of the co-circularity. A similar experiment with similar result was already reported in [86].

7.5 Probability of co-occurrence of edges

In the previous section we proved that the co-occurrence histograms verify a co-circularity rule. Moreover, from the histograms we were able to extract co-circular curves which are in accordance with the $SE(2)$ cortical inspired sub-Riemannian space introduced in Chapter 4 and with the association fields which is a psychophysical model. It is clear that the histograms encode much richer information that just co-circularity. The objective of this section is to model the full histogram as a probability density function in the space $\mathbb{R}^2 \times S^1$.

We propose modelling the probability of co-occurrence of edges with the same forward-backward Fokker-Planck equation introduced in Chapter 6. The main reason why we conjectured that the same model proposed for completion of existing boundaries can be also proposed for the description of co-occurrence of edges in natural images is that there is the possibility that the geometrical structure of the cortex is partially learned from the structure of the natural images, inheriting the same geometrical properties, and the same natural propagator.

7.5.1 Comparison with the fundamental solution of the Fokker-Planck equation

The goal of this section is to compare the probability of co-occurrence of edges in natural images with the Fokker-Planck fundamental solution introduced in the previous chapter. We recall that the space of position and orientations $\mathbb{R}^2 \times S^1$ in which edges are represented can be identified with the rototranslation group $SE(2)$ with a suitable sub-Riemannian geometrical structure. This space models the functional architecture of the primary visual cortex while the time independent Fokker-Planck equation was proposed as a model of the lateral connectivity. Here, the same forward-backward fundamental solutions of equations (6.19) and (6.20) are proposed as a model of the co-occurrence histogram.

The numerical computation of the fundamental solution was performed using COMSOL multiphysics, a commercial finite element method solver. The implementation details are described in the previous chapter. Shown in Fig. 6.6 top is an iso-surface (the surface defined by a constant value of the 3D function) of a numerically computed fundamental solution with arbitrary parameters. It is clear the relation between the fundamental solution and the integral curves (which projected on the xy-plane are co-circular). The

fundamental solution can be thought as a density distributed around the surface defined by the integral curves, a thicker version of that surface.

We have interpolated the mesh on which is computed the solution in a rectangular grid ($65\times65\times32$ large) in order to compare it with the histogram. The only free parameter of the model is σ, the value of the standard deviation in curvature σ appearing in equation 6.12 on edges belonging to natural images. The best fit between the experimental and simulated distributions has been accomplished by minimizing the mean square error by varying the parameter σ of the equations (6.19) and (6.20). With a discretization step of 0.01 into a range of $0 < \sigma < 10$, the minimum error value result $E_m < 2\%$ showing at a quantitative level that the model accurately represents the experimental distribution. This is in agreement with the Mumford hypothesis on contour reappearance [62]. Moreover, the minimizer corresponding to E_m results $\sigma = 1.73$ pixels^{-1}. This value can be considered as the natural constant of the stochastic process underlying edge distribution in natural images.

Fig.7.17 shows a visualization of the resulting histogram (blue) and the computed fundamental solution of the FP (red). The isosurface corresponding to the 2 % of the maximum has been selected and a surface rendering visualization has been adopted. The qualitative resemblance of the two datasets is evident. As mentioned before, both distributions are thick versions of the surface generated by the set of integral curves shown in Fig.4.8. The torsion of the isosurfaces is visible in both datasets.

Note that the sum of the backwards and forwards fundamental solutions are symmetrical with respect to the origin by construction. This property is also verified by the histogram H, i.e. given a pair of contour points it is sufficient to switch the reference between them to verify the central symmetry.

Another numerical comparison between the two distributions is visualized in Fig.7.18 and Fig.7.19, where isocontour plots at constant θ are shown. Every θ constant slice corresponds to a configuration of two contours differing by θ in their orientations. The drop off in correlation strength with distance present in the figures is strictly related to both the spread of the integral curves of the deterministic model, and the diffusion term defined by the variance σ.

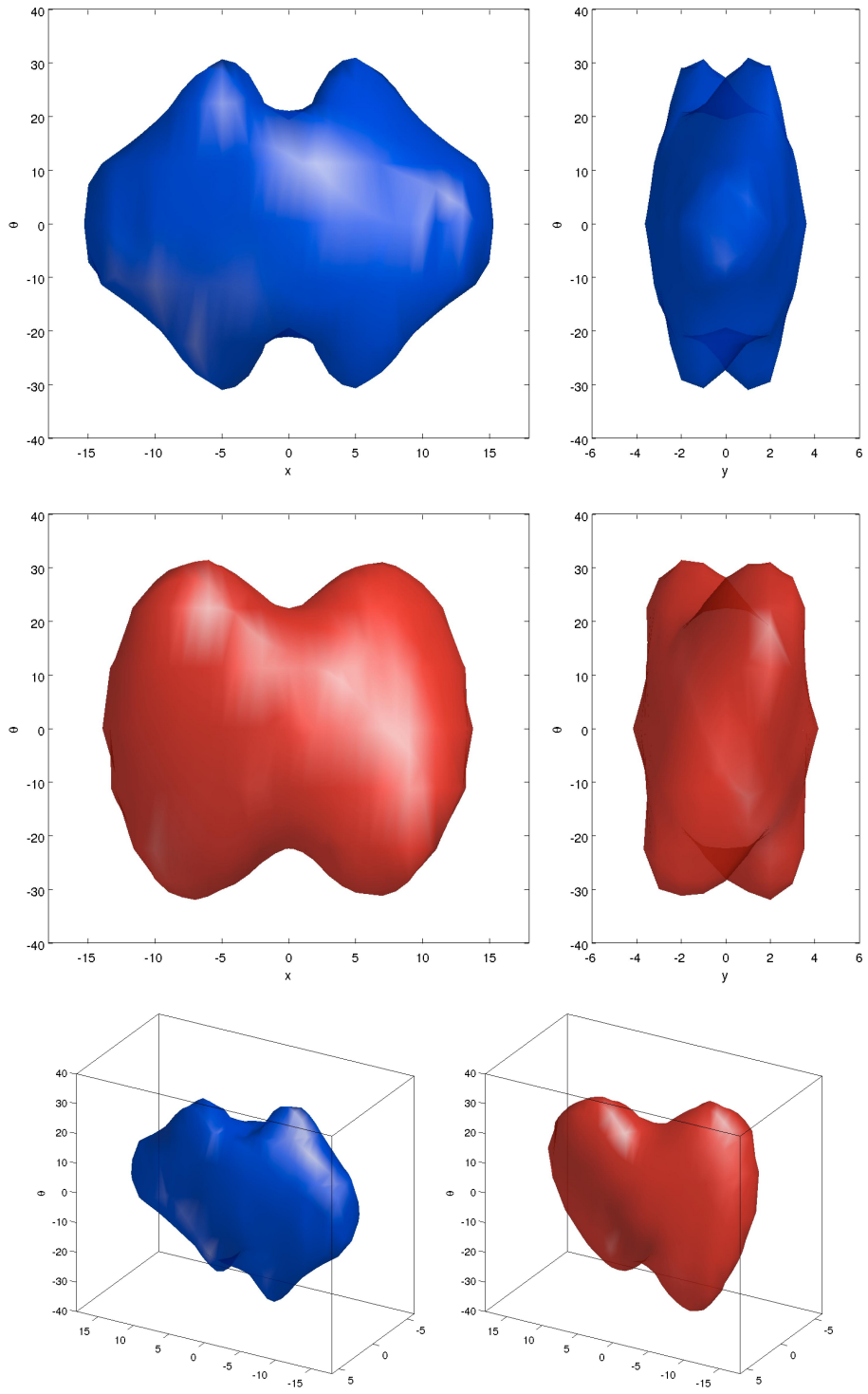


Figure 7.17: Top rows: isosurface visualization of the histogram H (blue) and the FP fundamental solution (red) of the orthogonal projection on the $x\theta$-plane (left) and on the $y\theta$-plane (right). Bottom Row: idem with orthographic projection.

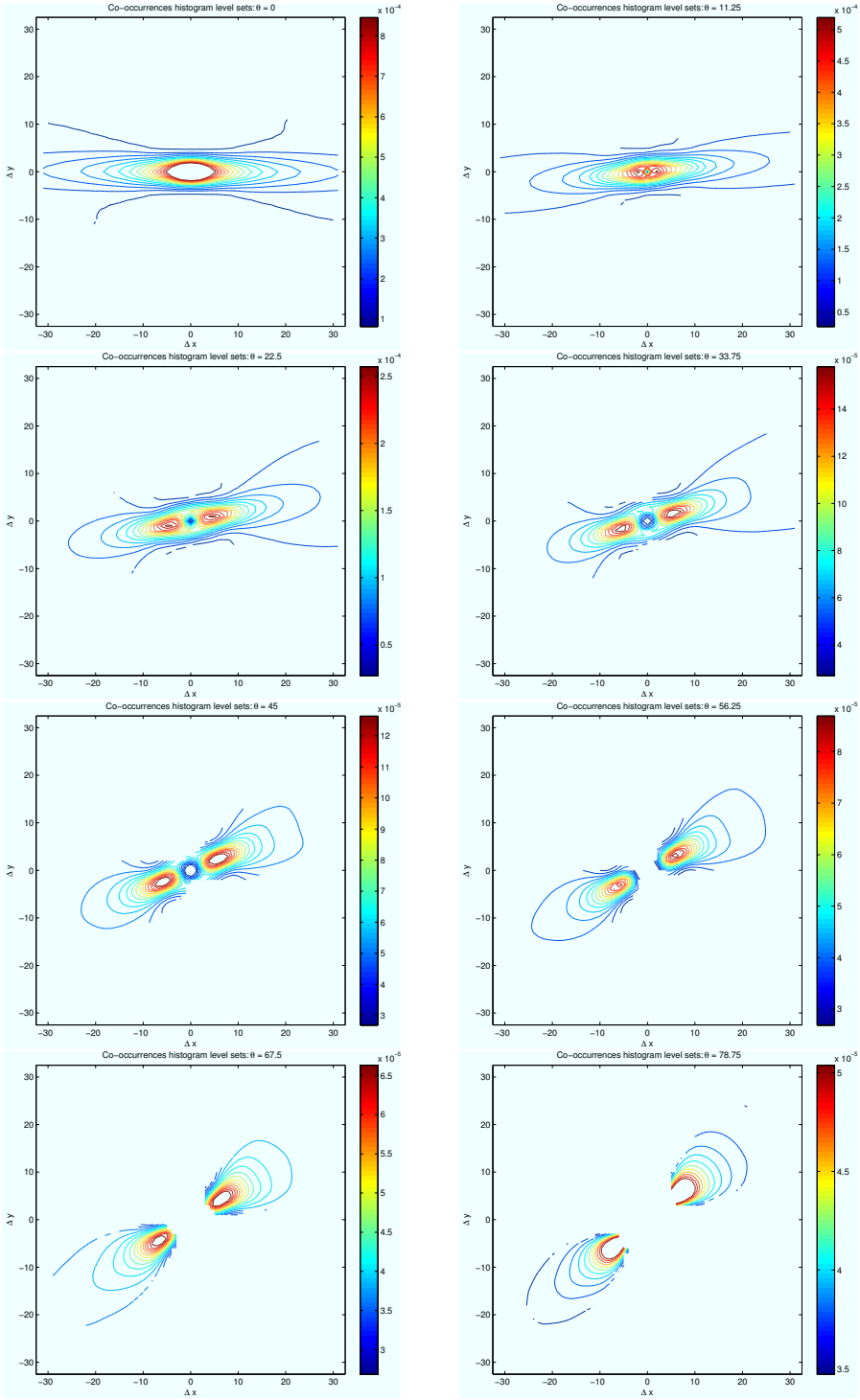


Figure 7.18: Isocontour visualization of the histogram H at different angles θ.

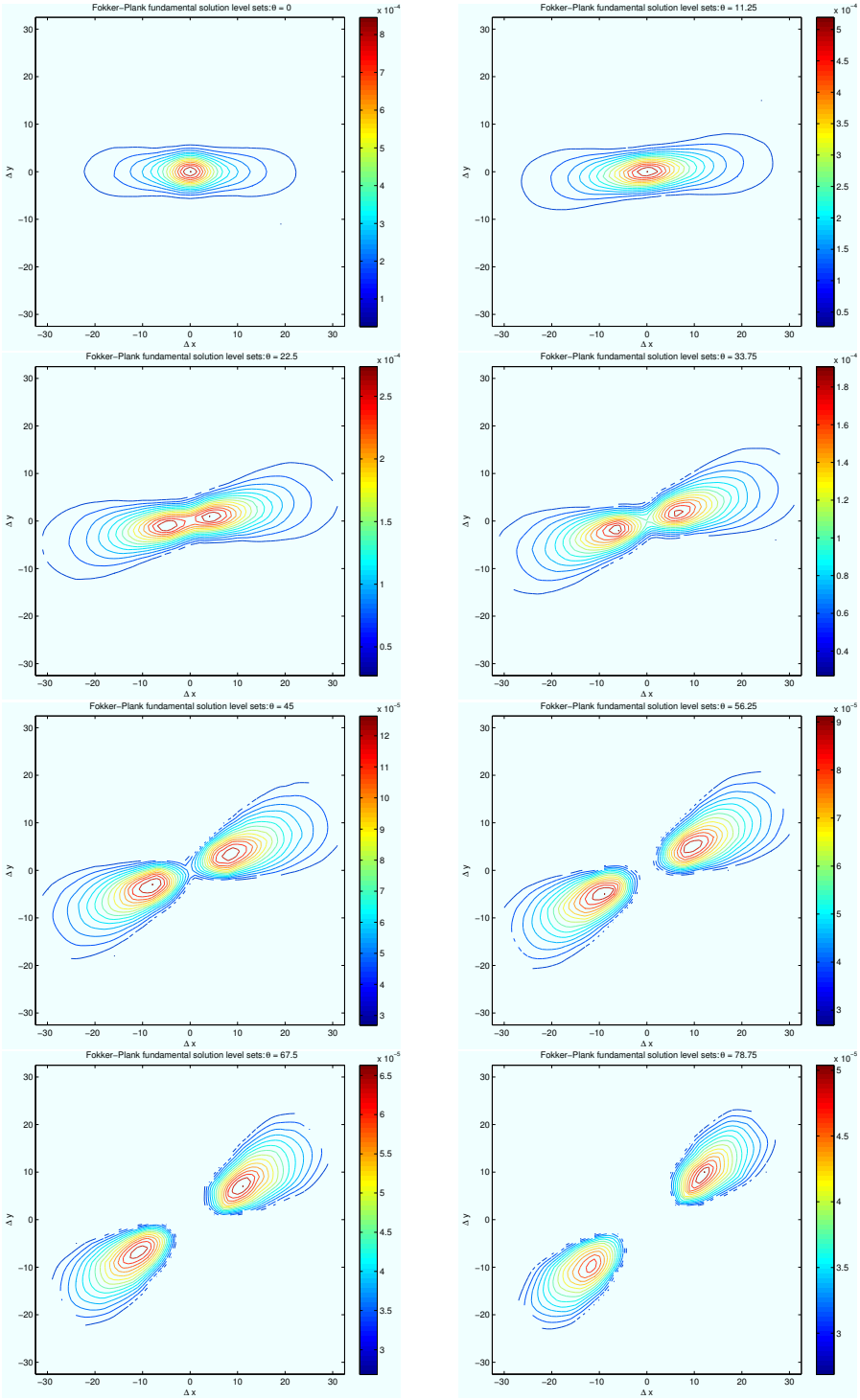


Figure 7.19: Isocontour visualization of the Fokker-Planck fundamental solution at different angles θ.

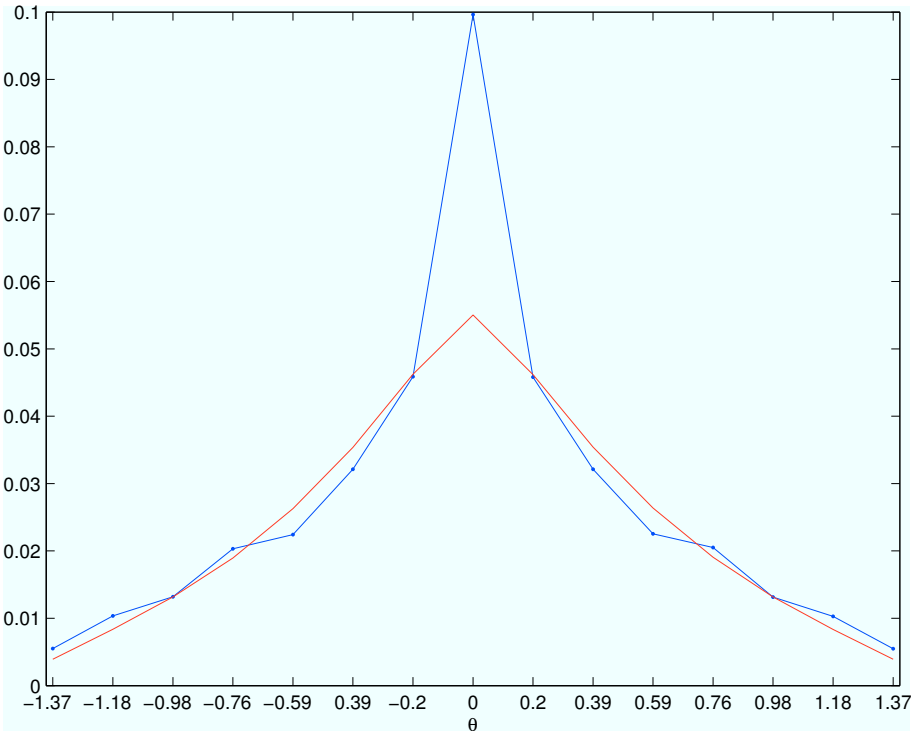


Figure 7.20: Comparison between the number of co-occurrences in every θ constant slice in H (after normalization) and the corresponding value in the Fokker-Planck distribution. The value obtained from the histogram is plotted in blue while that from the Fokker-Planck is in red. There is a big discrepancy between the histogram and the Fokker-Planck solution for $\theta = 0$ which we think is due to the existence of a parallel structure.

Plotted in blue in Fig.7.20 is the amount of co-occurrences of the histogram H for each θ constant slice and in red that of the integral for the Fokker-Planck distribution at a constant θ slice after performing a suitable rescaling which is explained below. The two distributions are almost coincident for all values of the θ angle strictly less than $|\frac{\pi}{2}|$ and different from 0 (collinear configuration).

Over 10% of the co-occurrences of H correspond to collinear edges ($\theta = 0$). This is explained by the fact that in the natural image database there are many sharp straight edges corresponding mostly to buildings. The same argument explains the two peaks present at $\theta = \pm\frac{\pi}{2}$, orthogonal contours.

As outlined by Fig. 6 in the paper of Olshausen and Simoncelli [87] where they modelled the same kind of distribution, the distributions had a sharp peak at 0 and much longer tails than that of a Gaussian density. Field [27] argued that the representation corresponding to these densities, in which

most neurons had small amplitude responses, had an important neural coding property, which he called sparseness. Our invariant model still underestimates the peak at 0, while it fits the tails of low intensities. The underestimation at 0 can probably be explained by a violation of the rotation invariance of the image we postulated here. Indeed the probability density is stronger in the collinear direction. The range of considered values of θ exclude automatically the presence of parallelism effects, which are statistically significant for larger values of theta, as shown in [51] as well as in [32]. The modelling of parallelism in terms of the Lie groups theory has been faced at a purely theoretical level in [84], and will be studied in terms of image statistics in Chapter 9.

7.6 Discussion

To our understanding this chapter contains the main contribution of this thesis. We developed the connection between the rototranslation model for the interaction between oriented neurons in visual cortex and the edge co-occurrence statistics in natural images. Firstly, we checked that the association fields are encoded in the co-occurrence probability distribution. Our pattern (Fig. 7.16) is the proof of a co-circularity hypothesis which was already successfully tested in literature [86]. Then, inspired by Mumford [62], the fundamental solution of the time independent Fokker-Planck operator introduced in Chapter 5 was used to bridge between the geometric model and the edge statistics. To our knowledge, the only work in literature which proposed a relationship between Mumford's direction process and was [4], but the statistics only were used to "tune" a computer vision algorithm for perceptual completion.

A reasonable critic to the methods proposed here is that they are based on the rotation-translation invariance hypothesis. We are applying that hypothesis to compute our histogram and the models used to predict it derive from the same symmetry principle. Nevertheless, the invariance seems reasonable if we visually inspect the 4D histogram (Fig. 7.10). Moreover, if the histograms are computed using white noise as input images (instead of the natural scenes), no structure appears. This result was reported by Krüger in [51].

The approach proposed in this chapter is extended in Chapter 8 to a different Lie Group, the affine group, so it can take into account the scale, another low level visual features. We expect in future research to extend the method to consider other low level features as curvature, stereotopic vision, etc.

Chapter 8

Scale

8.1 Introduction

This chapter presents an extension of the previous one where the feature of scale is added to the feature of boundary orientation. This leads us to consider the affine group (the Lie group invariant under rotations, translations and scaling transformations). The extension of the cortical model was provided by Sarti, Citti and Petitot in [84] where a symplectic structure was introduced. The scale takes the role of the distance from a boundary. Then, it becomes possible to take into account the interior of the objects. Using considerations similar to that in [14], two sub-Riemannian structures are used as models for the long range connectivity. One is inherited from the $SE(2)$ model which models the co-axial connectivity. The second one is responsible for the trans-axial connectivity and is morphologically different from the previous one. Then, we develop a new method for computing natural image statistics so the scale feature may be taken into account. This leads to the consideration of a 6D histogram. This chapter concentrates in the trans-axial connectivity since the coaxial connectivity was already studied in the previous chapter. Analyzing the histogram it is clear the presence of a pattern coherent with the model. Then, in analogy with the methodology used in Chapter 7, a differential operator for modelling the trans-axial connectivity is introduced.

8.2 The symplectic structure

8.2.1 Relation with the $SE(2)$ model

In Chapter 4 (following [14]) the hypercolumnar structure of the visual cortex is modelled as a fiber bundle. The 2-dimensional retinal plane (x, y) is the base manifold M of the fiber bundle, and the engrafted variable is the orientation θ of a simple cell with an odd-symmetric receptive profile. The vertical fibers over each point emulate the hypercolumnar architecture so that the cortical space is identified with the total space of the fiber bundle isomorphic to $\mathbb{R}^2 \times S^1$. Taking into consideration the appropriate group law, the fibration is identified with the group $SE(2)$. A visual stimulus represented in the 3D space is interpreted geometrically as a lifting where the third coordinate is the local orientation at the point. The connectivity in the group is given by the directions of the left-invariant vector fields $X_1 = (\cos(\theta), \sin(\theta), 0)$ and $X_2 = (0, 0, 1)$. A natural sub-Riemannian structure is defined by introducing a metric on the plane spanned by X_1, X_2 which are non-commutative. The projection on the xy-plane of the integral curves of the vector fields X_1 and X_2 were interpreted as a model of the association fields [29]. Along these curves a parallel transport operation was performed.

In [84], the authors extended the previous work to a 4D geometrical model by considering the organization, connectivity and functionality of even symmetric simple cells. The cortex is now a fiber bundle were orientation and scale are the engrafted variables, so that its total space is $\mathbb{R}^2 \times S^1 \times \mathbb{R}$. The space is identified with the affine group in the plane (the group invariant to rotations, translations and scale transformations) with a symplectic structure. The connectivity is given by the differential of the one form $\omega = e^{-\sigma}(-\sin\theta dx + \cos\theta dy)$. In this case two sub-Riemannian structures are considered whose integral curves are responsible, respectively, for trans-axial and coaxial connections.

8.2.2 The receptive profiles and the hypercolumnar architecture

The retinal plane (or equivalently the visual plane) is identified with the plane $\mathbb{R}^2$ with coordinates (x, y). The scale dimension is introduced taking into consideration simple cells with even-symmetric receptive profiles instead of the odd receptive profiles. Let φ_0 be an even simple cell modelled as the real part of a Gabor wavelet in local coordinates (ξ, η):

$$\varphi_0(\xi, \eta) = e^{-(\xi^2+\eta^2)} \cos(2\eta). \tag{8.1}$$

The premise is the same as that in the previous chapters, for each point (x, y) in the retinal plane there are cells with receptive profiles in the cortex with all possible orientation θ and scales σ. Then, a generic receptive profile of a simple cell is obtained by centering the mother kernel φ_0 at the retinal point (x, y), rotating around its own axis by an angle θ and scaling it by a value proportional to σ. This is achieved by the following change of variables:

$$\begin{cases} \xi = e^{-\sigma}\left((\bar{\xi} - x)\cos\theta + (\bar{\eta} - y)\sin\theta\right) \\ \eta = e^{-\sigma}\left(-(\bar{\xi} - x)\sin\theta + (\bar{\eta} - y)\cos\theta\right) \end{cases}. \tag{8.2}$$

The generic receptive profile now writes as:

$$\varphi_{x,y,\theta,\sigma}(\bar{\xi}, \bar{\eta}) = e^{-(\xi^2+\eta^2)}\cos(2\eta). \tag{8.3}$$

Each cell in the cortex can be represented by the coordinates $(x, y, \theta, \sigma) \in \mathbb{R}^2 \times S^1 \times \mathbb{R}$ which are the parameters of the receptive profiles. Each receptive profile was obtained by translation, rotation and scaling of a mother kernel. Precisely, we call

$$A_{x,y,\theta,\sigma}(\xi, \eta) = \begin{pmatrix} x \\ y \end{pmatrix} + e^{\sigma} \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix}.$$

Then each receptive profile can be expressed as

$$\varphi_{x,y,\theta,\sigma}(\bar{\xi}, \bar{\eta}) = \varphi_{0,0,0,0}A^{-1}(\bar{\xi}, \bar{\eta}).$$

The set of transformations A with the standard composition law form the affine group. As in Chapter 4 the space of parameters $\mathbb{R}^2 \times S^1 \times \mathbb{R}$ can be identified with the affine group.

8.2.3 The lifting into the fiber bundle

The overall output of the ideally modelled cortex with a visual stimulus $I(x, y)$ (a function of activation on the retinal layer) is given by the integration of the stimulus weighted by the receptive profiles, plus a maximum selectivity mechanism. Let $O_{\theta,\sigma}$ be the result of the filtering:

$$O_{\theta,\sigma}(x, y) = \int I(\bar{\xi}, \bar{\eta})\varphi_{x,y,\theta,\sigma}(\bar{\xi}, \bar{\eta})d\bar{\xi}d\bar{\eta}, \tag{8.4}$$

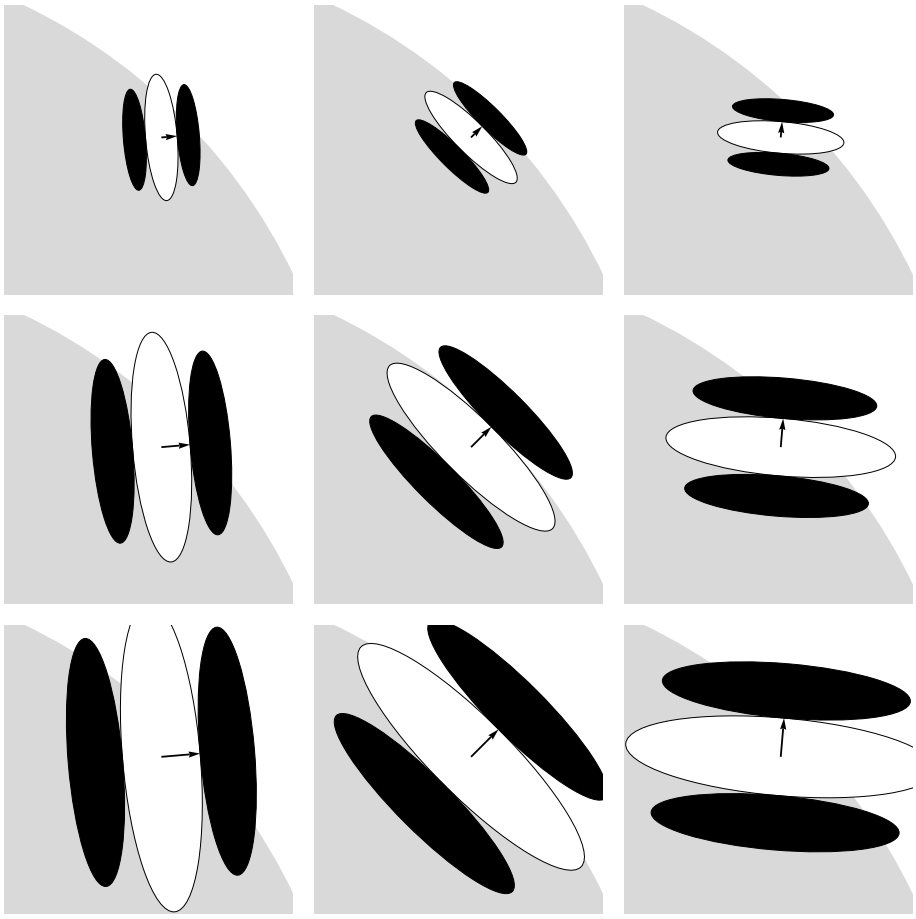


Figure 8.1: Simple cells act on the image detecting the tangent vector of optimal orientation (columns) and scale (rows). Each filter can be described by the 1-form ω_3 which is represented as a black arrow within the images.

Then, the cortex associates to each point (x, y), a couple (θ_m, σ_m) where the function O is maximal.

The map from the retinal point (x, y) to the cortical position $(x, y, \theta_m, \sigma_m)$ is interpreted as a lifting of the input stimulus into the fiber bundle $\mathbb{R}^2 \times S^1 \times \mathbb{R}$.

The geometrical interpretation of the lifting was analyzed on cartoon images and generalizes the $SE(2)$ model. If $\sigma = 0$ the model is reduced to that of the previous chapters as the distance to the boundary is 0, and θ_m expresses the orientation of the boundary. In [84], it was proved that the output function $O_{\theta,\sigma}(x, y)$ reaches a local maximum at the point (θ_m, σ_m), where $d(\sigma_m) = \frac{1}{\sqrt{2}} e^{\sigma_m}$ denotes the distance of (x, y) from the nearest boundary of the image I, and θ_m denotes the orientation of this boundary at the point

where the distance is achieved (see Fig. 8.2).

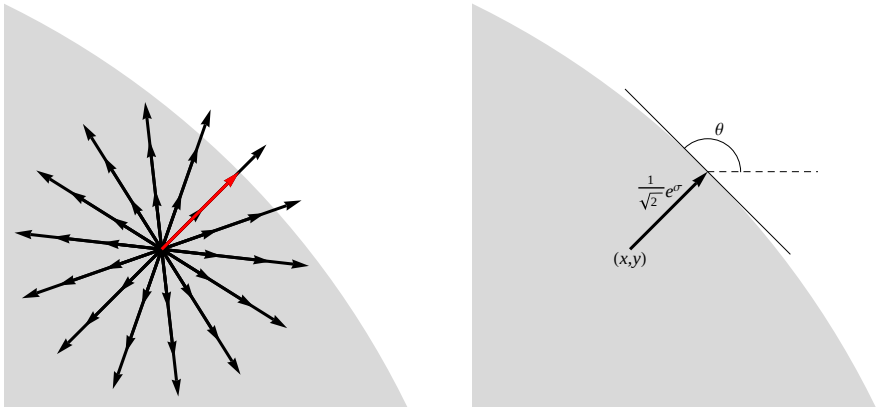


Figure 8.2: The simple cell centered in (x, y) takes the maximal activity $O_{\theta_m,\sigma_m}(x, y) = \max_{(\theta,\sigma)} O_{\theta,\sigma}(x, y)$ at a point where σ_m is, up to a constant, the logarithm of the distance to the nearest boundary and θ_m is the direction of this boundary

8.2.4 Left invariant vector fields

The Lie algebra associated to the affine group is spanned by the following left invariant vector fields:

$$\begin{aligned} X_{\sigma,1} &= e^{\sigma}(\cos\theta\partial_x + \sin\theta\partial_y) \\ X_{\sigma,2} &= \partial_\theta \\ X_{\sigma,3} &= e^{\sigma}(-\sin\theta\partial_x + \cos\theta\partial_y) \\ X_{\sigma,4} &= \partial_\sigma \end{aligned}$$

Let us explicitly note that the vectors $X_{\sigma,1}$, $X_{\sigma,2}$, $X_{\sigma,3}$, have the same direction as the corresponding vector fields X_1, X_2, X_3, introduced in the previous chapters. The two vectors $X_{\sigma,1}$, $X_{\sigma,3}$ have been modified in order to take into account the scale.

We will now introduce to this Lie algebra a symplectic structure, which is an extension of the sub-Riemannian structure. The two vector fields $X_{\sigma,1}$, $X_{\sigma,2}$, images of the generators of the horizontal plane are coupled within this structure, and we can find a linear transformation on the horizontal plane, a matrix J such that

$$X_{\sigma,1} = JX_{\sigma,2}, \quad X_{\sigma,2} = -JX_{\sigma,1}.$$

The vector field X_3, uncoupled in the subriemannian structure, is here lifted to the vector field $X_{\sigma,3}$, and coupled with the vector $X_{\sigma,4}$. The coupling is realized as

$$X_{\sigma,4} = JX_{\sigma,3}, \quad X_{\sigma,3} = -JX_{\sigma,4}.$$

The matrix J is a rotation in the 4D space, and plays the same role as the multiplication by i in the complex space. This generalization of the complex structure is called the symplectic structure. It is then characterized by the presence of two different contact structures, with independent sub-Riemannian metrics. And the horizontal planes are respectively that spanned by $\{X_{\sigma,1}X_{\sigma,2}\}$ and by $\{X_{\sigma,3}X_{\sigma,4}\}$.

The connectivity in this case (formally named symplectic structure) is given by means of the 2-form $d\omega_3$ obtained by differentiating the 1-form ω_3 with respect to all its variables:

8.2.5 Integral Curves

If we consider the first part of the symplectic form (the plane spanned by $\{X_{\sigma,1}, X_{\sigma,2}\}$), we reduce ourselves to the previous model. The integral curves defined as:

$$\begin{cases} \dot{\gamma}(t) = X_{\sigma,1}(\gamma(t)) + kX_{\sigma,2}(\gamma(t)) \\ \gamma(0) = (x_0, y_0, \theta_0, \sigma_0) \end{cases} \tag{8.5}$$

model the trans-axial connections. The explicit parametric solution is

$$\begin{cases} x(t) = \frac{1}{k}(\sin(kt + \theta_0) - \sin(\theta_0)) + x_0 \\ y(t) = -\frac{1}{k}(\cos(kt + \theta_0) - \cos(\theta_0)) + y_0 \\ \theta(t) = kt + \theta_0 \\ \sigma(t) = \sigma_0 \end{cases}. \tag{8.6}$$

For $k = 0$ its projection on the xy-plane is the x-axis and for $k \neq 0$ it is a circle of radius $\frac{1}{k}$ tangent to the x-axis.

In the same way, we can consider the integral curves of the vector fields $\{X_{\sigma,3}, X_{\sigma,4}\}$,

$$\begin{cases} \dot{\gamma}(t) = X_{\sigma,3}(\gamma(t)) + kX_{\sigma,4}(\gamma(t)) \\ \gamma(0) = (x_0, y_0, \theta_0, \sigma_0) \end{cases}. \tag{8.7}$$

The solution is

$$\begin{cases} x(t) = -\frac{\sin(\theta_0)}{k} e^{\sigma_0}(e^{kt} - 1) + x_0 \\ y(t) = \frac{\cos(\theta_0)}{k} e^{\sigma_0}(e^{kt} - 1) + y_0 \\ \theta(t) = \theta_0 \\ \sigma(t) = kt + \sigma_0 \end{cases}. \tag{8.8}$$

Its projection on the xy-plane is independent of k and orthogonal to the direction θ_0. For the k variable the integral curve is the line of slope k in the fixed *vertical* plane spanned by $\{X_{\sigma,3}, X_{\sigma,4}\}$.

The projection of these two classes of integral curves on the xy-plane is shown in Fig. 8.3. These patterns were intended in [14] as a model of long range excitatory connections. Two different regions may be distinguished. The one plotted in blue along the direction of the axis of orientation (co-axial) which spreads out the model of the association fields. This structure was already modelled in the sub-Riemannian $SE(2)$ model.

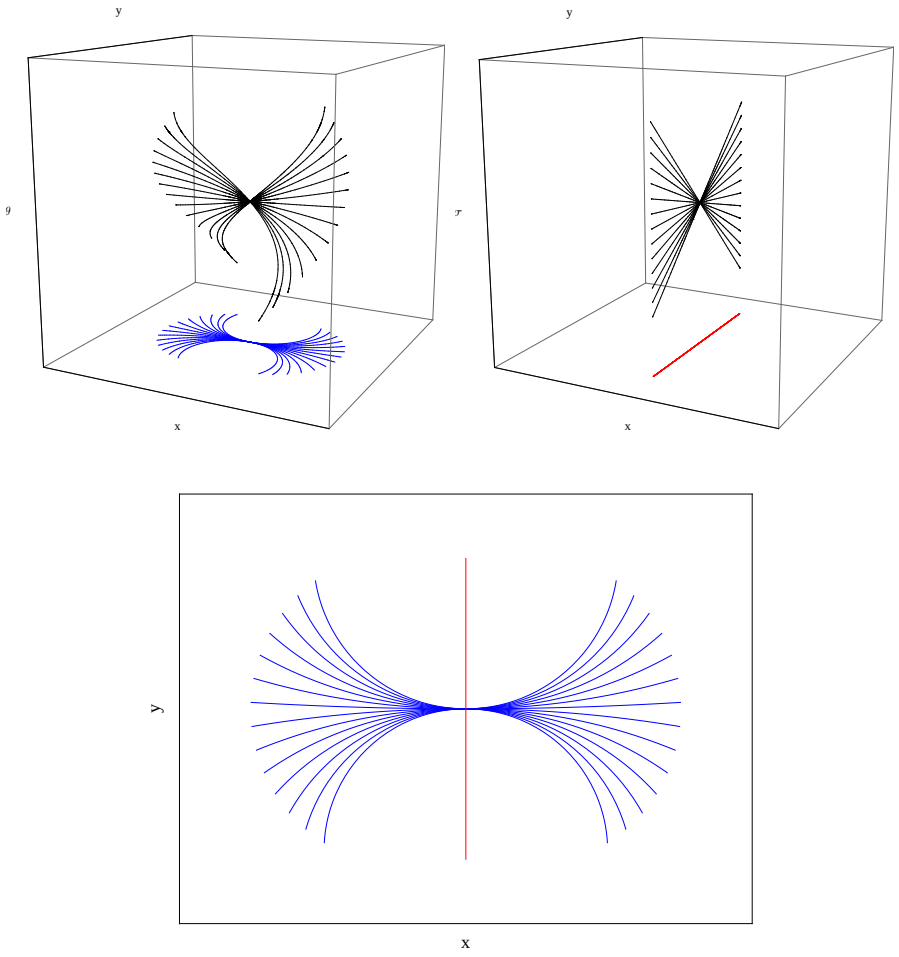


Figure 8.3: Visualization of the integral curves of the affine group with the symplectic structure. Shown are the fields $X_{\sigma,1}$, $X_{\sigma,2}$ (upper-left) and $X_{\sigma,3}$, $X_{\sigma,4}$ (upper-right), and their projections on the (x, y) plane (bottom).

The symplectic structure considered in this chapter adds a second set of

connections which are confined to a narrow zone extending orthogonally to the axis (trans-axial, plotted in red). The trans-axial connections are represented here as integral curves of the vector fields $X_{\sigma,3}$, $X_{\sigma,4}$. The reason why co-axial connections spread out in a fan while trans-axial connection are more spatially focused is that the field $X_{\sigma,1}$ and $X_{\sigma,2}$ do not commute and their commutator is linearly independent from them:

$$[X_{\sigma,1}, X_{\sigma,2}] = -X_{\sigma,3},$$

while the vectors $X_{\sigma,3}$, $X_{\sigma,4}$ do not commute but their commutator linearly depends on them:

$$[X_{\sigma,3}, X_{\sigma,4}] = -X_{\sigma,3},$$

In Fig. 8.3 it can be seen that the integral curves of $X_{\sigma,1}$ and $X_{\sigma,1}$ are not planar (top-left) while integral curves of the $X_{\sigma,3}$, $X_{\sigma,4}$ belong to the plane spanned by themselves (top-right) and their projection is the red line only.

8.3 Natural image statistics

This section is devoted to the computation of the distribution of orientation and scale in natural images lifted in the 4D space introduced in the previous section.

8.3.1 Methodology

Each image I in the data base is lifted to a 4D space by means of a filter bank of even symmetric kernels. The filter bank is indexed by the parameters scale σ and orientation θ. The lifting maps each point (x, y) in the image to a point (x, y, θ, σ) in $\mathbb{R}^2 \times S^1 \times \mathbb{R}$. The same procedure applied in Chapter 7 is applied here for the computation of the co-occurrence histogram, so we refer to Section 7.3 for a detailed explanation. The natural image database is the same that was used in the previous chapter. The only methodological difference is that we consider even symmetric filters indexed by scale and orientation. More precisely we used second derivative of Gaussian filters $\varphi_{\sigma,\theta}(x, y)$:

$$\varphi_{\sigma,\theta}(x, y) = \frac{\partial^2 G_\sigma}{\partial \tilde{y}^2}(x, y), \tag{8.9}$$

where:

- G_σ is the 2D gaussian function:

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

- $\tilde{y} = -x\sin\theta + y\cos\theta$ so the orientation of the filter is θ;
- σ is the scale parameter.

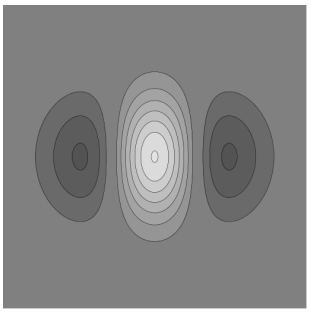


Figure 8.4: Shown in the figure are some level sets of a second directional derivative of gaussian kernel $\varphi_{\sigma,\theta}(x,y)$ (eq. 8.9, $\theta = \frac{\pi}{2}$). This is the mother kernel from which is obtained the filter bank used in the construction of the 6D histogram.

Depicted in Fig. 8.4 is the kernel $\varphi_{\sigma,\theta}(x,y)$, for an arbitrary scale σ and orientation $\theta = \frac{\pi}{2}$. Each kernel used for the detection is obtained by a scaling and a rotation of $\varphi_{\sigma,\theta}$. Note that $\varphi_{\sigma,\theta} = \varphi_{\sigma,\theta+\pi}$, as the function is even symmetric. Then, the angle θ takes values in the interval $[0,\pi)$ and not in the whole circle.

There are 8 different orientations and 10 different scales considered here. The sampling of the scale space is done in such a way that the scale grows one pixel in each sample from 3 to 12 pixels. This means that the scale σ takes values from the set $\{3px, 4px, \ldots, 11px, 12px\}$. The orientations of the filters are equidistributed in the interval from 0 to π: $\theta \in \{0, \frac{\pi}{8}, \ldots \frac{7\pi}{8}\}$. The filters were implemented with the steerable architecture [30] as in the previous chapter. The threshold to the output of the filters was set empirically. We considered at most a 16 pixels distance for the co-occurrence of edges. Each detected point is represented by 4 magnitudes (x, y, θ, σ), the first 2 are the position in pixels in the image, the third is its orientation and the last is its scale (which in this framework is interpreted as the distance in pixels to a boundary).

The co-occurrence histogram is 6D (after imposing translation invariance in the variables (x,y)). The coordinates are the relative position, as well as

the orientation and scale of each point: $(\Delta x, \Delta y, \theta_c, \theta_p, \sigma_c, \sigma_p)$. With these parameters, the size of the 6d histogram is $33 \times 33 \times 8 \times 8 \times 10 \times 10$.

All the computations were implemented in C language. In the following section we provide visualization of the histogram.

8.3.2 Results

Visualizing 6D data is at least a challenging problem. Probably, the easiest way is to fix some of the parameters and visualize the resulting histogram. Fig. 8.5 shows the first attempt. In there is visualized the 5D histogram resulting from setting $\sigma_c = 3px$, its lowest possible value. Shown in the figure is a stack of planes. Each plane corresponds to a different value of σ_p, then, in each plane is visualized a 4D histogram corresponding to fixed values of σ_c and σ_p. For space restrictions, only 6 values of σ_p are plotted. The value of σ_p increases from the bottom to the top. The slices with constant σ_c and σ_p are 4D histograms which are plotted as a grid of matrices, using the same technique used in Fig. 7.10. In the grid the orientation of one of the edges varies along columns, while the orientation of the other along rows. Each one of the small panels inside the 4D histogram corresponds to a slice of the 6D histogram with the four last parameters θ_c, θ_p, σ_c and σ_p fixed. It is hard to extract conclusions from this image but it gives an intuition about the 6 parameters.

The stack of Fig. 8.5 is presented in Figure 8.6 with a different visualization. Each panel correspond to a different value of σ_c and σ_p indicated on the top. Within each of them, a 4D histogram is plot using the same visualization technique as in Fig. 7.10 in the previous chapter.

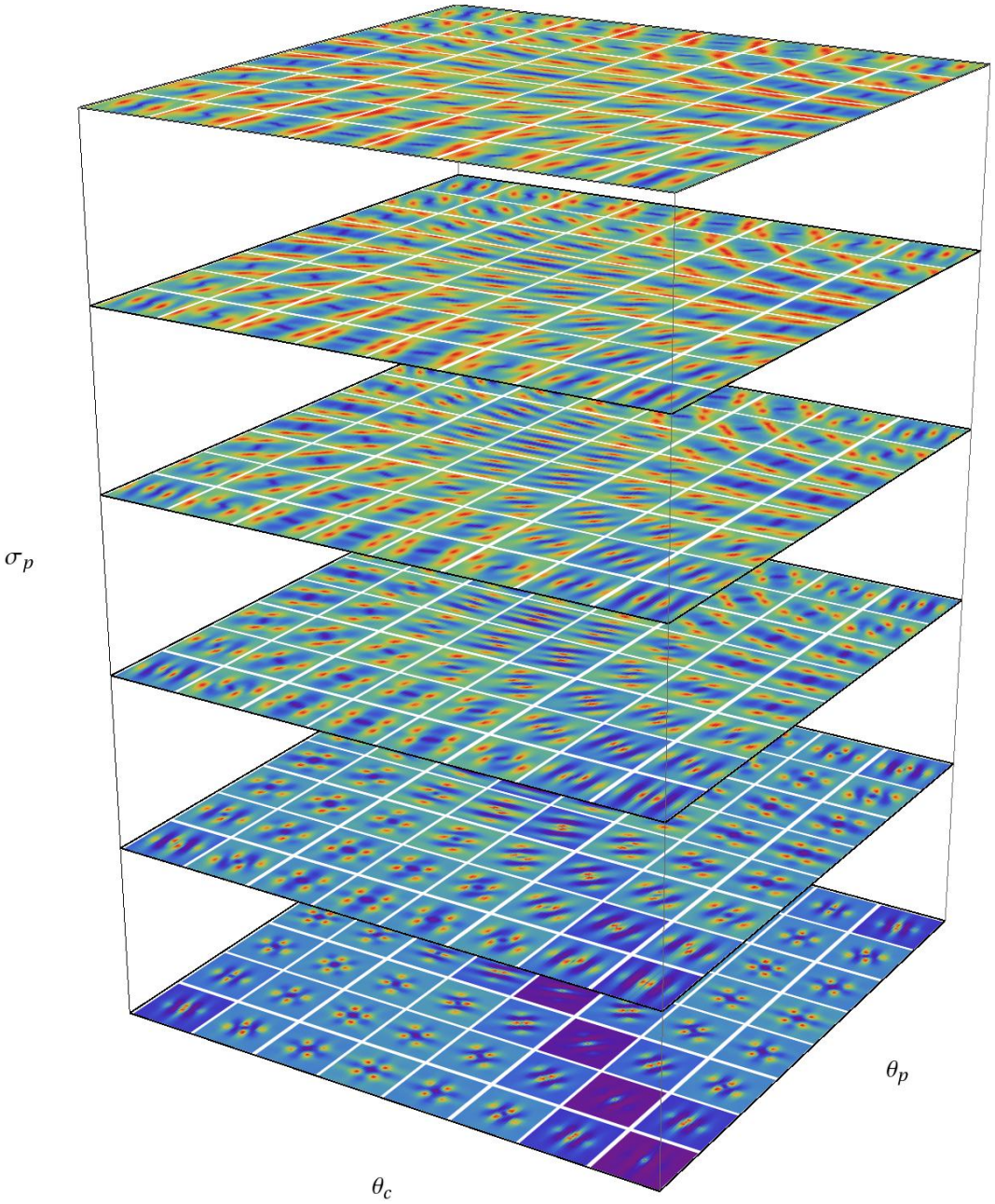


Figure 8.5: Visualization of the 5D histogram of co-occurrences for the case in which σ_c (the scale of the reference edge) is constant and equal to 3 px. In the stack of images, each plane correspond to a different value of σ_p. Then, each plane contains a slice of the histogram for which σ_c and σ_p are constant (a 4d histogram). For space restrictions, only the histograms corresponding to the lowest 6 values of σ are plotted. In Fig. 8.6, each plane of the stack is visualized individually.

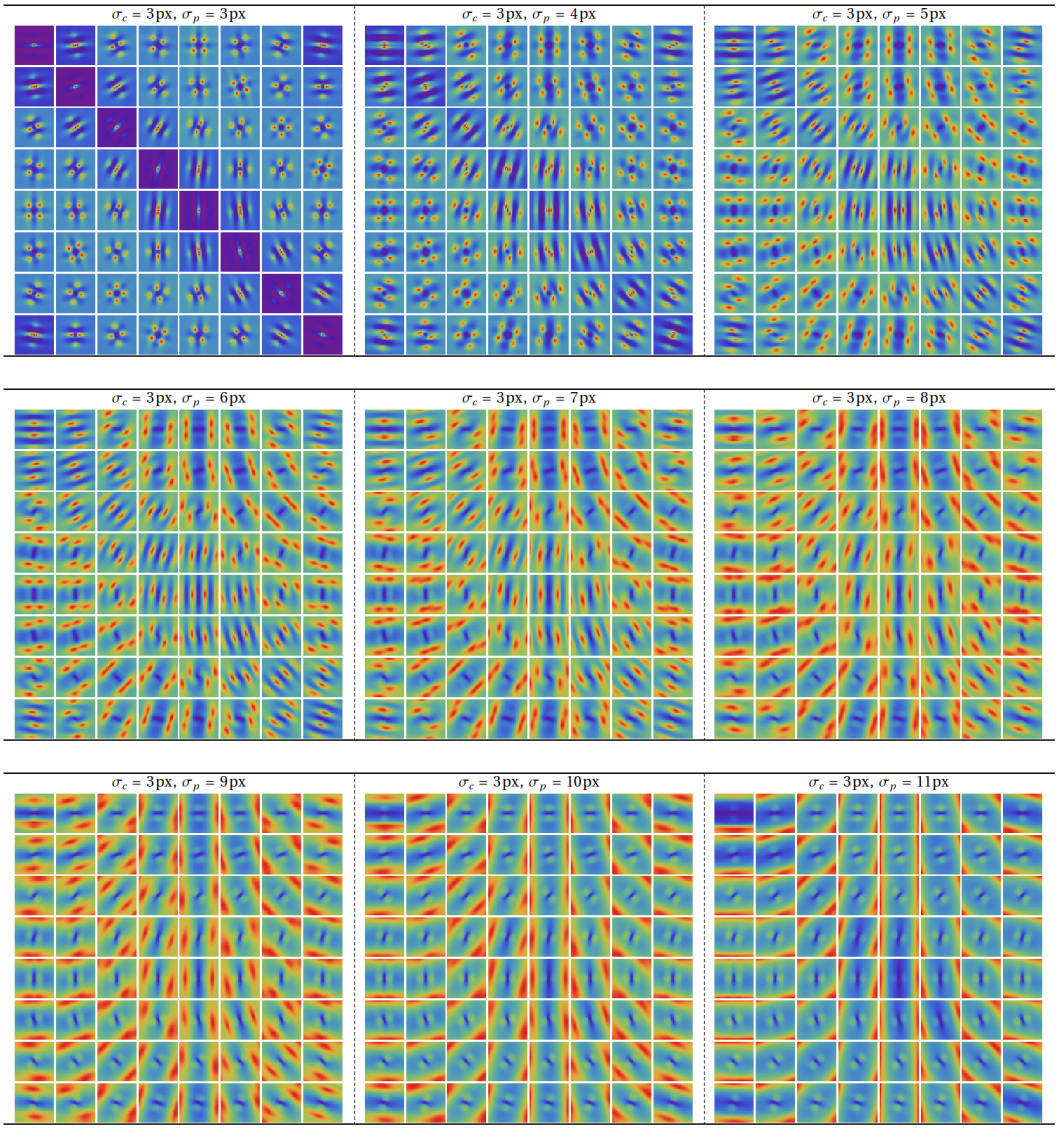


Figure 8.6: Visualization of the histogram of co-occurrences. Visualized in each of the nine large panels is a section of the histogram where σ_c and σ_p are constant. Across the large panels, the value of σ_c remains constant at its lowest possible value. The value of σ_p increases first by rows and then by columns.

8.3.3 Interpretation

Let us first consider a slice of the histogram, for sigma fixed. Some of the characteristic of this histogram are indeed similar to that one. Let us analyze for example the one corresponding to $\sigma_c = \sigma_p = 3px$, shown on the left-top corner of Fig. 8.6. These data correspond to the minimum scale, for both the reference and the secondary edge. Theoretically we expect it to be similar to the one visualized in Fig. 7.10. Indeed the integral curves of $X_{1,\sigma} + kX_{2,\sigma}$ (visualized on Fig. 8.3) for σ fixed, are the same as those of the rototranslation model (see Fig. 4.7). Qualitatively, it can be observed that the histograms are similar. The main difference is that the new one is π-periodic, while that of the previous chapter was 2π-periodic. As a result, instead of 2 branches there are 4, as in the data reported by [86] which we reproduce in Fig. 7.3.

Another characteristic observed in each of the 4D panels is invariance by rotation. Within a panel one can empirically verify that each column is approximatively equal to the next one up to a rotation and a shift on the small panels. The same happens within the columns.

We can now focus on the behavior on the plane $X_{\sigma,3}$ $X_{\sigma,4}$. The vector $X_{\sigma,3}$ is perpendicular to the contours, while $X_{\sigma,4}$ is the direction of the scale axis. For simplicity, in this analysis we set $\theta_c = \theta_p = 0°$ and $\sigma_c = 3px$. Fig 8.7 illustrates this intersection and the actual plane is shown in Fig. 8.8.

One can observe the presence of two branch-like structures. The principal (the strong one), is form by two straight branches starting from the origin. This structure is coherent with the connectivity predicted by the symplectic model, as it is along one of the integral curves of $X_{\sigma,3} + kX_{\sigma,4}$. Intuitively, the reason for the existence of this structure is well explained by the central column of drawings in Fig. 8.1. If a kernel responds maximally to one edge at a certain scale, the same edge is going to be detected by filters parallel to the first one and centered along the direction perpendicular to the axis of the filter. There is also a weaker structure at a lower angles which does not come out from the origin. This structure is not predicted by the model and we think it may be due to the violation of the constant piecewise hypothesis which is implicit in our interpretation of the scale feature.

Summarizing, we conjecture that the stronger branch-type structure is due to the distance to the boundary in piecewise constant images, while the second structure could be due to the presence of lines and textures in natural images. In Section 8.3.4 a simple experiment to check this last conjecture is proposed. Taking into account that in the model we considered only piecewise constant images (cartoon), the model is only able to predict the first structure.

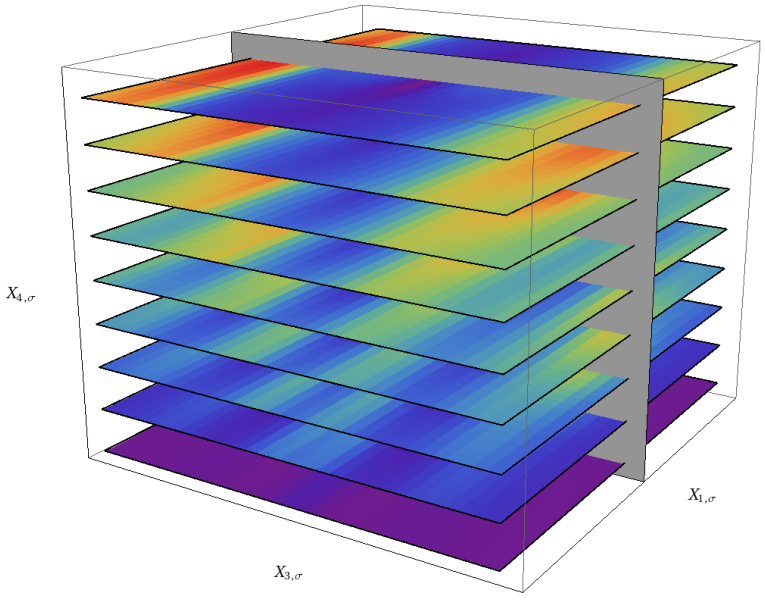


Figure 8.7: The intersection of the histogram with the plane $\{X_{\sigma,3}, X_{\sigma,4}\}$ (plotted in gray). For simplicity, we set $\theta_c = \theta_p = 0°$ and $\sigma_c = 3px$. Then, each plane of the stack is that on the left-top corner of each big panel in Fig 8.6.

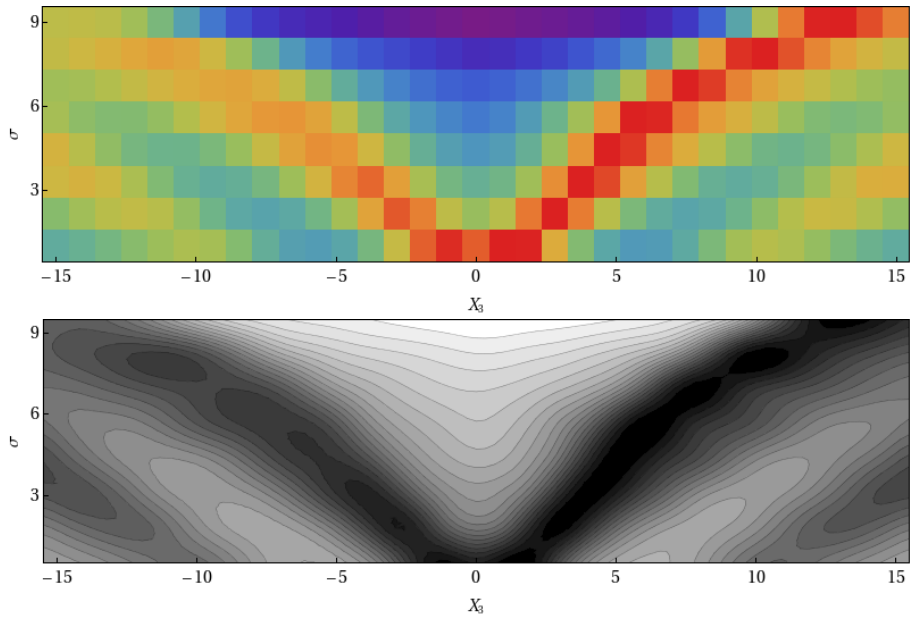


Figure 8.8: The intersection of the histogram with the plane $\{X_{\sigma,3}, X_{\sigma,4}\}$. On the upper figure the color is proportional to the probability while at the bottom some level curves are visualized.

We can model the main structure with a propagation in the direction of the integral curve of the structure on the $X_{\sigma,3}$ $X_{\sigma,4}$ plane. The corresponding equation is a pure advection of the two directions $X_{\sigma,3} - X_{\sigma,4}$ and $X_{\sigma,3} + X_{\sigma,4}$ which leads to the wave operator (the equation comes trivially because of the conjugate product):

$$(X_{\sigma,33})^2 - (X_{\sigma,44})^2 = 0. \tag{8.10}$$

where $X_{\sigma,33} = X_{\sigma,3}(X_{\sigma,3})$ and $X_{\sigma,44} = X_{\sigma,4}(X_{\sigma,4})$.

The solution to (8.10) was simulated using COMSOL and visualized in Fig. 8.9.

Note we did not have any spreading in the histogram, as predicted by the model, due to the linear dependence of the commutator $[X_{\sigma,3}, X_{\sigma,4}]$ from the vectors $X_{\sigma,3}, X_{\sigma,4}$. When modelling the trans-axial connectivity, (8.10) is the analogue of the Fokker-Planck operator we used for modelling the co-axial connectivity in the past chapters.

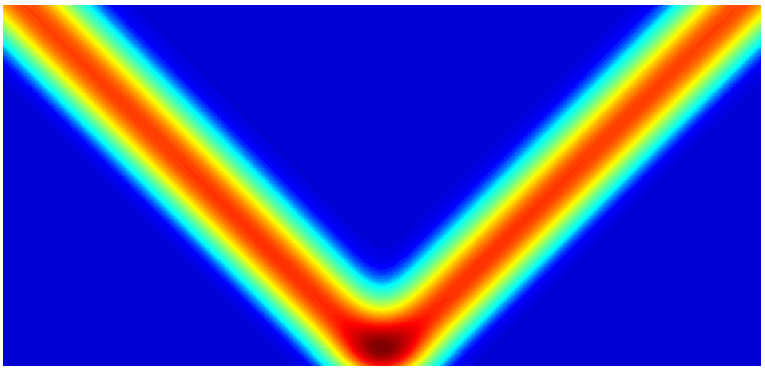


Figure 8.9: The solution of eq. (8.10) which models the structure at 45° shown in Fig. 8.8. This density was computed using COMSOL.

8.3.4 Cartoon Database

In order to support our conjecture we tested the methodology on a database of randomly generated cartoon images. The scope of this simple experiment is to prove the absence of the secondary structure seen in the previous histogram. In this way, we verified that this structure was due to the presence of lines, textures or light gradients in the natural images. These artifacts were not considered in the symplectic model. Nevertheless, one may argue that the secondary structure may be due to some other reason, for example the bank of filters that was selected. If this other conjecture is true, then the same

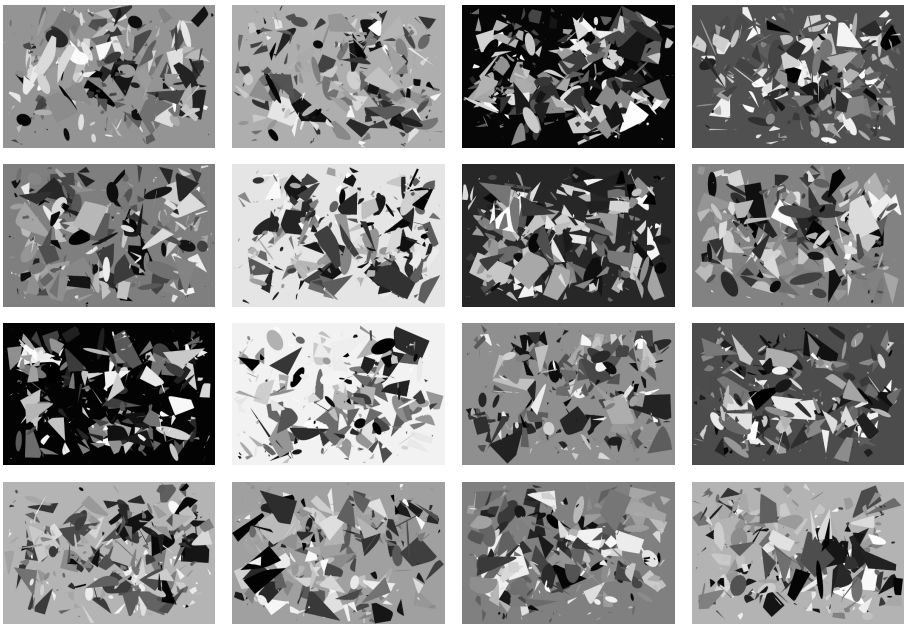


Figure 8.10: Examples of the randomly generated database of cartoon images. The images are large 1536×1024

structure should appear when the histograms are computed on the piece-wise constant database.

Some examples of synthetic images are presented in Fig. 8.10.

The parameters used in the computation of the co-occurrence histograms are the same as in the previous section. The histograms were computed using 100 randomly generated images of size 1536px×1024px. A part of the resulting histogram is visualized in Fig. 8.11, while the plane $\{X_{\sigma,3}, X_{\sigma,4}\}$ is depicted in Fig. 8.12. Analyzing this last figure, one can check that the additional structure is no longer present, proving our conjecture.

This experiment proves that the structure of the histogram strongly depends on the set of images, and it is not dependent on the specific choice of the type of filters.

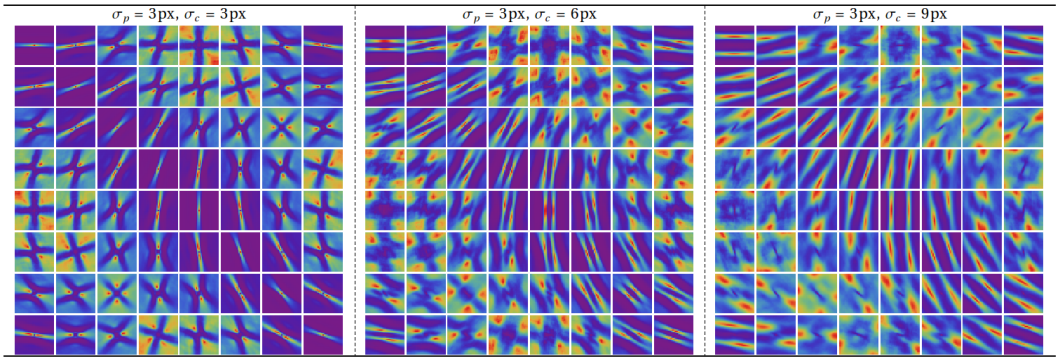


Figure 8.11: The histogram computed on the database of 100 random cartoon images. This figure is analogue to Fig. 8.6. The values of σ_c and σ_p are indicated in the top of each panel.

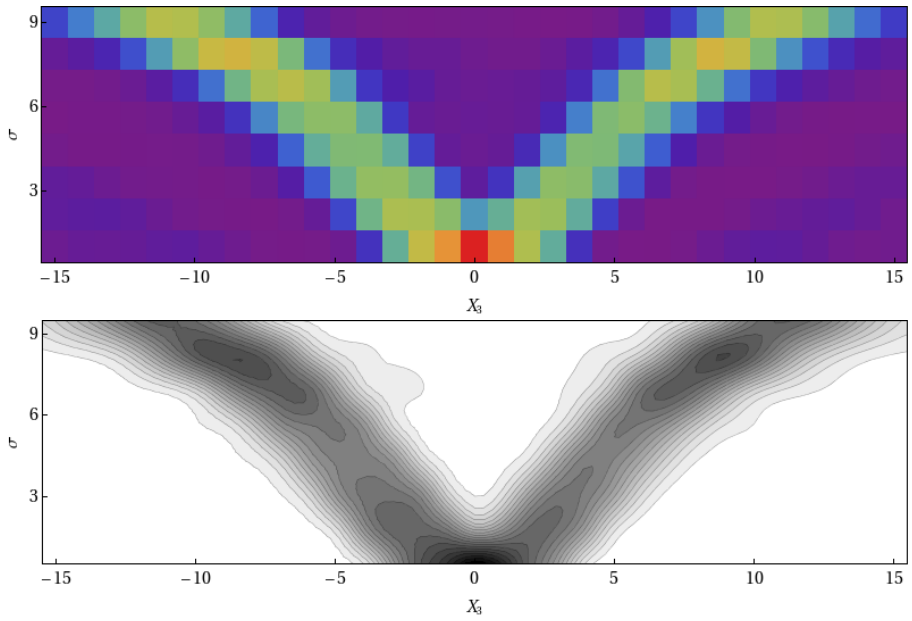


Figure 8.12: Visualization of the plane $\{X_{\sigma,3}, X_{\sigma,4}\}$ of the histogram computed in the cartoon database. The secondary structure observed in Fig. 8.8 almost disappeared.

8.4 Discussion

This chapter begins by reviewing a model of the cortex introduced in [84] which takes into account the scale as an engrafted variable in addition to

the orientation. The model is an extension of the $SE(2)$ model of the cortex used in the previous chapters. Now the group considered is the affine group, endowed with a symplectic structure which models two types of cortical connections, co-axial and trans-axial. The first structure is the same already described in the past chapters, so we focused in the second one.

It is important to outline that the construction of the cortical inspired space, specially of the geometrical structures therein, is analogue to that presented in Chapter 4. Firstly, a group of symmetries is chosen and identified with an appropriate fiber bundle. The coordinates of the cortical space are the parameters of the receptive profiles of the cells which takes into account the considered features. Equivalently, each coordinate may be intended as the element of the group operation which if applied to a mother kernel generates each receptive profile. The differential structure comes out by considering the kernel of a differential form which approximates the receptive profile. This same methodology may be extended to take into consideration other features.

The main contribution of this chapter is adding another source of experimental data, the statistic of natural images. This was achieved by extending the method presented in Chapter 7 so the scale can be taken into account. In that chapter, prove was given that the connectivity patterns present in the $SE(2)$ model (the fun of integral curves of the vector field $X_1 + kX_2$ and the Fokker-Planck's fundamental solution) were in accordance with the statistic of co-occurrences of oriented contours. In this chapter, we checked that the geometrical structure predicted by the symplectic model for modelling the trans-axial connections is encoded in a 6d histogram, which results from computing the statistics of co-occurrences of edges (which are considered with their orientation and scale). The main difference between the co-axial and the trans-axial models of connectivity is that second structure does not spread out. By spreading we intend the brush like shape of the association fields. We propose to model the propagation along the trans-axial field as the pure advection process 8.10. The trans-axial direction (indicated in this thesis by X_3), if projected to the retinal plane, is perpendicular to the contours. The mechanism of perceptual completion in this direction is called filling-in. Unlike the perceptual completion of contours in which the process is very well understood from both the neurophysiological and psychophysical point of view, the processes underlying the filling-in are not yet clear [72]. There is still much work to do in this area and the we think that the methodologies presented in this chapter are at least promising. However, still more physiological and phenomenological mechanisms need to be understood.

Finally, some considerations can be done on the scale feature. In the framework of the symplectic model, the scale is interpreted as the distance from a

contour. From a signal point of view a contour is a discontinuity. This interpretation only has a sense if the image is piecewise constant. Such an hypothesis may seem too restrictive, as natural images are clearly much more complex. However, piecewise constant functions are the output of an image segmentation [63, 60] which is the base of human perception according to Gestalt psychologists [46]. The input for the experiments are natural images which do not verify the cartoon model. This why we hypothesize that the secondary structure observed in Fig. 8.8 is due to the presence of texture. In order to confirm this, we computed the statistic on a database of synthetic cartoon type images and verified that the secondary structure practically disappeared.

Chapter 9

Ladder

9.1 Introduction

In Chapter 2 devoted to phenomenology, a recent psychophysical experiment of May and Hess [57] is reviewed. In there, the authors carefully compare the saliency of two perceptual structures: the snake and the ladder. In the snake configuration, Gabor patches are arranged tangentially to a path and the experiment tests a combination of the Gestalt laws of good continuation and proximity. In the ladder configuration the patches are orthogonal to the path and the Gestalt law of parallelism and proximity are tested. The targets are immersed in a background of identically shaped but randomly distributed elements. The two-alternative force choice method (2AFC) was used, in which an observer should indicate which one of two stimuli presented in a trial contains a target, with the target presented in only one of the two stimuli.

Previous works demonstrated that ladders are usually harder to identify than snakes [29, 37, 8, 52, 56], but [57] was the first work to test this systematically for different parameters. They reported data in agreement with previous works for the snake configuration and a surprising new characteristic of the ladder association field: the saliency of ladders is independent from the separation of elements in a wide range of the visual field. We will show that this result is in agreement with the sympletic geometrical model of the cortex introduced in Chapter 8 and with the statistics of natural images.

The chapter is organized as follows. We start by re-analyzing the data provided in [57] which is interpreted as a measurement of the association fields for snakes and ladders, and comparing the results with the integral curves of the symplectic model projected on the image plane. In section 9.2 we

use the framework developed in Chapter 7 for estimating a new distribution of co-occurrence of edge probabilities. We want to test the occurrence of parallel edges, so this analysis could have been done on the same histograms computed in Chapter 7, and the conclusions would have been similar. However, since it provides a slight improvement, we decided not to take into account polarity of edges. We show that this estimated probability distribution is compatible with the psychophysical experiment and with the symplectic model.

9.2 Psychophysical data

In this section we provide reinterpretation of the data from [57], already presented in Chapter 2. The authors reported the percentage of correct responses of two subjects for each combination of parameters as supplementary material. The separation between elements s (measured in degrees of the visual field) took values in $\{1.09°, 1.54°, 2.18°, 3.08°\}$, the carrier wavelength cw (same units) took values between $\{0.193°, 0.273°, 0.385°, 0.545°\}$ and the angle α took values between $\{0°, 10°, 30°, 40°\}$.

This work is focused only in the separation s and the angle difference α between Gabor elements. Therefore we decided to average the proportion of correct responses between all possible carrier wavelengths. This means that each trial with the same value of separation, angle path and type of stimulus, was considered a realization of the same experiment. We also take the average values between the two subjects. Even though, since we are only interested in qualitative interpretation, the average is not really important. Indeed, the considerations we will provide are also valid for one subject or a single wavelength.

Once the simple dimensionality reduction explained in the previous paragraph has been applied, our data is a function of two parameters s and α and takes values between 0 and 1, the percentage of correct responses to the 2AFC experiment. A value of 1 implies that all the responses are correct and therefore the target is always detected. A value near 0.5 is equivalent to a completely random choice and the target can not be distinguished from the background.

The parameters s and α completely define the relative position between elements. We performed a simple change of variables in order to use a local polar coordinate system with respect to the position and orientation of one of the patches (see Fig. 9.1). The coordinate change is then defined as:

$$(r,\theta) = \begin{cases} \left(s, \frac{\alpha}{2}\right) & \text{for the snake} \\ \left(s, \frac{\pi-\alpha}{2}\right) & \text{for the ladder} \end{cases} . \tag{9.1}$$

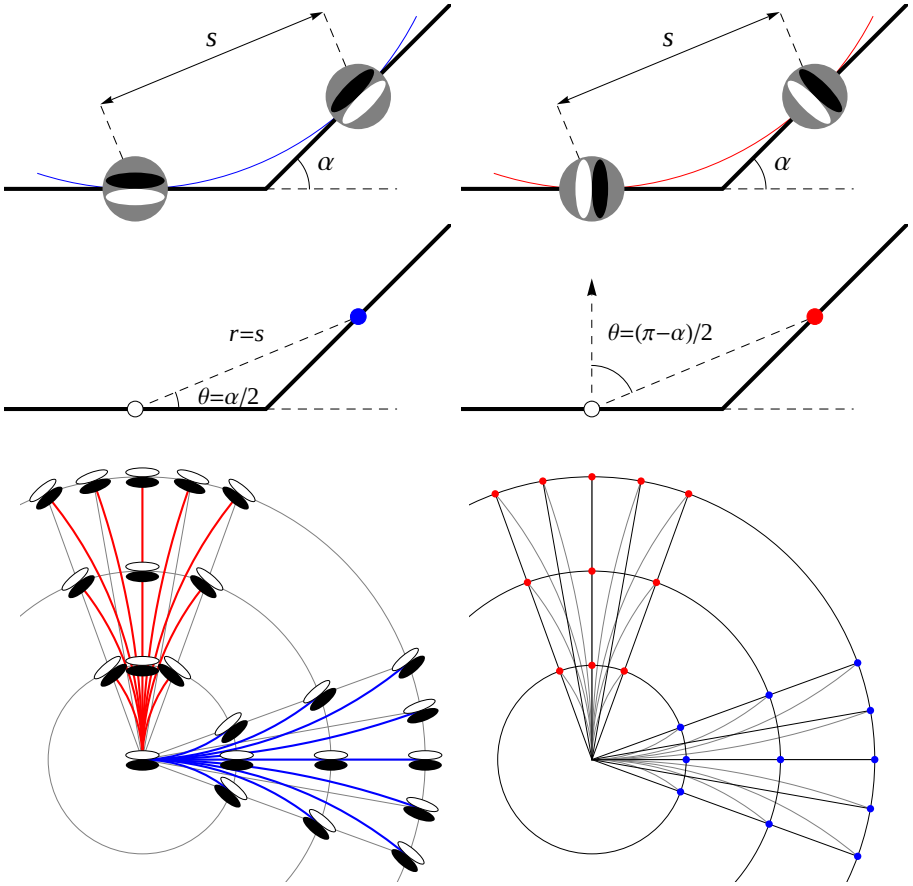


Figure 9.1: In blue it is shown the snake configuration and in red the ladder. The data reported in [57] are in the coordinates (s, α) indicated in the upper figures. We will consider the relative positions between elements in the polar coordinates (r, θ) defined as in eq. (9.1). This lets us represent both configurations on the same axes as shown within the bottom panel.

Fig. 9.2 presents the data after performing the collapse on the carrier wavelength parameter. The blue and red plots correspond to the snake and ladder configurations respectively. Show on the top row are the percentages of correct detections of the target as the separation between elements changes. Each single curve corresponds to a different value of α. The value of α for each plot is indicated on the left-bottom panel in the coordinates system (r, θ)

defined in (9.1). Plotted on the last panel is the averaged data for all possible angle differences.

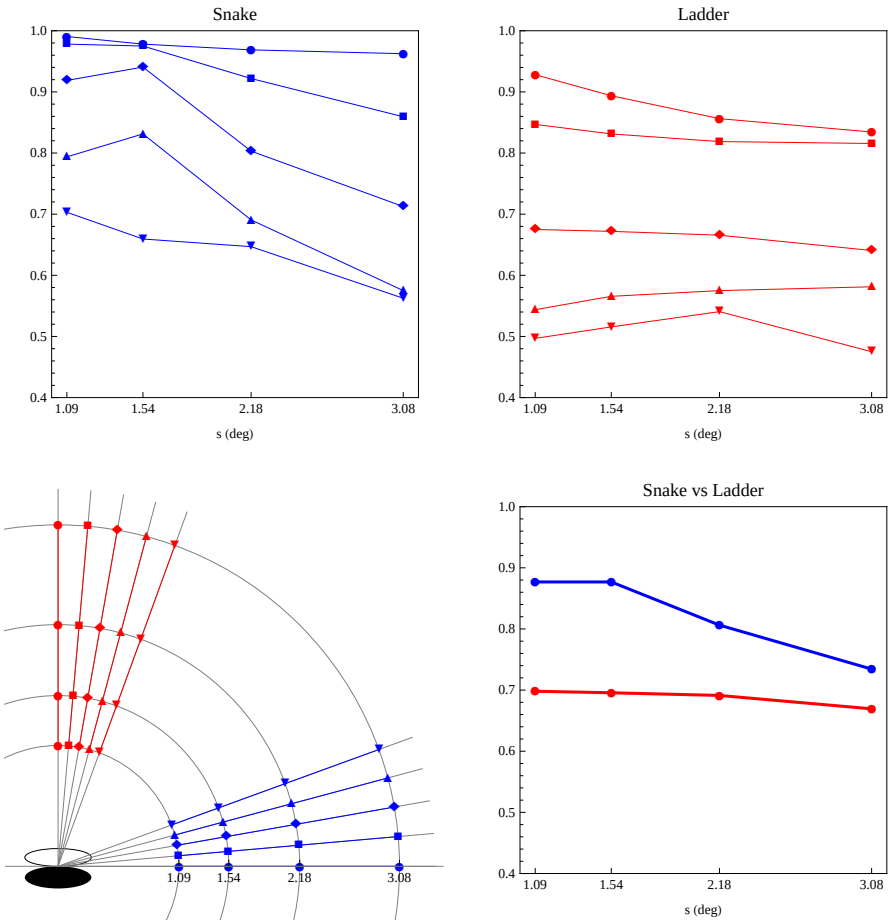


Figure 9.2: Reinterpretation of results reported by May and Hess [57]. The blue plots correspond to the snake configuration while the red plots to the ladder. The top row of panels shows the percentage of correct responses as the separation s grows, each of the five line plots corresponds to a different value of angle α (if α is constant all the Gabor patches must be tangent to the same arc of circumference). The respective angles are indicated in the left-bottom panel using the polar coordinates (9.1). The points correspond to relative positions between the patches. Plotted in the bottom right panel are the averages for both the ladder and snake detection performance.

Analyzing the graphics one can check May and Hess observations (see Section 2.5). From Fig. 9.2 we can infer that the subject's performance:

- decays as the angle difference grows;
- is better at detecting snakes than at detecting ladders;
- decays detecting snakes when the separation between elements grows;
- detecting ladders is almost independent from the distance between the elements.

An important property of the change of coordinates (9.1) is that we are able to plot the snake and the ladders data on the same polar axes. Indeed, in the space (r, θ) the output of the psychophysical experiment is a two dimensional function. In Fig. 9.3 (left) we used a color scheme to visualize it. The relative positions and orientations with respect to the central patch are indicated. The wavelength of the color is proportional to the amount of correct responses. The colors have been interpolated between the discrete data obtained from [57]. One can observe some differences between the areas corresponding to the snake and to the ladder. In the snake part of the image, one can see how the color spreads from the center approximately into the directions indicated by the orientations of the patches. On the other hand, in the ladder configuration the color is propagated radially without spreading. The significant perceptual structure is the parallel configuration, a small variation from this configuration strongly affects the ability to detect it.

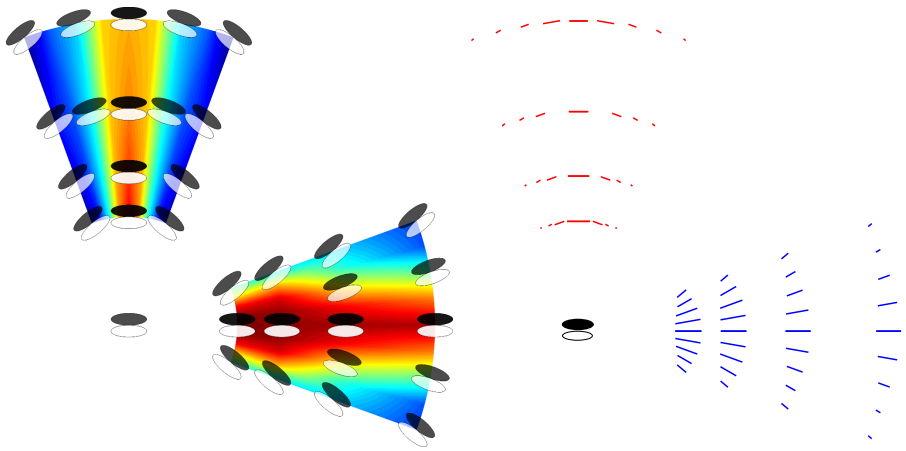


Figure 9.3: Reinterpretation of results reported by May and Hess [57].

In Fig. 9.3 (right) another visualization technique is used. The segments are centered at the points (r, θ) of the experiment. The orientation of the

segments is the relative orientation between two patches α. The length of the segments is proportional to the saliency of the configuration.

The behaviors of these association fields strongly agree with those of the integral curves of the symplectic model when projected onto the image plane (see Fig. 8.3). In the co-axial configuration, the fan of integral curves spreads as it happens with the saliency of the snakes (see Fig. 9.3). On the contrary, the trans-axial integral curves remains confined to a single curve and there is no spreading. A similar effect can be observed with the saliency of ladders. When the angle varies from the parallel configuration, the perception drastically decays. The agreement between the trans-axial integral curves and the experiment of May and Hess is even stronger. In the configuration, the separation between elements does not affect the saliency. In the model, this is interpreted by the fact that the respective integral curves do not spread.

9.3 Image Statistics

9.3.1 Computation of the image statistics

This section is devoted to the computation and analysis of the distribution of co-axial and parallel contours in natural images. The methodology for computing the histogram is the same as the one used in Chapter 7.

The probability of co-occurrence of edges is estimated by means of a co-occurrence histogram whose coordinates are the relative position and orientation of significant edges. The procedure for constructing the histogram, and the natural database are identical to that described in Chapter 7 (see Section 7.3). In this instance we do not take into account the polarity of contours. Therefore, the only methodological difference is that we used a local edge energy $E(\theta)$ instead of the odd symmetric filter $G(\theta)$ (eq 7.1). The edge energy combines the output of the odd symmetric filter $G(\theta)$ with that of an even symmetric filter $H(\theta)$:

$$E(\theta) = G^2(\theta) + H^2(\theta).$$

The angle θ is the orientation of the axis of the filter and it is discretized in 16 different orientations between 0 and π. As before, the kernels are steerable [30] and the parameters are the same as those reported in [86]. We considered for the co-occurrence edges a maximum distance of 64 pixels.

With these parameters, the dimension of the 4D histogram is 129×129×64×64. This histogram is visualized in Fig. 9.4.

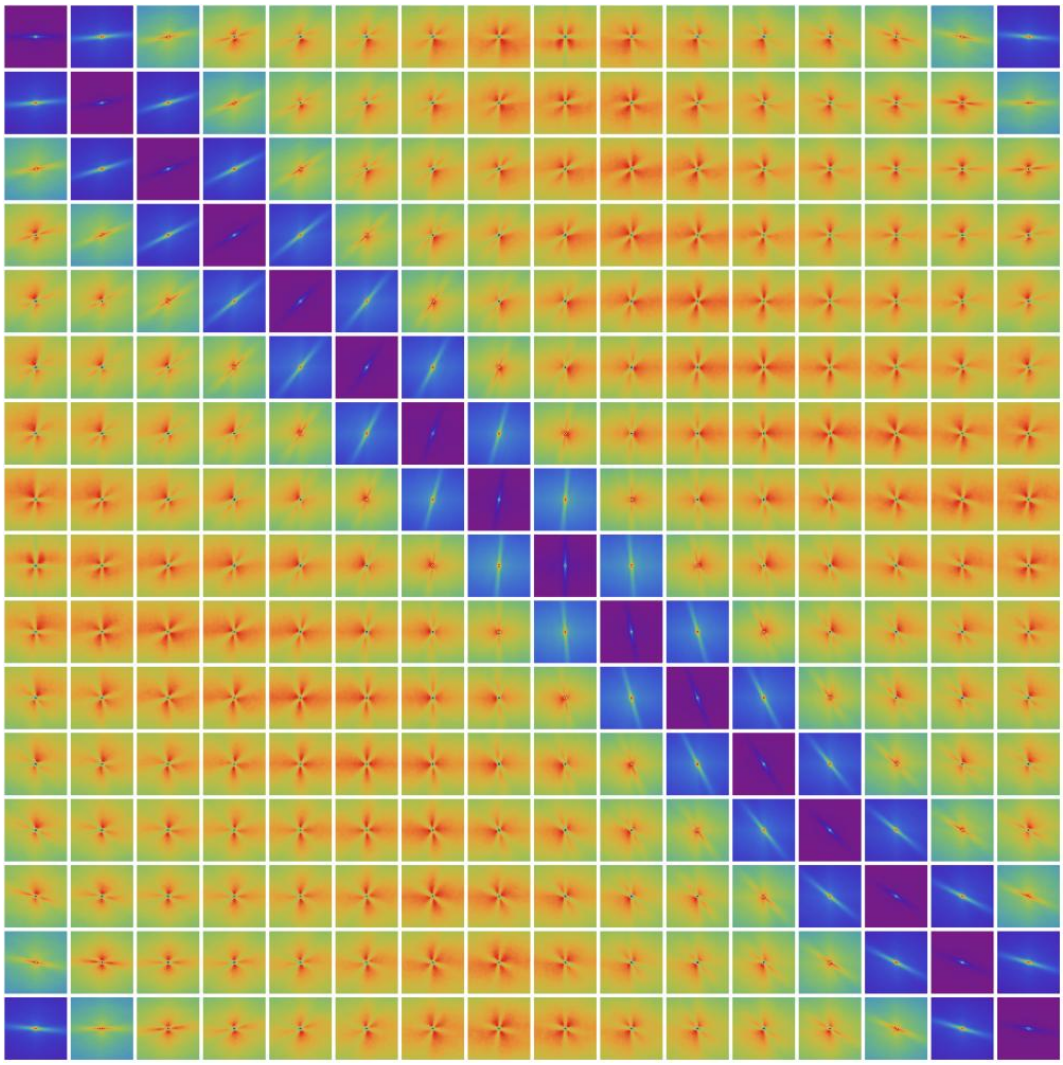


Figure 9.4: The 4D histogram computed with the edge energy (visualized with the same technique as in Fig. 7.10). Each panel corresponds to a different configuration of the two edges. The co-linear configuration is in the diagonal. The orientation of the central edge changes across the rows while the orientation of the relative edge changes across the columns. Each histogram is equalized singularly so that the full range of the colormap is exploited.

In our implementation, we have set the variance of the DoG filter to $\sigma = 1$ and chosen a support of 7 pixels as reported in [86]. The orientations were discretized at 16 different values between 0 and π.

9.3.2 Analysis

In this section we provide comparison between the psychophysical and statistical data of saliency. The results can be analyzed in Fig. 9.5. Shown in the upper panels is the resulting histogram for two edges with the same orientation (left) and some of its level curves (right). Let us first note that two different structures are clearly visible. An elongated co-axial correlation is present with high probability. This is due to points belonging to the same contour, in Chapter 7 this structure was extensively analyzed. A second isotropic structure characterizes the behavior of parallel covariances at low probability values.

The bottom panels of Fig. 9.5 show the plotting of the histogram values along the co-axial direction (blue plot) and trans-axial direction (red plot). The two lines are indicated in the histogram with their respective colors. The panel on the right is a zoom of the panel on the left, in the area indicated by the dashed rectangle. This graph should be compared with Fig. 9.2 (bottom-right panel), as the same properties are observed. The probability of the co-axial configuration is stronger than that of the trans-axial one. When the separation grows, the co-axial probability decays while the trans-axial probability remains approximately constant. Identical considerations to those realized in the previous section can be done in order to compare the statistical data with the symplectic model.

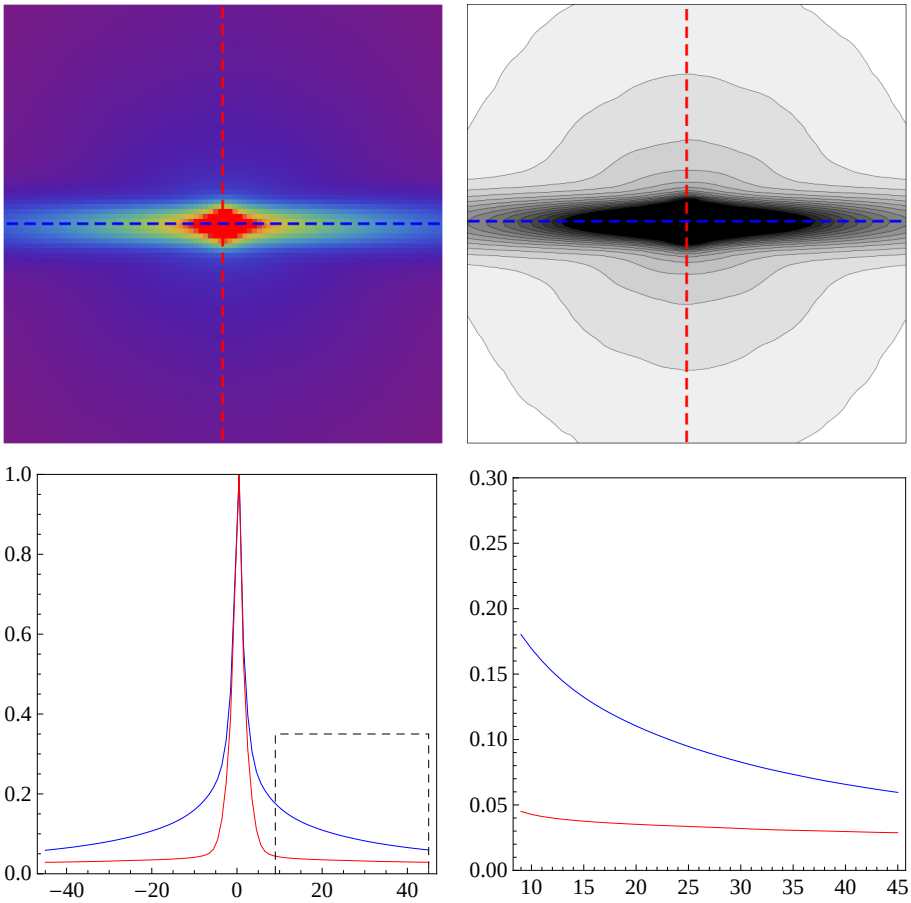


Figure 9.5: Visualization of the histogram: color density plot (upper-left) and level lines (upper right). The value of the histogram along the blue lines (co-axial) and red lines (trans-axial) are plotted in the bottom panels. The right panel corresponds to a zoom image of the area inside the dashed rectangle drawn above the left panel.

9.3.3 Discussion

This section has been inspired from the psychophysical results of the experiment presented by [57]. In there, the human performance in detecting contours and ladders was tested and systematically measured. One of the conclusions of the authors of the experiment was the following (page 19 of [57]):

> *"increasing the separation between the elements had a disruptive effect on the detection of snakes but had no effect on ladders, so that as separation increased, performance on the two types converged"*

The human ability to detect snakes is closely correlated with the perceptual process underlying integration and grouping of contour. This aspect has been faced with mathematical models of the functional architecture of V1 in Chapter 4 and with the statistics of natural images in Chapter 7. On the other hand, the grouping of oriented perceptual units parallel between them (ladder) is less understood. The reviewed psychophysical experiment proved that, within certain limits, the distance between the elements does not affect integration. This distance is measured along the direction perpendicular to the orientation of the patch, the trans-axial direction indicated by the vector X_3. This behavior can be explained in terms of the mathematical model of trans-axial connectivity introduced in the previous chapter. In fact integral curves in the direction X_3 do not spread out and remain confined to the axes perpendicular to the contour. This feature can explain why the salience of ladders is not spread and it is conserved (Fig. 9.3). Not only the perception of the ladders is weaker than the perception of snakes, it is independent from the distance between the patches, and approximately confined to the direction X_3. Something similar occurs with the statistics of natural images. First, the co-occurrence of parallel contours is weaker than the co-occurrence of the co-linear ones. However, it is statistically significant. Second, the probability of co-occurrence of this configuration does not depend on the distance between the contours. We can conclude that also for the feature of ladder there is good agreement between psychophysical measurement, statistics of natural images and Lie group based mathematical model.

Chapter 10

Conclusions

The aim of this thesis is to integrate three different scientific areas related to low level vision: phenomenology of perception, visual neurosciences and statistics of natural images. It is shown how the morphologies emerging in these three fields are very correlated one to the other and can be formalized using an unified, powerful mathematical setting, the theory of Lie groups. In the classical approach of naturalization of phenomenology of vision (Petitot and Citti-Sarti) the experimental heritage of phenomenology of perception has been reconsidered at the light of new discoveries of neurophysiology and modelled with suitable mathematical instruments.

The main contribution of this thesis has been to add a new empirical domain, namely the statistics of natural images and to integrate it with the phenomenological and neurophysiological data. The thesis proved the existence of strong isomorphisms between the morphologies observed in the different fields.

Suitable techniques to compute edge co-occurrences histograms in a large database of natural images have been proposed and discussed. In particular, great evidence has been produced that resulting data were in good agreement with the geometrical model of the cortex in $SE(2)$ and with the psychophysical data.

It has been observed that the natural mathematical instrument to model image statistics was a stochastic framework in Lie groups. These instruments have been adopted with success also to model non deterministic association fields and propagation of contours, showing how the methodologies set up in one domain, resulted well suitable for modelling in other domains.

The methods have been fully developed within the $SE(2)$ group, because of the large literature already present in this specific case. But the thesis

was not limited to this case and the approach introduced has been extended to the affine group. This group was recently used to model the functional architecture of the visual cortex taking into account the orientation and scale features. In the present study we have shown that recent psychophysical experiments fits very well with statistics of natural images in the affine group and both morphologies are well modelled by Lie group theory.

A discussion has been provided about which geometrical structure is encoded by the visual cortex. Three candidates were in order:

1. A section of the fiber bundle, encoding an orientation vector field as a function of the position in the retinal plane;
2. A second order tensor field as a function of the position encoding more information about orientation distributions;
3. A distribution of probability defined on the full fiber bundle as a function of the of position and orientation θ and corresponding to a infinite dimensional tensor field on the retinal plane.

Even if there is not a unique answer to the problem of geometrical encoding we have stressed the evidence that statistical distributions of co-occurrences naturally live in the full Lie group structure and that this is a much richer structure that the vectorial or the tensorial one.

Anyway this thesis leaves more open questions than answers:

- how these methods perform in processing of real images;
- how the physics of real neurons can be introduced in the Lie based theoretical framework;
- how the methods perform in higher dimensionality Lie groups dealing with other low level visual features;
- how the problem of binding and global figural emergence can take advantage from these architectures.

We hope that this is just the beginning of a fruitful field of new researches.

References

[1] D. Alexander, P. Sheridan, P. Bourke, O. Konstandatos, and J. Wright. Global and local symmetry of the primary visual cortex: Derivation of orientation preference. *http://www.mhri.edu.au/dma*, 2000.

[2] L. Ambrosio and S. Masnou. On a variational problem arising in image reconstruction. *Internat. Ser. Numer. Math., 147*, 1:17–26, 2005.

[3] J. August. *The curve indicator random field*. PhD thesis, Yale University, New Haven, CT, USA, 2001. Director-Steven W. Zucker.

[4] J. August and S. Zucker. Sketches with curvature: The curve indicator random field and markov processes. *IEEE Trans. Pattern Anal. Mach. Intell*, 25(4):387–400, 2003.

[5] C. Ballester, M. Bertalmío, V. Caselles, G. Sapiro, and J. Verdera. Filling-in by joint interpolation of vector fields and gray levels. *IEEE Transactions on Image Processing*, 10(8):1200–1211, 2001.

[6] O. Ben-Shahar and S. Zucker. Geometrical computations explain projection patterns of long-range horizontal connections in visual cortex. *Neural Computation*, 16(3):445–476, 2004.

[7] M. Bertalmío, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *SIGGRAPH*, pages 417–424, 2000.

[8] P. Bex, A. Simmers, and S. Dakin. Snakes and ladders: the role of temporal modulation in visual contour integration. *Vision Research*, 41(27):3775–3782, December 2001.

[9] T. Bonhoeffer and A. Grinvald. Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature*, 353(6343):429–431, October 1991.

[10] U. Boscain, J. Duplaix, J.-P. Gauthier, and F. Rossi. Anthropomorphic image reconstruction via hypoelliptic diffusion. *ArXiv e-prints*, June 2010.

[11] W. Bosking, Y. Zhang, B. Schofield, and D. Fitzpatrick. Orientation Selectivity and the Arrangement of Horizontal Connections in Tree Shrew Striate Cortex. *J. Neurosci.*, 17(6):2112–2127, 1997.

[12] T. Chan and J. Shen. Mathematical models for local nontexture inpaintings. *Journal of Applied Mathematics*, 62(3):1019–1043, 2001.

[13] Pascal Chossat and Olivier Faugeras. Hyperbolic planforms in relation to visual edges and textures perception. *PLoS Comput Biol*, 5(12):e1000625, 12 2009.

[14] G. Citti and A. Sarti. A cortical based model of perceptual completion in the roto-translation space. *Journal of Mathematical Imaging and Vision*, 24(3):307–326, may 2006.

[15] M. Crair, E. Ruthazer, D. Gillespie, and M. Stryker. Ocular dominance peaks at pinwheel center singularities of the orientation map in cat visual cortex. *J Neurophysiol*, 77(6):3381–3385, June 1997.

[16] J. Daugman. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision research*, 20(10):847–856, 1980.

[17] G. DeAngelis, I. Ohzawa, and R. Freeman. Receptive-field dynamics in the central visual pathways. *Trends in neurosciences*, 18(10):451–458, October 1995.

[18] A. Desolneux, L. Moisan, and J-M Morel. Computational gestalts and perception thresholds. *Journal of Physiology - Paris*, 97, 2002.

[19] R. Duits, M. Duits, M van Almsick, and B ter Haar Romeny. Invertible orientation scores as an application of generalized wavelet theory. *Pattern Recognition and Image Analysis*, 17:438–438, 2007.

[20] R. Duits and E. Franken. Left-invariant parabolic evolutions on se(2) and contour enhancement via invertible orientation scores part i: Linear left-invariant diffusion equations on se(2). *Quart. Appl. Math.*, 68:255–292, 2010.

[21] R. Duits and E. Franken. Left-invariant parabolic evolutions on se(2) and contour enhancement via invertible orientation scores part ii: Nonlinear

left-invariant diffusions on invertible orientation scores. *Quart. Appl. Math.*, 68:293–331, 2010.

[22] R. Duits and M. van Almsick. The explicit solutions of linear left-invariant second order stochastic evolution equations on the 2d euclidean motion group. *Quart. Appl. Math.*, 66:27–67, 2008.

[23] V. Barocas E. Sander. Comparison of 2d fiber network orientation measurement methods. *J Biomed Mater Res A*, 88(2):322–31, February 2009.

[24] J. Elder and R. Goldberg. Ecological statistics of Gestalt laws for the perceptual organization of contours. *J. Vis.*, 2(4):324–353, 8 2002.

[25] S. Esedoglu and R. March. Segmentation with depth but without detecting junctions. *Journal of Mathematical Imaging and Vision*, 18:7–15, 2003.

[26] L. C. Evans. An introduction to stochastic differential equations. Lecture Notes.

[27] D. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America. A, Optics and image science*, 4(12):2379–2394, December 1987.

[28] D. Field and A Hayes. Contour integration and lateral connections of v1 neurons. *In LM Chalupa and JS Werner (Eds.), The visual neurosciences*, 2:10691079, 2004.

[29] D. Field, A. Hayes, and R. Hess. Contour integration by the human visual system: Evidence for a local association field. *Vision Research*, 33(2):173–193, January 1993.

[30] W. Freeman and E. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:891–906, 1991.

[31] W. Geisler. Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology*, 59(1):167–192, 2008.

[32] W. Geisler, J. Perry, B. Super, and D. Gallogly. Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, 41(6):711 – 724, 2001.

[33] C. Gilbert, A. Das, M. Ito, M. Kapadia, and G. Westheimer. Spatial integration and cortical dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 93(2):615–622, January 1996.

[34] A. Grinvald, D. Shoham, A. Shmuel, D. Glaser, I. Vanzetta, and E. Shtoyerman. In-vivo optical imaging of cortical architecture and dynamics. *Modern Techniques in Neuroscience Research. U. Windhorst and H. Johansson Springer*, pages 893–969, 1999.

[35] D. Hannula, D. Simons, and N. Cohen. Imaging implicit perception: promise and pitfalls. *Nat Rev Neurosci*, 6(3):257–255, 2005.

[36] R. Hess and D. Field. Integration of contours: new insights. *Trends in Cognitive Sciences*, 3(12):480 – 486, 1999.

[37] R. Hess, T. Ledgeway, and S. Dakin. Impoverished second-order input to global linking in human vision. *Vision Research*, 40(24):3309–3318, October 2000.

[38] R. Hladky and S. Pauls. Minimal surfaces in the roto-translation group with applications to a neuro-biological image completion model. *Journal of Mathematical Imaging and Vision*, 36:1–27, 2010.

[39] W. Hoffman. Higher visual perception as prolongation of the basic lie transformation group. *Mathematical Biosciences*, 6:437 – 471, 1970.

[40] L. Hormander. Hypoeliptic second order differential equations. *Acta Mathematica*, 119:147–171, 1968.

[41] D. Hubel. *Eye, Brain, and Vision (Scientific American Library, No 22)*. W. H. Freeman, 2nd edition, May 1995.

[42] D. Hubel and T. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160:106–154, January 1962.

[43] D. Hubel and T. Wiesel. Ferrier lecture: Functional architecture of macaque monkey visual cortex. *Royal Society of London Proceedings Series B*, 198:1–59, May 1977.

[44] L. Iverson and S. Zucker. Logical/linear operators for image curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:982–996, 1995.

[45] J. Jones and L. Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *J Neurophysiol*, 58(6):1233–1258, December 1987.

[46] G. Kanisza. *Organization in Vision: Essays on Gestalt Perception*. Praeger, New York, NY, 1979.

[47] M. Katkov, M. Tsodyks, and D. Sagi. Analysis of a two-alternative force-choice signal detection theory model. *Journal of Mathematical Psychology*, 50:411–420, 2006.

[48] H. Knutsson. Representing local structure using tensors. In *The 6th Scandinavian Conference on Image Analysis*, pages 244–251, Oulu, Finland, June 1989. Report LiTH–ISY–I–1019, Computer Vision Laboratory, Linköping University, Sweden, 1989.

[49] J. Koenderink. The structure of images. *Biological Cybernetics*, 50(5):363–370, August 1984.

[50] J. Koenderink. *Solid shape*. MIT Press, Cambridge, MA, USA, 1990.

[51] N. Krüger. Collinearity and parallelism are statistically significant second-order relations of complex cell responses. *Neural Processing Letters*, 8(2):117–129, October 1998.

[52] T. Ledgeway, R. Hess, and W. Geisler. Grouping local orientation and direction signals to extract spatial contours: Empirical tests of association field models of contour integration. *Vision Research*, 45(19):2511–2522, September 2005.

[53] D. Martin, C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(5):530–549, 2004.

[54] L. Martinez and J. Alonso. Complex receptive fields in primary visual cortex. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry*, 9(5):317–331, October 2003.

[55] S. Masnou and J. M. Morel. Level lines based disocclusion. In *ICIP (3)*, pages 259–263, 1998.

[56] K. May and R. Hess. Dynamics of snakes and ladders. *Journal of Vision*, 7(12):1–9, 2007.

[57] K. May and R. Hess. Effects of element separation and carrier wavelength on detection of snakes and ladders: Implications for models of contour integration. *Journal of Vision*, 8(13):1–23, October 2008.

[58] G. Medioni. Tensor voting: Theory and applications, 2000.

[59] R. Miikkulainen, J. Bednar, Y Choe, and J Sirosh. *Computational Maps in the Visual Cortex*. Springer, 2005.

[60] J. M. Morel and S. Solimini. *Variational methods in image segmentation*. Birkhauser Boston Inc., Cambridge, MA, USA, 1995.

[61] M. Morrone and D. Burr. Feature detection in human vision: A phase-dependent energy model. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 235(1280):221–245, 1988.

[62] D. Mumford. Elastica and computer vision. In *Algebraic geometry and its applications (West Lafayette, IN, 1990)*, pages 491–506. Springer, New York, 1994.

[63] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.

[64] M. Nitzberg and D. Mumford. The 2.1-D sketch. In *International Conference on Computer Vision*, pages 138–144, 1990.

[65] M. Nitzberg, D. Mumford, and T. Shiota. *Filtering, Segmentation, and Depth*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1993.

[66] K. Ohki, S. Chung, P. Kara, M. Hubener, T. Bonhoeffer, and R. Reid. Highly ordered arrangement of single neurons in orientation pinwheels. *Nature*, 442(7105):925–928, August 2006.

[67] B. Oksendal. *Stochastic Differential Equations: An Introduction with Applications (Universitext)*. Springer, December 2005.

[68] C. Olman, P-F Van de Moortele, J. Schumacher, J. Guy, K. Ugurbil, and E. Yacoub. Retinotopic mapping with spin echo bold at 7t. *Magnetic Resonance Imaging*, 28(9):1258 – 1269, 2010.

[69] P. Olver. *Applications of Lie Groups to Differential Equations*. Springer-Verlag, 2000.

[70] F. Orabona, G. Metta, and G. Sandini. Learning association fields from natural images. *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW '06. Conference on*, pages 174–174, June 2006.

[71] P. Parent and S. Zucker. Trace inference, curvature consistency, and curve detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(8):823–839, 1989.

[72] L. Pessoa and P. De Weerd. *Filling-in: From perceptual completion to cortical reorganization.* Oxford University Press, New York. USA, 2003.

[73] J. Petitot. The neurogeometry of pinwheels as a sub-Riemannian contact structure. *Journal of Physiology-Paris*, 97(2-3):265–309, May 2003.

[74] J. Petitot. *Neurogeometrie de la Vision Modeles, Mathematiques et Physiques des Architectures Fonctionnelles.* Ecole Polytechnique, 2008.

[75] J. Petitot and Y. Tondut. Vers une neurogéométrie. fibrations corticales, structures de contact et contours subjectifs modaux, 1999.

[76] J. Petitot, F. Varela, B. Pachoud, and J-M. Roy, editors. *Naturalizing phenomenology: issues in contemporary phenomenology and cognitive science.* Stanford University Press, 1999.

[77] D. Ringach and R. Shapley. Reverse correlation in neurophysiology. *Cognitive Science*, 28(2):147–166, 2004.

[78] H. Risken. *The Fokker-Plank Equation. Methods of Solution and Applications.* Springer-Verlag, Berlin, 1989.

[79] G. Sanguinetti, G. Citti, and A. Sarti. Image completion using a diffusion driven mean curvature flow in a sub-riemannian space. *VISAPP*, 2:46–53, 2008.

[80] G. Sanguinetti, G. Citti, and A. Sarti. Implementation of a model for perceptual completion in r2xs1. In *Computer Vision and Computer Graphics. Theory and Applications*, volume 24, pages 188–201. Springer Berlin Heidelberg, 2009.

[81] G. Sanguinetti, G. Citti, and A. Sarti. A model of natural image edge co-occurrence in the rototranslation group. *Journal of Vision*, 10(14), 2010.

[82] A. Sarti and G. Citti. A cortical based model of perceptual completion in the roto-translation space. In *Proceeding of the Workshop on Second Order Subelliptic Equations and Applications. Cortona*, June,15-21 2003.

[83] A. Sarti and G. Citti. Non-commutative field theory in the visual cortex. *Computer Vision: From Surfaces to 3D Objects. (ed) Tyler, CW. CRC Press: Boca Raton, FL, USA.*, 2011.

[84] A. Sarti, G. Citti, and J. Petitot. The symplectic structure of the primary visual cortex. *Biological Cybernetics*, 98(1):33–48, 2008.

[85] A. Sarti, R. Malladi, and J. Sethian. Subjective surfaces: a method for completing missing boundaries. *Proceedings of the National Academy of Sciences of the United States of America*, 97(12):6258–63, 2000.

[86] M. Sigman, G. Cecchi, C. Gilbert, and M. Magnasco. On a common circle: Natural scenes and gestalt rules. *Proceedings of the National Academy of Sciences, USA*, 98:1935–1940, 2001.

[87] E. Simoncelli and B. Olshausen. Natural image statistics and neural representation. *Annu Rev Neurosci*, 24(1):1193–1216, 2001.

[88] K. Sinan, F. Worgotter, and N. Krüger. Statistical analysis of local 3d structure in 2d images. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:1114–1121, 2006.

[89] K. Thornber and L. Williams. Analytic solution of stochastic completion fields. *Biological Cybernetics*, 75(2):141–151, 1996.

[90] M. van Almsick, R. Duits, E. Franken, and B. ter Haar Romeny. From stochastic completion fields to tensor voting. In Ole Olsen, Luc Florack, and Arjan Kuijper, editors, *Deep Structure, Singularities, and Computer Vision*, volume 3753 of *Lecture Notes in Computer Science*, pages 124–134. Springer Berlin / Heidelberg, 2005.

[91] J. van Hateren. Processing of natural time series of intensities by the visual system of the blowfly. *Vision Research*, 37:3407–3416, 1997.

[92] J. van Hateren and H. Snippe. Information theoretical evaluation of parametric models of gain control in blowfly photoreceptor cells. *Vision Research*, 41(14):1851–1865, June 2001.

[93] Y. Wang, Y. Zhou, D. Maslen, and G. Chirikjian. Solving phase-noise fokker-planck equations using the motion-group fourier transform. *Communications, IEEE Transactions on*, 54(5):868 – 877, May 2006.

[94] M. Wertheimer. Untersuchungen zur lehre von der gestalt ii. *Psycologische Forschung*, 4:301–350, 1923.

[95] L. Williams and D. Jacobs. Stochastic completion fields: A neural model of illusory contour shape and salience. In *ICCV*, pages 408–415, 1995.

[96] L. Williams and D. Jacobs. Local Parallel Computation of Stochastic Completion Fields. *Neural Comp.*, 9(4):859–881, 1997.

[97] J. Wilson, H.; Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal*, 12(1):1–24, 1972.

[98] W. Zhu and T. Chan. A variational model for capturing illusory contours using curvature. *Journal of Mathematical Imaging and Vision*, 27:29–40, 2007.

[99] J. Zweck and L. Williams. Euclidean group invariant computation of stochastic completion fields using shiftable-twistable functions. *J. Math. Imaging Vis.*, 21(2):135–154, 2004.