Formal models of visual perception based on cortical architectures

Marta Favali

To cite this version:

Marta Favali. Formal models of visual perception based on cortical architectures. Numerical Analysis [math.NA]. Université Pierre et Marie Curie - Paris VI; Università degli studi (Bologne, Italie). Dipartimento di matematica, 2017. English. NNT : 2017PA066094 . tel-02105088v1

HAL Id: tel-02105088
https://theses.hal.science/tel-02105088v1

Submitted on 20 Apr 2019 (v1), last revised 30 Apr 2019 (v2)

UNIVERSITÉ PIERRE ET MARIE CURIE
ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA

DOCTORAL SCHOOL BRAIN-COGNITION-BEHAVIOUR
DOCTORAL SCHOOL IN MATHEMATICS

PREPARED AT CAMS, CNRS-EHESS (PARIS) AND DM - DIPARTIMENTO DI MATEMATICA (BOLOGNA)

Formal models of visual perception based on cortical architectures

Marta Favali

Doctoral thesis in Neuroscience and Mathematics.

Directed by Alessandro Sarti and Giovanna Citti.

Defended publicly the 18th of April 2017 in front of a jury composed of:

Advisor	Alessandro Sarti	CAMS (EHESS-CNRS), Paris, France
Advisor	Giovanna Citti	DM, Bologna, Italy
Secondary-advisor	Laurent Bonnasse-Gahot	CAMS (EHESS-CNRS), Paris, France
Reviewer	Mauro Ursino	DEI, Bologna,Italy
Reviewer	Raul Serapioni	DM, Trento, Italy
Examiner	Olivier Marre	UPMC, INSERM, CNRS, Paris, France
Examiner	Daniel Bennequin	IMJ-PRG, Paris, France
Examiner	Bart M. ter Haar Romeny	TU/e, Eindhoven, The Netherlands

UNIVERSITÉ PIERRE ET MARIE CURIE
ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA

ÉCOLE DOCTORALE CERVEAU-COGNITION COMPORTEMENT
ÉCOLE DOCTORALE DE MATHÉMATIQUES

CAMS, CNRS-EHESS (PARIS) ET DM - DIPARTIMENTO DI MATEMATICA (BOLOGNE)

Modèles formels de la perception visuelle basés sur des architectures corticales

Marta Favali

Thèse de doctorat en Neurosciences et Mathématiques.

Dirigée par Alessandro Sarti et Giovanna Citti.

Présentée et soutenue publiquement le 18 Avril 2017 devant un jury composé de:

Directeur de Thèse	Alessandro Sarti	CAMS (EHESS-CNRS), Paris, France
Directrice de Thèse	Giovanna Citti	DM, Bologne, Italie
Co-enchadrant	Laurent Bonnasse-Gahot	CAMS (EHESS-CNRS), Paris, France
Rapporteur	Mauro Ursino	DEI, Bologne, Italie
Rapporteur	Raul Serapioni	DM, Trento, Italie
Examinateur	Olivier Marre	UPMC, INSERM, CNRS, Paris, France
Examinateur	Daniel Bennequin	IMJ-PRG, Paris, France
Examinateur	Bart M. ter Haar Romeny	TU/e, Eindhoven, Pays-Bas

Keywords: **Cortical architecture, encoding, decoding, fMRI activity,image analysis, mathematical modelling, neural models, neuromathematics, neurogeometry, perceptual grouping, quantitative gestalt, retinal spectral clustering, visual cortex.**

Ai miei genitori e a Daniele.

Acknowledgements

I have always thought that the “word of gratitude” would have denoted the end of this path and it is with different emotions that I write these words.

My first words will be for my advisors. I would like to thank Alessandro Sarti and Giovanna Citti for sharing with me during these years their passion, their expertise and their time. You have introduced me into this fascinating world of neuroscience and mathematics, by encouraging me to achieve my goals and allowing me to do my best. Thanks for your honesty and patience during days (and nights) spent for working together. Thanks for having believed in me and for having given to me the possibility to be part of an amazing project. I would also like to thank Laurent for his precious help and for sharing his expertise with me.

In those years I had the chance to meet several exceptional collegues and (now) friends. Thanks to Benedetta for her support in this adventure, for our walks in the hall of EHESS, and for the thousands of coffee taken together. Thanks to Emre for his hospitality and availability, and for the thousands of beer taken together. Thank you both for your friendship. I would equally like to thank Samaneh, It has been a pleasure working together. I am also grateful to all the guys in Bologna, for having welcomed me in the “Math family”, you made me feel at home. Thanks to EHESS group, in particular to Thomas, Nathalie and Sandrine.

Thanks to all the friends that I met in Paris, you have been (and you will be) my family in those years. I am also very grateful to all the Italian friends who made this period less difficult thanks to their love. Thanks to whom always present, who came and visited me, who spent their evenings on Skype. I am lucky to have you all in my life. Thanks to Cate, Ceci, Marghe, Ari and all the “Santasofiese” team, to Sassi, Giuli, Marti, Chiara, Licia, Claudia, Silvia for the “Cesenate”group and to Meli, Vale and Babi for the “Bagnese” team.

A special thank goes to Luca, who lived with me this experience. I would never had the courage to come to Paris wthout you. You have always been my support and I will never forget it. Thanks to your family, it has been (and will always be) a part of me.

I miei ultimi ringraziamenti vanno alla mia famiglia, alla mamma Irma, al babbo Norberto, ai i fratelli e agli splendidi nipoti. Grazie di tutti i sacrifici che avete fatto per me, non solo in questi anni ma da sempre. Grazie per non avermi mai ostacolato ma per avermi dato la forza di affrontare questa avventura. Grazie per essermi stati e starmi sempre accanto. Spero di rendervi sempre orgogliosi e questo lavoro è per tutti voi.

This project has received funding from the European Union's Seventh Framework Programme, Marie Curie Actions- Initial Training Network, under Grant Agreement No. 607643, "Metric Analysis For Emergent Technologies (MAnET)".

Contents

1 *Introduction*

The objective of this thesis is to develop mathematical models of visual perception based on cortical architectures and to apply them to reproduce phenomenological experiments as well as to process natural images. We primarily focus on low level vision tasks and in particular we are interested in the problem of grouping and of individuation of perceptual units. In this setting we will face the problem of the reconstruction of illusory figures and the detection of retinal vessels in optical images. Then we consider the problem of encoding and decoding of the fMRI signal from in vivo acquired brain activity of visual cortex. This allows to estimate the structure of the cortex of a specific human patient and eventually to reconstruct the visual stimulus from fMRI activity, in a so called "brain reading" strategy. The difference between our approach and the state of the art literature consists in using previously defined neuromathematical models of the cortices as a-priori knowledge to regularise the in vivo estimated structure. Even if it is a long term objective, we propose a first approach to improve the results in this field.

The entire work of this thesis has been developed taking into account results from phenomenology of perception on one hand and results of neurophysiology on the other.

In the field of the phenomenology of perception, at the beginning of the last century, the theory of the Gestalt psychology [Wertheimer, 1938, Kohler, 1947, Koflka, 1935] defined the integration of contours and in particular they defined grouping laws underlying perception. These are crucial in the construction of visual objects: points with characteristics in common can be grouped together to form a new, larger visual object. Many psychophysical experiments have been proposed to measure the quantitative parameters of these laws. A particular interest of this thesis is the concept of association fields introduced by Field et al. [1993] which encodes different Gestalt principles (for example, good continuation and proximity). They showed that stimulus co-linearity and co-circularity play an important role for the feature of grouping. Their study showed how chances of perceiving the curvilinear path were high if the orientation of its features was the one tangent at that point and collapsed as their relative orientation deviated from being tangent.

On the other hand, in neurophysiology, an impressive amount of experiments confirm that the problem of grouping and of boundary detection is performed by the primary visual cortex (V1). The fundamental structures of V1 implemented in the neural circuitry are closely related to contour grouping [Hubel, 1995].

A mathematical framework, based on differential instruments, has been introduced to formalize these findings. The first geometrical models are due to Koenderink and van Doorn [1987], who underlined the differential action of perceptual mechanisms, and by Hoffman [1989], who described V1 as a fiber bundle equipped with a contact structure. This fiber bundle is the mathematical structure ideally modelling both the retinotopic and hypercolumnar structure. More recently, Petitot and Tondut [1999], reconsidering the fiber bundle model of Hoffman, proved that it is coherent with contemporary psychophysical and neurophysiological findings, and it is able to describe the association fields from one side and the functionality of simple cell from the other. Other model that considered a differential geometry approach was introduced by Zucker [2006]. In Citti and Sarti [2006], Citti and Sarti reconsidered this cortical structure as a Lie group equipped with a sub-Riemannian metric. They proposed to model the functional architecture as Lie groups, showing the relation between geometric integral curves, association fields, and cortical properties. This method has been implemented in Sanguinetti et al. [2008] and Boscain et al. [2012]. Exact solution of the Fokker-Planck equation has been provided by Duits and Van Almsick [2008] and their results have been applied by Duits and Franken [2009] to image processing.

These differential models are local, hence insufficient to explain the problems of grouping and constitution of a percept, since a perceived form is characterized by a global consistency.

Different authors qualitatively defined this consistency as pregnancy or global saliency [Merleau-Ponty and Smith, 1996], but only a few quantitative models have been proposed [Koch and Ullman, 1987]. In particular spectral approaches for image processing were proposed by Perona and Freeman [1998], Shi and Malik [2000], Weiss [1999], Coifman and Lafon [2006]. In Sarti and Citti [2015] it is shown how this spectral mechanism is implemented in the neural morphodynamics, in terms of symmetry breaking of mean field neural equations. In that sense, Sarti and Citti [2015] can be considered as an extension of Bressloff et al. [2002].

Our results fall within this approach and in particular the one introduced in Sarti and Citti [2015], but the main original contributions of the thesis can be described as follows:

- quantitative fitting between the computed kernels and the experimental ones is performed, in order to validate the model;
- the model is enriched exploiting the role of the polarity feature;
- the model is extended; including the feature of intensity and curvature, finally considering a 5-dimensional kernel;
- these models of cortical connectivity are applied to the problems of identification and reconstruction of natural images from human brain activity, adding into this analysis neurophysiological constraints.

The thesis is organized as follows:

Chapter 3 starts by briefly reviewing the Gestalt theory and some of its basic laws, that describe how elements tend to be perceptually grouped and made salient. The chapter continues by describing one key psychophysical experiment related to contour organization which inspired the concept of association fields: it is the classical result of Field et al. [1993]. The chapter ends by describing the visual cortex, introducing the visual pathways, the receptive fields and receptive profiles. The main structures implemented by neural circuitry are described: the layered, the retinotopic, the hypercolumnar structure. Finally, the pinwheel structure, that is the real topological implementation of the hypercolumns, and the horizontal connectivity are introduced. These structures will be relevant to the models presented in the following Chapters.

Chapter 4 reviews several differential models of the visual cortex, particularly focuses on the models of Petitot and Tondut [1999] in the Heisenberg group and of Citti and Sarti [2006] in the Rototranslation group. A differential structure is considered, in order to model the long range 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 association fields. Then stochastic models of cortical connectivity are presented [Mumford, 1994, Williams and Jacobs, 1997b, August and Zucker, 2000, 2003], defining the connectivity kernels.

We underline our contribution to the model, clarifying how it is neurally implemented in the cortex, extending it in high generality and providing a solid numerical implementation.

We conclude the chapter describing the problem of individuation of perceptual units.

In Chapter 5 we perform a quantitative validation of the model of cortical connectivity comparing the proposed kernels with neurophysiological data of horizontal connectivity [Bosking et al., 1997, Angelucci et al., 2002]. We make a comparison between the fundamental solution of the Fokker Planck equation with experimental data of Bosking et al. [1997], Ben-Shahar and Zucker [2004] and Gilbert et al. [1996], showing how the stochastic paths are implemented in the neural network. In particular, we consider the distribution of a tracer through lateral connection modeling each injection with stochastic paths. The bouton distributions are realizations of a stochastic process, in particular of a random walk in $\mathbb{R}^2 x S^1$ space. We show how the probability density obtained as a combination of Fokker Plank is an integration of stochastic paths. Moreover we propose to use also the Subelliptic Laplacian kernel, in order to account for the variability of connectivity patterns.

The chapter ends with a quantitative validation of these kernels, comparing to an experiment of Gilbert et al. [1996]. The link between the connectivity kernel and cell's response is underlined. This work is published in Favali et al. [2016b].

Chapter 6 is devoted to the generalization of the model of cortical connectivity. As defined in Chapter 4, the visual cortex has a modular structure and its cells are capable to extract several features. We present here the inclusion in the model of cortical connectivity of the feature of intensity and an extension to a 5 dimensional kernel in the lifted space of positions, orientations, intensities and curvatures. The results in this chapter are published in Favali et al. [2016a], Abbasi-Sureshjani et al. [2016a].

Chapter 7 presents applications of these models to the problem of individuation of perceptual units. The first application is in the analysis of illusory figures: we present the identification of perceptual units in experiments similar to the ones of Field et al. [1993] and in Kanizsa figures, underlying the role of polarity and comparing the behavior of the different kernels previously presented. Then, the method is applied to the analysis of retinal images, to face the problem of grouping during the tracking of blood vessels. We underline how the features of intensity and curvature become relevant to correctly identify the perceptual units in these images. These works represent original contributions of this thesis and are published in Favali et al. [2016b,a], Abbasi-Sureshjani et al. [2016a].

Chapter 8 describes how to combine the theory of cortical models with encoding and decoding techniques. The chapter starts recalling the functional MRI and describing an overview of brain encoding and decoding results in fMRI analysis [Chen et al., 2014]. The problem of identification of natural images from human brain activity and the results of Kay et al. [2008] on this topic are presented. A modified version of the encoding model is described, adding sparsity of the representation, a nonlinearity of logaritmic type to consider the structure of the cortex and a regularization in the subriemannian cortical structure. This represents an original contribution in the thesis. Image identification performances with our method are then described. The chapter continues by considering the problem of reconstruction of the stimulus, primarily using simple and artificial images. The inverse mapping is described and the fMRI response is represented from the flattened cortex to the retinal coordinates. Then, the reconstruction of images from Gabor wavelets and from the contrast energy is presented, underlying how the representation provided by a population of complex cells encodes the phase information that is needed for object recognition. The chapter ends defining the reconstruction of images from fMRI activity, comparing our method and the results obtained by Naselaris et al. [2009].

The last Section is devoted to the conclusions.

2 *Résumé*

L'objectif de cette thèse est de développer des modèles mathématiques de perception visuelle basés sur des architectures corticales et de les appliquer pour reproduire des expériences phénoménologiques ainsi que pour traiter des images naturelles. Nous nous concentrons principalement sur les tâches de vision de bas niveau et en particulier nous sommes intéressés par le problème du groupement et de l'individuation des unités perceptives. Dans ce contexte, nous ferons face au problème de la reconstruction des figures illusoires et de la détection des vaisseaux rétiniens dans les images optiques. Ensuite, nous considérerons le problème du codage et du décodage de l'activité cérébrale du cortex visuel obtenue par Imagerie par Résonance Magnétique fonctionnelle (IRMf). Ceci permet d'estimer la structure du cortex d'un patient spécifique et éventuellement de reconstruire le stimulus visuel de l'activité IRMf, dans une stratégie "de lecture du cerveau" (brain reading). La distinction entre notre approche et l'état de la littérature consiste à utiliser des modèles neuromathématiques du cortex comme connaissance a priori pour régulariser la structure estimée.

Même si c'est un objectif à long terme, nous proposons une première approche pour améliorer les résultats dans ce domaine. L'ensemble du travail de cette thèse a été développé en tenant compte des résultats de la phénoménologie de la perception d'une part et des résultats de la neurophysiologie de l'autre.

Dans le domaine de la phénoménologie de la perception, au début du siècle dernier, la théorie de la psychologie de la Gestalt a défini l'intégration des contours et en particulier Wertheimer [1938], Kohler [1947], Koflka [1935] ont défini le regroupement des lois de la perception.

Celles-ci sont cruciales dans la construction d'objets visuels : les éléments avec des caractéristiques en commun peuvent être regroupés pour former un nouvel objet visuel plus grand. De nombreuses expériences psychophysiques ont été proposées pour mesurer les paramètres quantitatifs de ces lois. Un intérêt particulier de cette thèse est le concept de champ d'association introduit par Field et al. [1993] lequel code différents principes de la Gestalt (par exemple, la bonne continuation et la proximité).

Ces auteurs ont montré que la co-linéarité de stimulus et la co-circularité jouent un rôle important dans la caractéristique du groupement. Leur étude a montré comment les chances de percevoir un chemin curviligne étaient élevées si l'orientation de ses éléments était tangente à ce chemin.

D'autre part, en neurophysiologie, une grande quantité d'expériences confirment que le problème du groupement et de détection des contours est effectué par le cortex visuel primaire (V1) [Hubel, 1995].

Un cadre mathématique, basé sur les instruments différentiels, a été introduit pour formaliser ces résultats. Les premiers modèles géométriques sont dus à Koenderink and van Doorn [1987], qui ont souligné l'action différentielle des mécanismes perceptifs, et à Hoffman [1989], qui a décrit V1 comme un faisceau de fibres équipé d'une structure de contact. Ce faisceau de fibres est la modélisation mathématique de la structure rétinotopique et de l'organisation en hypercolonnes. Plus récemment Petitot and Tondut [1999] reconsidèrent le modèle de faisceau de fibres de Hoffman, et prouvent qu'il est cohérent avec les résultats psychophysiques et neurophysiologiques contemporains, et qu'il est capable de décrire les champs d'association et la fonctionnalité des cellules simples.

Une autre modèle examiné dans l'approche de la géométrie différentielle a été introduit par Zucker [2006]. Dans Citti and Sarti [2006], Citti et Sarti reconsidérés cette structure corticale en tant que groupe de Lie équipé avec une métrique sous-riemannienn. Ils ont proposé de modéliser l'architecture fonctionnelle en tant que groupes de Lie, en montrant la relation entre les courbes intégrales géométriques, les champs d'association et les propriétés corticales. Cette méthode a été implémentée dans Sanguinetti et al. [2008] et Boscain et al. [2012]. La solution exacte de l'équation de Fokker-Planck a été fournie par Duits and Van Almsick [2008] et leurs résultats ont été appliqués par Duits and Franken [2009] au traitement d'image.

Ces modèles différentiels sont locaux, donc insuffisants pour expliquer les problèmes de groupement et de constitution d'une perception, car une forme perçue est caractérisée par une cohérence globale.

Différents auteurs ont qualitativement défini cette cohérence comme la saillance globale [Merleau-Ponty and Smith, 1996], mais seuls quelques modèles quantitatifs ont été proposés [Koch and Ullman, 1987]. En particulier, l'approche spectrale pour le traitement d'image a été proposée par Perona and Freeman [1998], Shi and Malik [2000], Weiss [1999], Coifman and Lafon [2006]. Sarti and Citti [2015] montrent comment ce mécanisme spectral est implémenté dans la morphodynamique neuronale, en termes de rupture de symétrie des équations neuronales de champ moyen. Dans ce sens, le modèle de Sarti and Citti [2015] peut être considéré comme une extension de Bressloff et al. [2002].

Nos résultats s'inscrivent dans cette approche et en particulier celle introduite dans Sarti and Citti [2015] et les principales contributions originales de la thèse peut être décrit comme suit:

- un ajustement quantitatif entre les noyaux calculés et les essais expérimentaux est effectué, afin de valider le modèle;
- le modèle est étendu en exploitant le rôle de la caractéristique de polarité;
- le modèle est étendu y compris la caractéristique d'intensité et de courbure, en considérant un noyau à 5 dimensions;
- ces modèles de connectivité corticale sont appliqués aux problèmes d'identification et de reconstruction d'images naturelles à partir de l'activité cérébrale humaine, en ajoutant à cette analyse des contraintes neurophysiologiques.

La thèse est organisée comme suit:

Le chapitre 3 commence par une brève revue de la théorie de la Gestalt et de quelques-unes de ses lois fondamentales, qui décrivent comment les éléments tendent à être perceptivement groupés et faits saillants. Le chapitre continue en décrivant une expérience psychophysique clé liée à l'organisation du contour qui a inspiré le concept des champs d'association : c'est le résultat classique de Field et al. [1993]. Le chapitre se termine par la description du cortex visuel, en introduisant les voies visuelles, les champs récepteurs et les profils récepteurs. Les structures principales des circuits neuronaux sont décrites : les couches, l'organisation rétinotopique, les hypercolonnes. Enfin, la structure “pinwheels” et la connectivité horizontale sont introduites. Ces structures seront importantes pour les modèles présentés dans les chapitres suivants.

Le chapitre 4 examine plusieurs modèles différentiels du cortex visuel, puis se concentre sur les modèles de Petitot and Tondut [1999] dans le groupe de Heisenberg et de Citti and Sarti [2006] dans le groupe Rototranslation. Ces modèles considerent une structure différentielle, afin de modéliser les connexions horizontales entre hypercolonnes. La structure est formalisée comme l'algèbre de Lie de la SE(2) avec une structure Sub-Riemannienne. Les courbes intégrales de ces champs vectoriels générateurs modélisent les champs d'association. Les modèles stochastiques de connectivité corticale sont alors présentés [Mumford, 1994, Williams and Jacobs, 1997b, August and Zucker, 2000, 2003], définissant les noyaux de connectivité.

Nous soulignons notre contribution au modèle, en précisant comment il est implémenté neuralement dans le cortex, en fournissant une exécution numérique solide.

Nous concluons le chapitre avec une présentation du problème de l'individuation des unités perceptuelles.

Dans le chapitre 5 nous effectuons une validation quantitative du modèle de la connectivité corticale comparant les noyaux avec des données neurophysiologiques de connectivité horizontale [Bosking et al., 1997, Angelucci et al., 2002].

Nous faisons une comparaison entre la solution fondamentale de l'équation de Fokker Planck avec les données expérimentales de Bosking et al. [1997], Ben-Shahar and Zucker [2004] et Gilbert et al. [1996], montrant comment les chemins stochastiques sont implémentès dans le réseau neuronal.

Nous considérons la distribution d'un traceur à travers la connexion latérale avec la modélisation stochastique de chacun des chemins d'injection.

Les distributions des boutons sont des réalisations de processus stochastique, en particulier d'une marche aléatoire dans l'espace $\mathbb{R}^2 x S^1$.

Nous montrons comment la densité de probabilité obtenu sous la forme d'une combinaison de Fokker Planck est une intégration de trajectoires aléatoires. De plus, nous proposons d'utiliser également le noyau Laplacien-elliptique, afin de tenir compte de la variabilité des modèles de connectivité.

Le chapitre se termine par une validation quantitative de ces noyaux, comparé à une expérience de Gilbert et al. [1996]. Le lien entre le noyau de la connectivité et de la réponse cellulaire est soulignée. Ce travail est publié dans Favali et al. [2016b].

Le chapitre 6 introduit la généralisation du modèle de connectivité corticale. Comme défini au chapitre 4, le cortex visuel a une structure modulaire et ses cellules sont capables d'extraire plusieurs caractéristiques. Nous présentons ici l'inclusion dans le modèle de connectivité corticale de la caractéristique d' intensité et une extension à un noyau à 5 dimensions dans l'espace des positions (2 dimensions), des orientations, des intensités et des courbures.

Les résultats de ce chapitre sont publiés dans Favali et al. [2016a], Abbasi-Sureshjani et al. [2016a].

Le chapitre 7 présente les applications de ces modèles au problème de l'individuation des unités perceptuelles. La première application est dans l'analyse de figures illusoires : nous présentons l'identification d'unités perceptives dans des expériences similaires à celles de Field et al. [1993] et de Kanizsa, en considérant le rôle de la polarité et en comparant le comportement des différents noyaux présentés précédemment.

Ensuite, la méthode est appliquée à l'analyse des images rétiniennes, afin d'examiner le problème du groupement des vaisseaux sanguins. Nous soulignons comment les caractéristiques d'intensité et de courbure deviennent fondamentales pour identifier correctement les unités perceptives dans ces images.

Ces travaux représentent des contributions originales de cette thèse et sont publiés dans Favali et al. [2016b,a], Abbasi-Sureshjani et al. [2016a].

Le chapitre 8 décrit comment combiner la théorie des modèles corticaux avec les techniques de codage et de décodage. Le chapitre commence avec une description de l'IRM fonctionnelle et il décrit une vue d'ensemble des résultats de codage et de décodage du cerveau dans l'analyse IRMf [Chen et al., 2014]. Le problème de l'identification des images naturelles à partir de l'activité du cerveau humain et les résultats de Kay et al. [2008] sur ce sujet sont présentés.

Une version modifiée du modèle d'encodage y est décrite, en ajoutant la sparsité de la représentation, une non linéarité de type logarithmique pour considérer la structure du cortex et une régularisation dans la structure cortical subriemannienne. Cela représente une contribution originale dans la thèse. Les performances d'identification d'image avec notre méthode sont ensuite décrites. Le chapitre continue en considérant le problème de la reconstruction du stimulus, principalement avec des images simples et artificielles. La méthode de “mapping inverse” est décrite, pour laquelle la réponse IRMf est représentée à partir du cortex jusqu’aux coordonnées rétiniennes. La reconstruction des images à partir du filtre de Gabor et de l’énergie de contraste est ensuite proposée.

Le chapitre présente enfin la reconstruction des images à partir de l’activité IRMf, en comparant notre méthode et les résultats obtenus par Naselaris et al. [2009].

La dernière section conclut par un discussion generale.

3 Phenomenology of perception and neurophysiology of the visual cortex

3.1 Phenomenology of perception

We start this chapter recalling the main rules of the Gestalt psychology, one of the most classical school of phenomenology of vision, which expressed local and global laws at the basis of grouping and constitution of percepts. Another crucial element which can contribute to the explanation of segmentation is the association field, introduced by Field, Hayes and Hess to describe which parts of the stimuli can be associated to the same perceptual unit. A detailed analysis of phenomenology of perception or of the visual cortex is out of the aim of this Chapter, but we will focus on the structures relevant to the models presented in the following ones. The neurological basis of these perceptual phenomena have to be searched in the functional architecture of the primary visual cortex (V1). We introduce the visual pathways, the receptive fields and receptive profiles and the layered, retinotopic and the hypercolumnar structure. Finally, the pinwheel structure and the horizontal connectivity will be described.

3.1.1 Gestalt Theory

An important process in visual perception is represented by perceptual grouping. Since the beginning of the last century, Gestalt theory began to define laws of perception, according to which distinct visual stimuli can be perceived as a single perceptual unit. It formulated both local and global laws which can explain the grouping process (see for example [Wertheimer, 1938, Kohler, 1947, Koflka, 1935] and for a recent review we quote: [Wagemans et al., 2012]). This process is fundamental in visual perception, when points have one or several characteristics in common they form a new and larger visual object. Gestalt psychologists formulated rules for perceptually significant image structure and describe how elements tend to be grouped together. The idea is that complex phenomena can be understood considering the idea of structure more than a single element. In particular, in order to individuate perceptual units, gestalt theory has introduced local and global laws, that describe the influence of global context in the perception of local features. Among the local laws we recall here the principle of good continuation, similarity and proximity:

- *good continuation*: elements aligned (or with comparable alignment) tend to form a continuous curve. As an example, in Figure 3.1 we clearly perceive an unique curve of dots crossing a black rectangle and not two distinct curves;

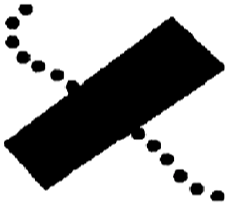


Figure 3.1: Examples of good continuation Gestalt laws. Adapted from: [Kanizsa, 1979].

- *similarity*: elements similar in color, texture, shape or orientation are grouped together (see Figure 3.2);

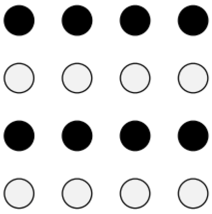


Figure 3.2: Examples of similarity Gestalt laws. Adapted from: [Kanizsa, 1979].

- *proximity*: elements that are close to each other and apart enough from the rest of the elements form a group or a cluster (see Figure 3.3).

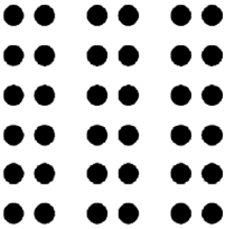


Figure 3.3: Examples of proximity Gestalt laws. Adapted from: [Kanizsa, 1979].

These properties describe how elements tend to be perceptually grouped and made salient. More than one grouping laws can play a role to the perception of a complex object. At local level, the central idea of the Gestalt psychology is that continuity is essential in perception. For example phenomena in which there is a phenomenological presence of boundaries without a physical stimulus describe the mechanisms of boundary completion. These local laws of good continuation and alignment have a central role in phenomena as the modal completion of Kanizsa presented in the next Section, but can not be sufficient to completely justify them. Indeed, in Figure 3.4 (left) we clearly perceive the square if we fix at the center of the figure, but the subjective countours completely disappear if we try to fix them, or if we focus on a packman inducer. As a consequence global laws are necessary to explain the perception. In the construction of percepts and in particular in the figure-ground articulation it is crucial a notion of saliency which denotes the relevance of a form with respect of a contextual frame, the power of an object to be present in the visual field.

Due to the perceptual grouping process the scenes are perceived as constituted by a finite number of figures and the saliency assigns a discrete value to each of them. In particular the most salient configuration pops up from the ground and becomes a figure [Merleau-Ponty and Smith, 1996]. Note that in case of continuous deformation of the visual stimulus, the salient figures can change abruptly from one percept to a different one [Merleau-Ponty and Smith, 1996]. This happens for example in Figure 3.4 where a regular deformation is applied to the Kanizsa square: we progressively perceive a more curved square, until it suddenly disappears and the 4 inducers are perceived as stand alone (see for example [Lee, 2001, Pillow and Rubin, 2002, Petitot, 2008]).

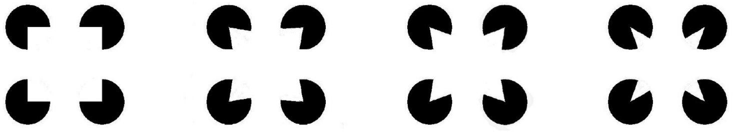


Figure 3.4: Deformation of visual stimulus, represented by squares with different angles between the inducers: the angle regularly decreases and we perceive regular deformations of the subjective Kanizsa square up to a certain value of curvature, when the square suddenly disappears and the inducers are perceived.

From the previous example, we can observe that the perceptual reality is composed by a discrete number of objects that do not necessarily depend on the existence of the correspondent physical objects. These phycological concepts are based on how the perception can influence the human behavior. Indeed, the visual stimulus is composed by indipendent points, but the process performed by the visual system induces the perception of perceptual units and objects, and a figure-ground segmentation, described in the following section.

3.1.2 Figure-ground segmentation

This articulation of the visual field in figure and ground, representing the minimal structure of visual perception, is one of the fundamental process at the base of visual experience and of the concept formulated by the Gestalt theory. It represents the tendency of the perceptual field to distinguish between what is primary and important in the figure and what it is secondary. The figure is defined as the elements delimited by a boundary that attracts our attention; the background is everything that appears distant and acts as a "frame" of the picture. Between the figure-ground segmentation laws we recall:

- *inclusion*: at the same conditions, the regions of the image that are enclosed or surrounded are perceived as the figure;
- *relative area*: regions that are smaller in area are more likely to be seen as figure;
- *orientation*: observers are more likely to perceive the figure on the side of the edge where the familiar object lies.

Summarizing, regions that are convex, symmetric, smaller in area, enclosed, or surrounded are more likely to be seen as figure than contiguous regions that are concave, asymmetric, larger in area, or surrounding. The Gestalt psychologists held that these properties for figure-ground perception were largely innate and did not depend upon an individual's past experience [Wertheimer, 1938].

When none of these conditions allow to distinguish what is salient in the figure, the perception becomes ambiguous. In the case of ambiguous figures the problem of the selection of figure-ground is particularly evident, as visualized in Figure 3.5. This optical illusion may be perceived either as a vase or as two human profiles facing each other, so it is not possible to perceive a more salient object in the scene.

The black regions in Figure 3.5 appear shapeless when they are seen as grounds to the white vase, while they appear shaped like profiles of faces when seen as figures. In a similar way, the white region appears shapeless when it is seen as the ground to the black profile faces, while appears to be a vase when it is

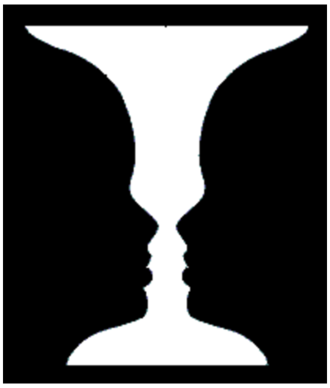


Figure 3.5: The Rubin vase. This illusion was created by the Danish psychologist Edgar Rubin.

seen as figure. Thus, regions appear shapeless when they are seen as grounds even though the same regions appear shaped when they are perceived to be figures.

3.1.3 Perceptual Completion

Kanizsa in Kanizsa [1979, 1980] underlined that visual perception represents a complex process which involves both the physical stimuli and their phenomenological organization, which in general do not concide. This happens in a very clear way in the subjective completion process, which, for this reason, can be considered a first crucial tool for understanding visual perception. Kanizsa defined in particular two modality of completion: modal and amodal completion.

In the first one, that is present for example in the famous Kanizsa square (see Figure 3.4 (left)), we perceive an image whose boundaries are not present in the physical stimulus with the full modality of vision: the square pops up from the background and we perceive it with a gray level different from the background. Another example is presented in Figure 3.6 (left), where a Kanizsa triangle is phenomenologically perceived even if the boundaries are not present in the image. There is an apparent contour separating the triangle from the figure, indeed the interior looks whiter than the background. This modal completion gives rise to the well known phenomenon of illusory boundaries or subjective contours. The amodal completion (see Figure 3.6 (right)) is much more common, since it is caused by any partially occluded objects. The figure perceived is a black circle occluded by the gray square. The circle is present in the visual field but its completion is performed without an illusory contour. In this case we perceive a completion without the modality of vision.

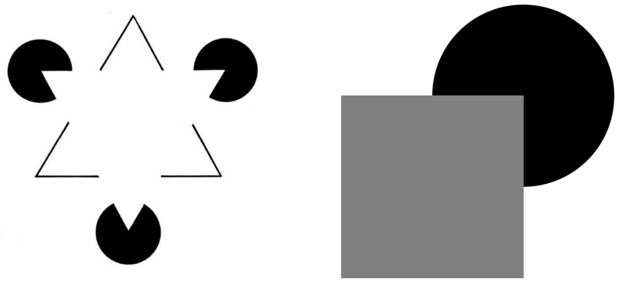


Figure 3.6: Left: the Kanizsa triangle as an example of modal completion. Right: an example of amodal completion. Adapted from: [Kanizsa, 1979].

3.1.4 Good continuation and association fields

In the previous Section it is described how in the problem of perceptual grouping the local law of good continuation plays a central role.

A number of results have been provided in order to refine the principles of psychology of form and assess neural correlates of the good continuation law. In particular, Grossberg and Mingolla [1985] introduced a "cooperation field" to model illusory contour formation. Similar fields of association and perceptual grouping have been produced by Parent and Zucker [1989]. In this contest, in the 1990s Kellman and Shipley provided a theory of object perception that specifically adressed perception of partially occluded objects and illusory contours [Kellman and Shipley, 1991, Shipley and Kellman, 1992, 1994]. Von Der Heydt et al. [1993] provided a theory of figural completion which can be applied to both illusory contour figures (as the Kanizsa triangle) and real images.

In the same years Field et al. [1993] introduced through psychophysical experiments the notion of *association fields*, to formalize the Gestalt principle of good continuation.

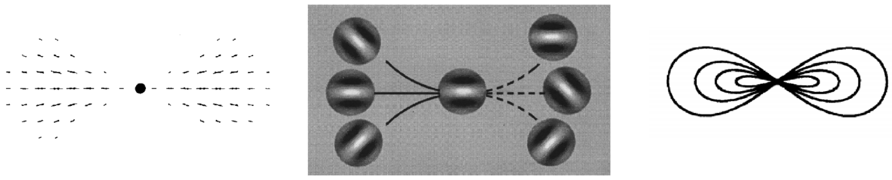


Figure 3.7: Field through the ages: the cooperative field of Grossberg and Mingolla [1985] (left), the association field of Field et al. [1993] (middle) and the two grouping fields of opposite orientation of Von Der Heydt et al. [1993] (right).

They developed a new approach to psychophysically investigate how the visual system codes contour continuity. The experiment introduced consists in showing to an observer an image composed by a small number of aligned Gabor patches (Figure 3.8 (b)) in a background of a large number of the same patches randomly distributed (Figure 3.8 (a)). The purpose was to test the ability of the subject to detect the perceptual units present in the visual stimulus and was repeated with different curves, changing the alignement of patches, their distances and their orientation.

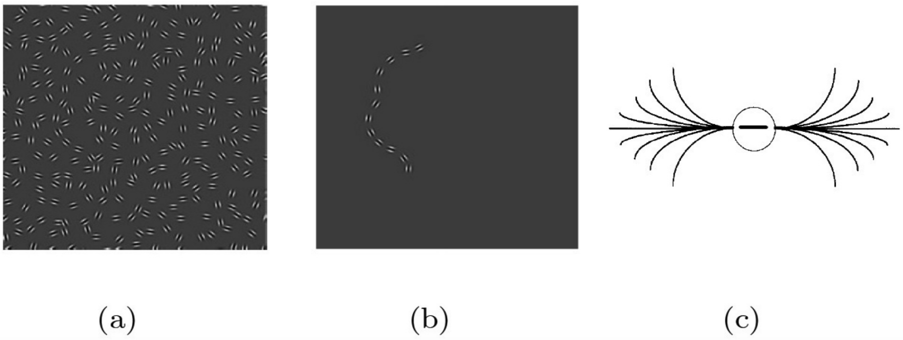


Figure 3.8: The stimulus proposed by Field, Hayes and Hess [Field et al., 1993] (a) and the perceptual unit present in it (b). In (c) the field lines of the association field [Field et al., 1993].

In this way they were able to describe the whole set of points and orientations which can be mutually connected. Through a series of similar experiments (see Figure 3.9), they constructed an association field, that describes the pattern of position - orientation elements of stimuli that can be associated to the same perceptual unit and represents the elements in the path which can be associated to the central point (see Figure 3.8 (c)). The stimulus in the central position can be joined with other stimula tangent to the lines in the figure but can not be joined with stimula with a different direction (see Figure 3.7 (middle)).

Based upon these results, they suggested that local interactions between contour elements follow specific rules and represent the basis for contour integration in humans.

3.2 The visual cortex

The origin of the previous described perceptual phenomena has to be found in the functionality of the primary visual cortex and its structures, which we describe here. The visual system has dedicated pathways through the multiple visual areas, that are related to separate functional measured properties as shape, color, motion and disparity [Hubel and Wiesel, 1977, 1962].

The cerebral cortex is the outermost layer of neural tissue in the two cerebral hemispheres and plays a central role in cognitive and sensory processing. It is composed by sensory, motor and association areas. The parts of the cortex that receive sensory inputs are the sensorial areas and the one that serves the sense of vision and receives the information from the visual path is the visual cortex. The visual pathway is represented in

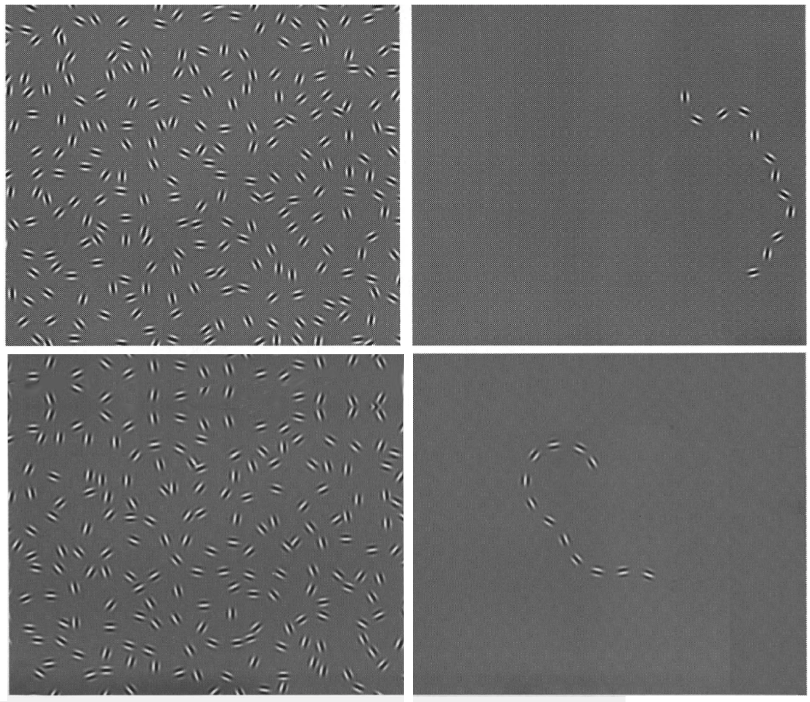


Figure 3.9: Test images of psychophysical experiments, used by Field et al. [1993]. The experiment consists of showing to subjects a grid made of oriented Gabor patches (left). In some images, the grid contains elements aligned over a path and other similar patches randomly distributed. In the other 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 (right).

Figure 3.10, where the left visual field is processed in the right half of the brain and viceversa [Zeki, 1993].

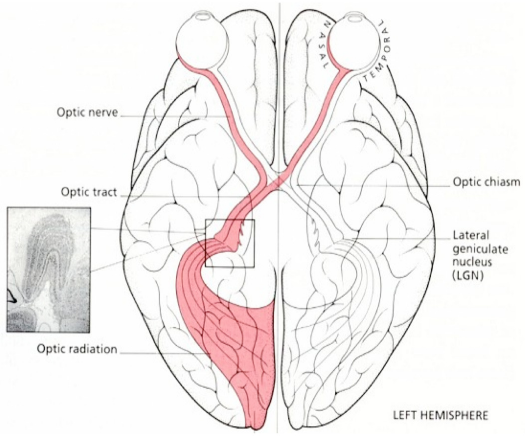


Figure 3.10: The visual pathway. The left visual field is processed in the right half of the brain. Source: [Zeki, 1993].

From the retina, the optic nerve runs into the central brain area and makes a connection in the Lateral Geniculate Nucleus (LGN), a specialized area of the thalamus. The LGN is the primary processing center for visual information received from the retina of the eye; in this area all incoming perceptual informations come together, as the visual, the tactile and the auditory ones.

The LGN consist of 6 layers; the top four layers have small cell bodies and form the parvo-cellular layers (Latin: parvus = small). The bottom layers have larger cell bodies and form the magno-cellular layers (Latin: magnus = big). The parvocellular layers contain shape and color information; the magnocellular layers are involved in mediating motion information. The mapping from the retina to the LGN is very precise, each layer is a retinotopic map of the retina.

Then, the axons of the cells of the LGN project the visual signal to the primary visual cortex (V1), also known as the striate cortex, a region in the calcarine sulcus in Brodmann area 17 (see Figure 3.11, left).

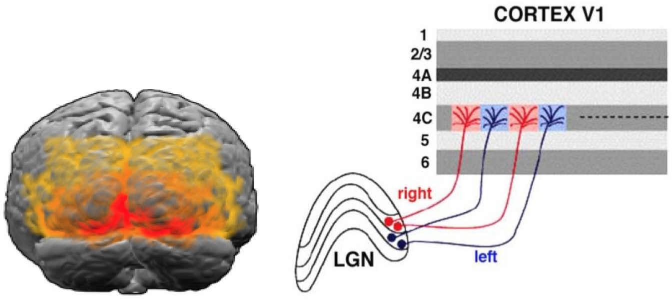


Figure 3.11: On the left the Brodmann area 17 (primary visual cortex) represented in red. Area 18 and 19 are shown respectively in orange and yellow. On the right the layered structure of V1. Most of the axons of LGN project to sublayer 4C. Source: [Cocci, 2014].

The visual cortex is composed by 6 layers in a retinotopic structure. Layer 4, which receives most visual input from the LGN, is further divided into 4 layers, labeled 4A, 4B, $4C\alpha$, and $4C\beta$. Sub-lamina $4C\alpha$ receives most magnocellular input from the LGN, while layer $4C\beta$ receives input from parvocellular pathways. The sublayer 4C is where most of the axons from the LGN arrive and where the concentration of oriented cells is higher as represented in Figure 3.11 (right).

From V1 projections go to the higher visual layers of the cortex as visual area V2, V3, V4 that correspond to Brodmann area 18 and 19 visualized in Figure 3.11 (left) and the mediotemporal (MT) layer [Wurtz and Kandel, 2000] as represented in Figure 3.12.

3.2.1 Simple cells in V1

The primary visual cortex processes the orientation of contours by means of the simple cells and other features of the visual signal (as estimation of motion direction) by means of complex cells. Every simple cell is characterized by its receptive field, classically defined as the domain of the retina to which the neuron is sensitive. The shape of the response of the cell in presence of a visual input is called receptive profile (RP) and can be reconstructed by electrophysiological recordings [Ringach, 2002]. In particular, Hubel and Wiesel first provided the char-

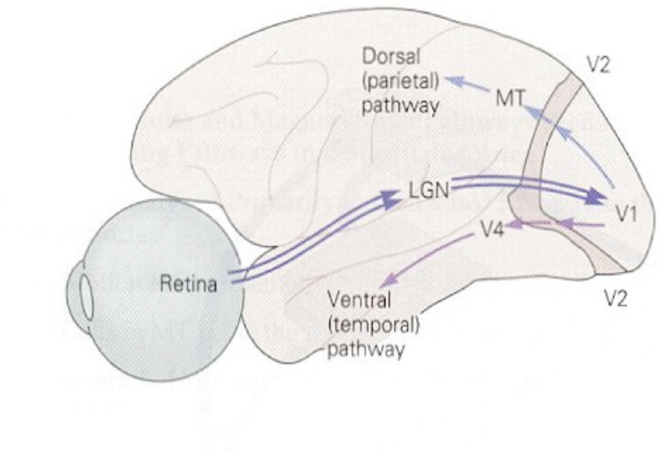


Figure 3.12: Diagram of the visual pathways. Adapted from [Kandel et al., 2000].

acterization of receptive fields in V1 based on their responses [Hubel and Wiesel, 1962].

Simple cells of V1 are sensitive to orientation and are strongly oriented. Most of the V1 simple cells are functionally involved in visual processing as orientation detectors. As was first noted by Koenderink [1984], the receptive field profiles of simple cells have a remarkable resemblance to Gaussian derivative kernels. Daugman proposed the use of Gabor filters in the modeling of the receptive fields of simple cells in the visual cortex of some mammals [Daugman, 1980]. Hence their RPs are interpreted as Gabor patches [Daugman, 1985, Jones and Palmer, 1987]. Precisely they are constituted by two coupled families of cells: an even and an odd-symmetric one (see Figure 3.13).

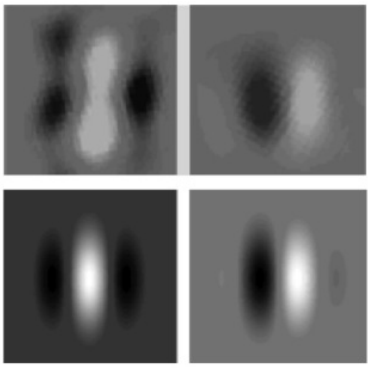


Figure 3.13: Receptive profile of a simple cell and its representation as a even-symmetric and odd-symmetric Gabor filters. Source: [Sarti and Citti, 2011].

The Gabor filter is a sinusoid modulated Gaussian (see Figure 3.14) and considering θ the orientation it has the following expression:

$$\psi(x, y) = e^{[-\frac{(\tilde{x}^2+\tilde{y}^2)}{2\sigma^2}+2\pi i\omega\tilde{y}]} \tag{3.1}$$

where:

$$\tilde{x} = xcos(\theta) + y\sin(\theta) \tag{3.2}$$

$$\tilde{y} = -xsin(\theta) + y\cos(\theta). \tag{3.3}$$

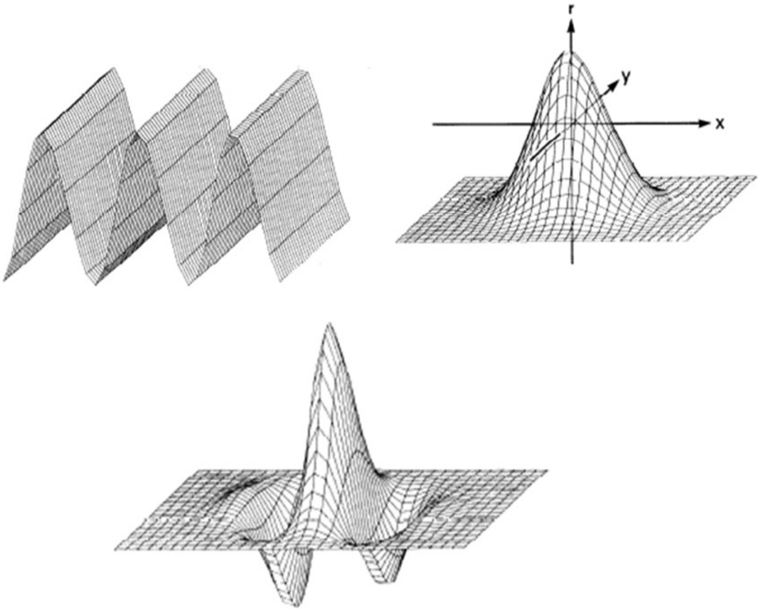


Figure 3.14: A Gabor filter (bottom) described as the product of a sinusoid (top left) and a Gaussian function (top right). Source: [Jones and Palmer, 1987].

The imaginary part of the Gabor filter models an odd-symmetric RP while the real part models an even one.

3.2.2 The functional architecture of V1

The functional architecture is defined as the spatial organization and as the connectivity between neurons in a cortical area. In V1 it is possible to identify the following structures:

- *the layered structure*: the cortex is formed by 6 horizontal layers and a number of sublayers (see Figure 3.11 (right));
- *the retinotopic structure*: what is near in the retina is near in the cortex. This is due to a topographic organization implying that the mapping from the retina to the cortex is preserved (see Figure 3.15). It is mathematically described by a logarithmic conformal mapping, as described in Section 3.2.2.1;
- *the hypercolumnar structure*: it organizes the cortical cells in columns corresponding to different parameters (like orientation, color) (see Figure 3.18). This is described in details in Section 3.2.2.2.

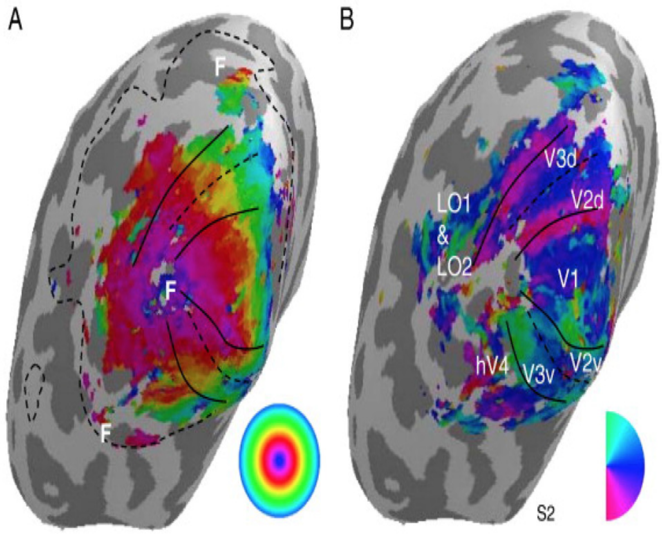


Figure 3.15: The retinotopic structure of the V1 recorded by fMRI. On the left is visualized the radial mapping while on the right is shown the polar angle (source: [Olman et al., 2010].

- *the horizontal connectivity*: it connects cells with the same orientation belonging to different hypercolumns, as described in Section 3.2.3.

3.2.2.1 The retino-cortical mapping

In this section, we describe the deformation that a signal on the retina obtains when represented on the cortex, estimating a model for the retino-cortical mapping.

The notion of retinotopy based on the anatomy of the visual cortex was first suggested by Polyak [1941], then Talbot and Marshall [1941] confirmed the hypothesis of the existence of a mathematical projection of the retina on the cortex using physiological methods. The study of Daniel and Whitteridge [1961] provided a source of quantitative data and a mathematical analysis of the retinotopic mapping has been presented in [Schwartz, 1977]. In [Schwartz, 1977] the retinotopic mapping of the striate cortex is mathematically described as a complex logarithmic mapping.

The principal quantitative measure of the structure of the cortical map is the *magnification factor* introduced by Daniel and Whitteridge [1961]. The representation of an image on the visual cortex is characterized by this factor: the portion of the image closer to the center of the field of view (the fovea) is strongly enlarged when mapped on the cortex [Tootell et al., 1988]. This factor decreases when the distance from the fovea (that is the retinal eccentricity) increases as represented in Figure 3.16.

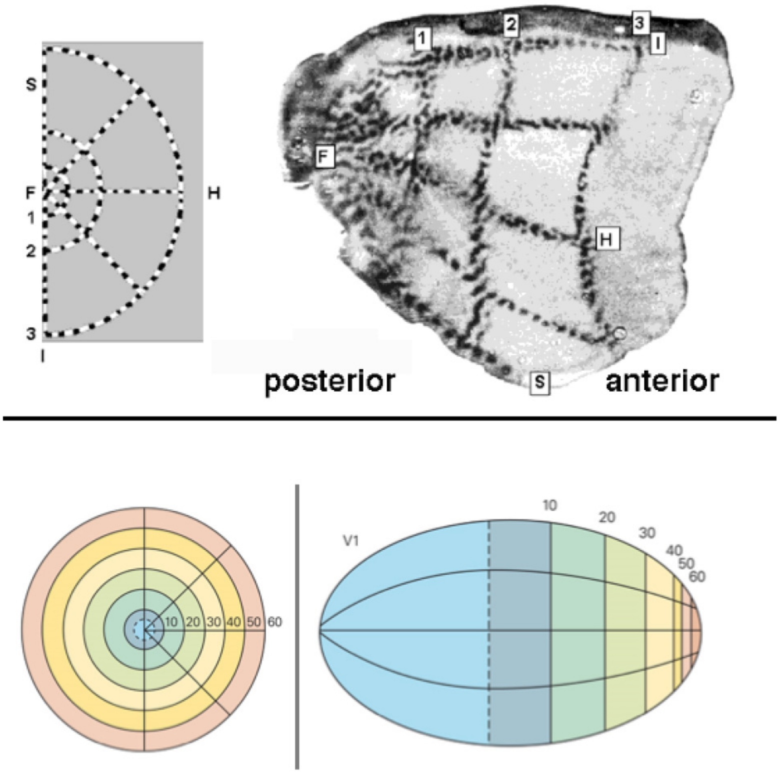


Figure 3.16: A visual stimulus (top left) and a portion of the cortex with corresponding activated regions (top right) [Tootell et al., 1988]. A schematic representation of the magnification factor property is visualized at the bottom.

Defined ρ the eccentricity, the magnification factor is determined as:

$$M_{a,k}(\rho) := \frac{k}{\rho + a} \tag{3.4}$$

where $(a, k) \in (0, 1]$ x $\mathbb{R}^+$ are two constants which determine the fit. In [Schwartz, 1980, 1977] this mapping was evaluated as a complex logarithmic function:

$$l_{a,k}(z) := klog(z + a), \quad (a, k) \in (0, 1]x\mathbb{R}^+ \quad and \quad z \in \mathbb{C} \tag{3.5}$$

and near the fovea the equation becomes:

$$l_{a,k}(z) \approx klog(a) + k\frac{z}{a}. \tag{3.6}$$

The motivation for considering a logarithmic function for the cortical mapping is based on noting that the magnitude of the cortical magnification factor is roughly inversely proportional to retinal eccentricity. In these equations, k is a normalization factor and a allows to discriminate between the linear (|z| « a) and the logarithmic (|z| » a) map. Thus the complex logarithm of a linear function of eccentricity provides a map from a linear foveal representation to a complex logaritmic surround (see Figure 3.17).

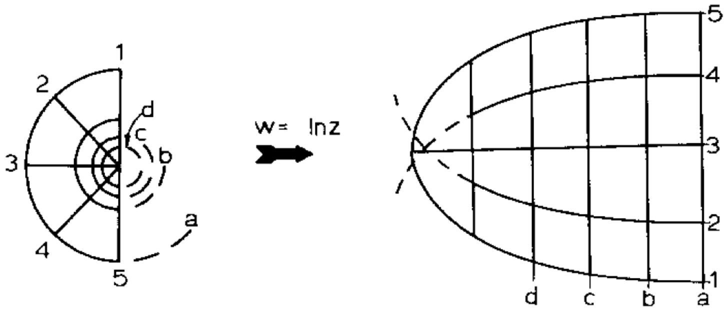


Figure 3.17: The retinotopic mapping under the logarithm function. Concentric circles and radial straight lines are mapped in equidistant cartesian grid on the cortex. Source: [Schwartz, 1977].

A comparison of Figure 3.17 with the experimental evidence displayed in Figure 3.16 suggests that this function provides a good model for the retino-cortical mapping.

In Montobbio [2016] it is shown a procedure designed in order to estimate suitable values for the constants a and k of the two-parameter logarithmic function $l_{a,k}$ as a model for the retino-cortical mapping.

3.2.2.2 The hypercolumnar structure

In the 70s Hubel and Wiesel discovered that the primary visual cortex is organized in the so called hypercolumnar structure [Hubel and Wiesel, 1962, 1977]. This means that for each retinal point (x, y) there is an entire set of cells each one sensitive to a specific orientation θ of the stimulus.

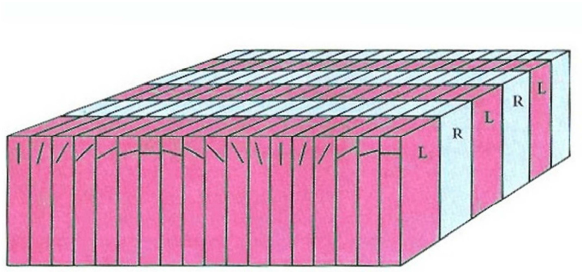


Figure 3.18: The classical Hubel and Wiesel cube scheme of V1. Cells belonging to the same column share similar receptive profile characteristics, the orientation hypercolumns are arranged tangentially to the cortical sheet. Source [Hubel, 1995].

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 $\mathbb{R}^2 x S^1$. At a certain scale and resolution, for each point of the retina (x, y) there exists a whole set of neurons in V1 maximally responding to every possible orientation θ. 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 tuned to the orientation given by the angle θ.

In other words, simple cells extract the orientation information at all locations and send a multi-orientation field to higher levels in the brain. Via the retinotopy, the retinal plane can be identified with the 2-dimensional plane $\mathbb{R}^2$. A visual stimuli I at the retinal point (x, y) activates the whole hypercolumnar structure over that point. All cells fire, but the cell with the same orientation of the stimulus is maximally activated, giving rise to orientation selectivity, as visualized in Figure 3.19.

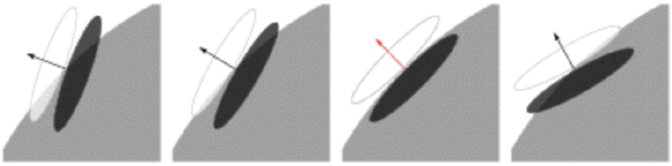


Figure 3.19: The maximal activity is observed for the simple cell sensitives to the direction of the boundary of the visual stimulus. Adapted from [Sarti et al., 2008].

In presence of a visual stimulus $I(x, y)$, the output of simple cells is computed as the integral of the receptive profile with the image:

$$h(x, y, \theta) = \int \psi_{x,y,\theta}(x', y')I(x', y')dx'dy'. \tag{3.7}$$

The cortex is equipped with a neural circuitry, called the intracortical circuitry, that is able to keep the direction of maximal response of the output of simple cells, achieved at a value $\bar{\theta}$. The lifted set is discrete and corresponds to the maximal activity of the output of simple cells. It selects the hypercolumns orientation of maximum output in response to a visual stimulus and to suppress all the others. The maximal activity is observed for the simple cell sensitive to the direction of the boundary of the visual stimulus.

In particular, in the hypercolumnar structure, it is possible to identify two types of connectivity between neurons. The *intracortical circuitry* is able to select the hypercolumns of orientation with maximum output in response to a visual stimulus. Cells with the same orientation belonging to different hypercolumns are connected by the *horizontal connectivity.*

The horizontal connections connect cells with the same orientation belonging to different hypercolumns. Correlation techniques have been used [Ts'o et al., 1986] to estimate the relation between connectivity and orientation of cells; recently techniques of optical imaging allowed to study the propagation of the neural signal via cortico-cortical connectivity [Bosking et al., 1997]. This connectivity allows to obtain the integration process, that is at the base of the formation of illusory contours

[Petitot and Tondut, 1999, Citti and Sarti, 2006].

3.2.2.3 The pinwheel structure

Optical imaging techniques were developed by Bonhoeffer et al. [1991] to study the layout of the orientation domains of the cortex. This in-vivo intrinsic-signal technique allowed to obtain the global map of orientations through the acquisition of activity from cells of the superficial layers of V1 (see Figure 3.20 (left)). It is possible to notice in the iso-orientation maps, that contain the areas that best respond to one orientation, the presence of singular points where all colors appear once. These points are arranged like spokes of a wheel and are called pinwheels see Figure 3.20 (right) [Petitot, 2003]. They found that the orientation centres from which the pinwheel-like organization of orientation preference originates are an important feature for organizing the representation of orientation in cortical area.

In Figure 3.20 two features of the orientation centres are shown: in all pinwheels each orientation appears once around the centre; the pinwheels exist in two forms, a clockwise (top right) and a counter-clockwise (top down)

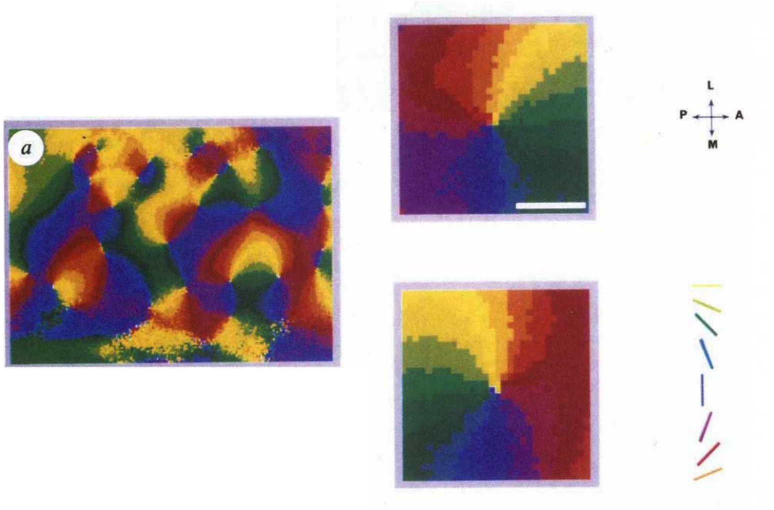


Figure 3.20: Left: colour-coded orientation preference map found by Bonhoeffer et al. [1991]. The preferred orientation for every location is coded according to the scheme shown on the right, where the yellow represents area responding best to a horizontal bar. Source: [Bonhoeffer et al., 1991].

Each orientation map can be obtained using a color scale where the color corresponds to the best orientation stimulus.

Orientation preference map of tree shrew's visual cortex is visualized in Figure 3.21 [Bosking et al., 1997].

The orientation preference maps contain both linear zones, which correspond to the orientation hypercolumns recorded with the electrodes, and pinwheel structure (see Figure 3.21).

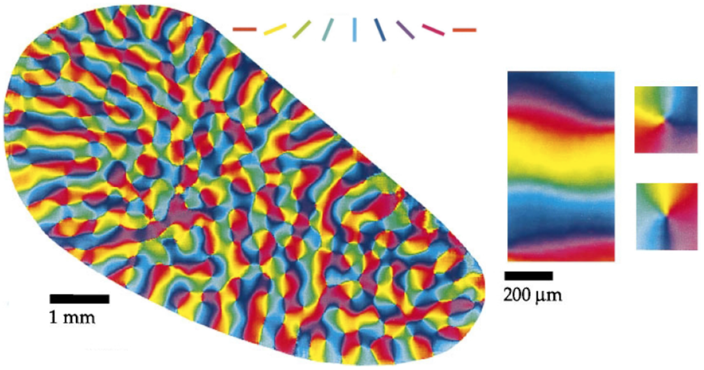


Figure 3.21: Left: orientation preference maps of tree shrew's visual cortex. Orientation preference of each location is color-coded. Right: the orientation preference maps contain both linear zones and pinwheel structure. Source: [Bosking et al., 1997].

Since conventional optical imaging is not capable of resolving the fine scale structure of the pinwheel centers, [Ohki et al., 2006] used the two-photon calcium imaging technique based on confocal microscopy to confirm that pinwheels exist as a real neurophysiological structure.

3.2.3 The horizontal connectivity

The 3D cortical structure is implemented in the 2D cortical layer as a pinwheel structure, which codes for position and orientations (see Figure 3.21 (left)).

From the neurophysiological point of view, there is experimental evidence of the existence of connectivity between simple cells of different hypercolumns. It is the so called long range horizontal connectivity, that is responsible for the cortico-cortical propagation of the neural activity between hypercolumns. These experiments revealed that the linked cells of different hypercolumns not only share the angle of tuning, but also the axis corresponding to the orientation is roughly the same. Bosking et al. [1997] clarified properties of horizontal connections on V1 of the tree shrew, measuring the cortico-cortical or horizontal connectivity by injecting a tracer (biocytin) in a simple cell and recording the trajectory of the tracer. In Figure 3.22 the propagation through the lateral connections is represented by black points. On the left, the bouton distribution shown over orientation preference maps after an injection into a site with a preferred orientation of 80 degree. The white symbols indicate the location of cells that took up the tracer. In black the boutons, found at sites with all orientation preferences near the injection site, but preferentially at sites with the same orientation preference as the injection site at longer distances. On the right, the results from an experiment in which an injection was made into a site with an orientation preference of 160 degree.

Bosking found that the propagation of the tracer is strongly directional and the direction of propagation coincides with the preferential direction of the activated cells. Moreover, he found a large variability of injections, which however have common stochastic properties as the direction of propagation. Combining the image of the propagation of a tracer through the lateral connections with the orientation maps obtained with optical imaging, they observed that nearby each neuron the connections are relatively isotropic but over larger distances they follow the orientation preferences.

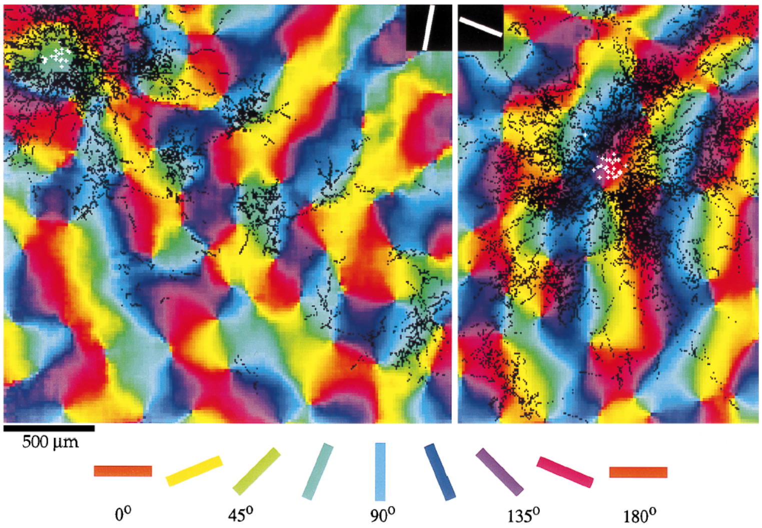


Figure 3.22: Left: bouton distribution shown over orientation preference maps after an injection into a site with a preferred orientation of 80 degree. Right: results after an injection into a site with an orientation preference of 160 degree. Source: [Bosking et al., 1997].

Angelucci et al. [2002] experimentally measured the lateral or horizontal connections of macaques, which is represented in Figure 3.23, showing a very isotropic morphology. Indeed, primates appear to have approximately isotropic horizontal connections (once ocular dominance is taken into account).

These experiments have shown that the propagation of a tracer is collinear to the preferred orientation of cells.

In the next Chapter we present a model for the functional structures described, showing that these are the basis of perceptual completion of contours.

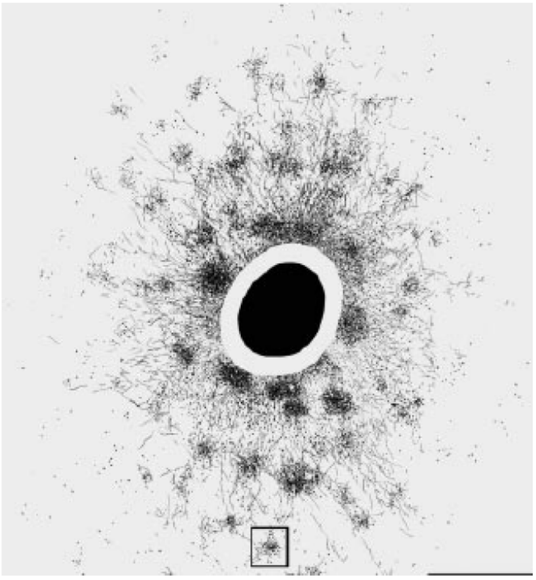


Figure 3.23: The isotropic connectivity map found by Angelucci et al. [2002] on macaques. The connectivity pattern is almost isotropic. Source: [Angelucci et al., 2002].

4 *Differential models of the visual cortex*

In this chapter we recall several models of the functional architecture of the visual cortex expressed in terms of differential geometry.

We focus in particular on the models of the visual cortex of Petitot and Tondut [1999] in the Heisenberg group and of Citti and Sarti [2006] in the Rototranslation group.

We will see how the functional organization of V1 implemented by long-range horizontal connections naturally induces a Sub-Riemannian structure within the Lie group of simple cells. In particular the position/orientation association fields are modelled with a family of horizontal integral curves in the $\mathbb{R}^2 \times S^1$ domain.

The propagation along the cortical connectivity is modelled by the propagation kernel of the structure, which are the fundamental solutions of Fokker Planck, Sub-Riemannian Laplacian and isotropic Laplacian equations.

It is well known that the fundamental solution of second order differential operators is expressed as the probability density of a suitable stochastic differential equation. This point of view will be considered here and the density kernel will be estimated with the efficient numerical technique of Markov Chain Monte Carlo methods (MCMC).

The classical mean field equation of Wilson and Cowan can be endowed with these geometric connectivity kernels. In [Sarti and Citti, 2015], instruments of spectral analysis have been applied to this modificated activity equation and they proved that the corresponding stable states coincide with the eigenvectors of the connectivity matrices, obtained by the density kernels. In this way the authors provide a model of grouping and figure ground segmentation, which will be further studied here.

4.1 The first differential models of V1

The term neurogeometry was introduced by Jean Petitot in Petitot and Tondut [1999] to indicate the intrinsic geometry immanent to neural connectivity, to be clearly distinguished by the geometry of the connectivity in the external ambient space. This term is at the base of a large class of models, which use instruments of differential geometry or group theory to model the behavior of the visual cortex starting from its functional architecture.

It is well known since the fundamental studies of Hubel and Wiesel [1962, 1977]) that V1 is one of the first physiological layers along the visual pathway to carry out geometrical measurements on the visual stimulus, decomposing it in a series of local feature components.

The first geometrical models of the functional architecture of the visual cortex in terms of differential geometry are due to Koenderink and van Doorn [1987] and Hoffman [1989]. The first author underlined the differential action of perceptual mechanisms, in particular with respect to jet spaces arising from linear filters, while the second author proposed to model the hypercolumnar structure defining fiber bundle structure and pointed out the central role of symmetries in perception expressing them in terms of Lie groups. In the fiber bundle the retinal plane (x, y) is the basis, while the fiber concides with the hypercolumnar variable θ. This fiber bundle is the mathematical structure ideally modelling both the retinotopic and hypercolumnar structure.

A variational approach to describe smooth edges was proposed by Mumford [1994], in terms of the elastica functional and stochastic path with random curvature at any point. His model produces a probability distributions in the space $\mathbb{R}^2 \times S^1$ of positions and orientations whose probability peaks follow elastica curves.

After that the studies mainly focused on the set of simple cells, responsible for the detection of position and orientation.

Williams and Jacobs [1997b] introduced a stochastic completion algorithm based on suitable probability kernel in the space of positions and orientations associated to the neural representation of images.

In the same 3D space, instruments of differential geometry and frame theory have been used by Zucker [2006] to propose models of boundary completion.

More recently Petitot and Tondut [1999] described the set of simple cells as a fiber bundle, relating the association fields of Field et al. [1993] with the contact geometry introduced by Hoffman [1989] and the elastica of Mumford [1994]. They identified the structure of this layer of cortical cells with the Heisenberg group, and performed contour completion in this structure minimizing a suitable Lagrangian functional. We will described their model in detail Section 4.1.1.

Then the problem of edge organization in images was addressed in terms of a stochastic process of the type of Mumford, introducing nonlinearities in order to take into account the role of curvature, by August and Zucker [2003, 2000]. Citti and Sarti [2006] and Sarti et al. [2008], proposed to describe the structure of the visual cortex as a Lie group with a Sub-Riemannian metric. We will describe their model in the Section 4.1.2.

In [Chossat and Faugeras, 2009] it is proposed that the cortex is modelled as a space of 2D symmetric tensor and the hypercolumns of orientations encode the structure tensors. The evolution of the activity in the cortex is governed by a Wilson-Cowan equation type [Chossat and Faugeras, 2009] operating under the appropriate mathematical space which takes into account the rotation and translation symmetries.

Finally we recall the work of Van Almsick et al. [2005], Franken et al. [2007], who proposed new models in the same Lie group and the results of Duits and Franken [2009, 2010a,b], Duits and Van Almsick [2008]. They defined suitable filters, inspired by the shape of cells, which transform any image, defined in R^2, into a function defined on $SE(2)$ or on an higher dimensional Lie group of features. In this way an invertible map, called *orientation score*, $W : L^2(R^2) \to L^2(SE(2))$ is derived from the set $L^2(R^2)$ of functions defined on R^2 to the set $L^2(SE(2))$ of functions defined on $SE(2)$. Enhancement of boundary, segmentation and completion are performed in $L^2(SE(2))$ and induced on the initial image I via the inverse of the map W.

4.1.1 The visual cortex modeled in the Heisenberg contact structure

The work of Petitot and Tondut in Petitot and Tondut [1999] can be considered as a problem of naturalizing phenomenological models, since they tried to justify the phenomenological model on the basis of the neurophysiological evidence.

They described the retinal structure with a plane R and called

M the cortical layer, they described the retinotopy by an isomorphism $q : R \to M$. They modelled the hypercolumn over each retinal position $(x, y) \in R$ as a full fibre of possible orientation u at (x, y), where $u \in R$ represents the direction of a boundary of an image mapped on the retina at the point (x, y) and so it is the angular coefficient of the tangent to the boundaries. Formally, if the boundary is described by a 2D curve $(x, y(x))$, then:

$$u = \frac{dy}{dx}. \tag{4.1}$$

In presence of a visual stimulus, all the hypercolumn over a retinal point (x, y) is activated and the simple cell sensible to the direction u has the maximal response. The retinal point (x, y) is lifted to the cortical point (x, y, u) and the whole curve is lifted to (x, y, u) in the 3-dimensional space $\mathbb{R}^3$. By condition (4.1), follows $dy = udx$ so that all the lifted curves lie in the kernel of the 1-form:

$$dy - udx = 0 \tag{4.2}$$

This 1-forms is the contact form which defines the Heisenberg contact structure. This model allows to lift only level lines or boundaries expressed in the form $y = y(x)$. In the next Section we will see how to overcome this limitation.

4.1.2 The visual cortex as the Rototranslation group

We will show that the visual cortex is naturally modelled as the Rototranslation group. This group, also known as the 2D Euclidean motion group $SE(2)$, is the group of rotations and translations.

4.1.2.1 The group law

In Section 3.2.2 we modeled the set of simple cells as a set of filters $\psi_{x,y,\theta}$, where $(x, y) \in \mathbb{R}$, represents the position on the retina and the orientation preference θ takes values in S^1. Hence this family of cells can be identified by the product space $\mathbb{R}^2 \times S^1$. A representation of this space is illustrated in Figure 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 θ of tuning. Given their retinotopic position (x_1, y_1) 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.

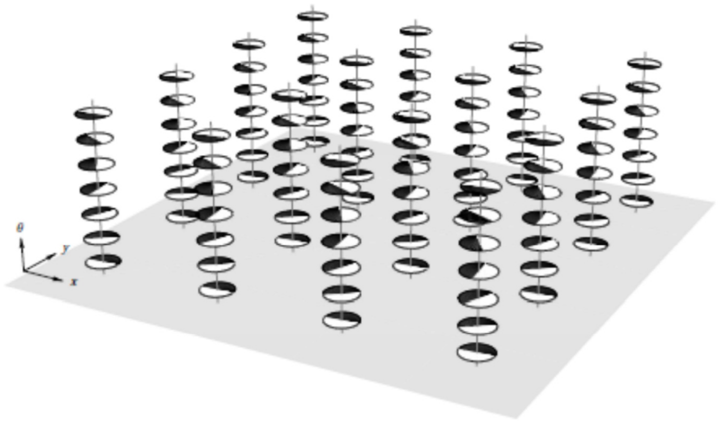


Figure 4.1: The visual cortex modelled as the roto-translation group, invariant under translations and rotations. Sorce: [Sarti and Citti, 2011].

It has been noted by Daugman [1980] and Bressloff and Cowan [2003] that this set of cells forms a group. Indeed, defined T_{x_1,y_1} the translation of the vector (x_1, y_1) and R_θ the rotation matrix of angle θ:

$$R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \tag{4.3}$$

a general element of the Rototranslation group is defined as the composition $A_{x_1,y_1,\theta} = T_{x_1,y_1} \circ R_\theta$ and its application to a point (x, y) brings to:

$$A_{x_1,y_1,\theta} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + R_{\theta_1} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}. \tag{4.4}$$

Proposition 1. *The set of all parameters* $\{g = (x, y, \theta) \in \mathbb{R}^2 \times S^1\}$ *forms a group with the operation induced by the composition of elements* $A_{x_1,y_1,\theta_1} \circ A_{x_2,y_2,\theta_2}$*:*

$$\begin{aligned} g_1 \circ g_2 = (x_1, y_1, \theta) +_R (x_2, y_2, \theta_2) = \\ = \left(\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + R_{\theta 1} \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \right)^T, \theta_1 + \theta_2 \right) \end{aligned} \tag{4.5}$$

Indeed the $+_R$ verifies the group operation axioms.

Definition 1. *The set* $\mathbb{R}^2 \times S^1$ *with the group law operation* $+_R$ *form a group called* $SE(2)$*.*

4.1.2.2 The lifting process

Due to the hypercolumnar structure, over each point (x, y) there is a whole ipercolumn of cells each with a different preferred orientation θ.

In presence of a visual stimulus I, characterized by edges or level lines, the whole hypercolumn fires, but the maximal response is provided by the cell whose characteristic orientation θ coincides with the angle $\theta = -\arctan(\partial_1 I, \partial_2 I)$ where $\theta \in [0, \pi]$. We call θ the orientation of the level line. This means that the vector field defined as:

$$X_\theta = \cos(\theta(x,y))\partial_1 + \sin(\theta(x,y))\partial_2 \tag{4.6}$$

is tangent to the level lines of I at (x, y).

Note that this representation, introduced in [Citti and Sarti, 2006], allows to remove the assumption that all level lines are expressed in the form $y = y(x)$, formulated in [Petitot and Tondut, 1999] (see Section 4.1.1).

This process associates to each retinal point (x, y) a cortical point identified by the variables $(x, y, \theta(x, y))$:

$$(x, y) \to (x, y, \theta). \tag{4.7}$$

In this way every two dimensional curve is lifted to a new curve in the 3D space, as shown in Figure 4.2. In blue it is represented a 2D curve and in red is shown its 3D cortical lifting in the Rototranslation group.

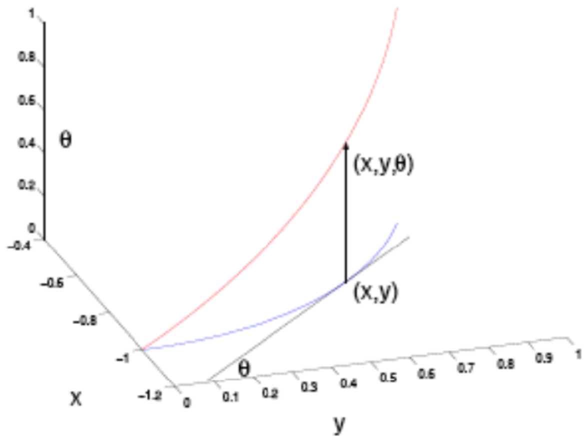


Figure 4.2: A contour represented by the blue curve is lifted into the Rototranslation group obtaining the red curve. Adapted from: [Sanguinetti et al., 2010].

It has been proved in Citti and Sarti [2006] that each lifted curve can be considered as integral curves of the vector fields:

$$\vec{X}_1 = (\cos\theta, \sin\theta, 0), \quad \vec{X}_2 = (0, 0, 1). \tag{4.1.8}$$

In particular the tangent vector of a lifted curve can not have component in the direction of the orthogonal vector $\vec{X}_3$:

$$\vec{X}_3 = (-\sin\theta, \cos\theta, 0) \tag{4.1.9}$$

As a consequence we will call admissible, the curves which are integral curves of these two vector fields with non vanishing coefficient in the direction X_1:

$$\dot{\gamma}(t) = (\dot{x}(t), \dot{y}(t), \dot{\theta}(t)) = \vec{X}_1(t) + k\vec{X}_2(t) \tag{4.1.10}$$

$$\dot{\gamma}(0) = (x_0, y_0, \theta_0). \tag{4.1.11}$$

It has been noted in Citti and Sarti [2006] that these curves, projected on the 2D cortical plane are a good model of the association fields (see Figure 4.3).

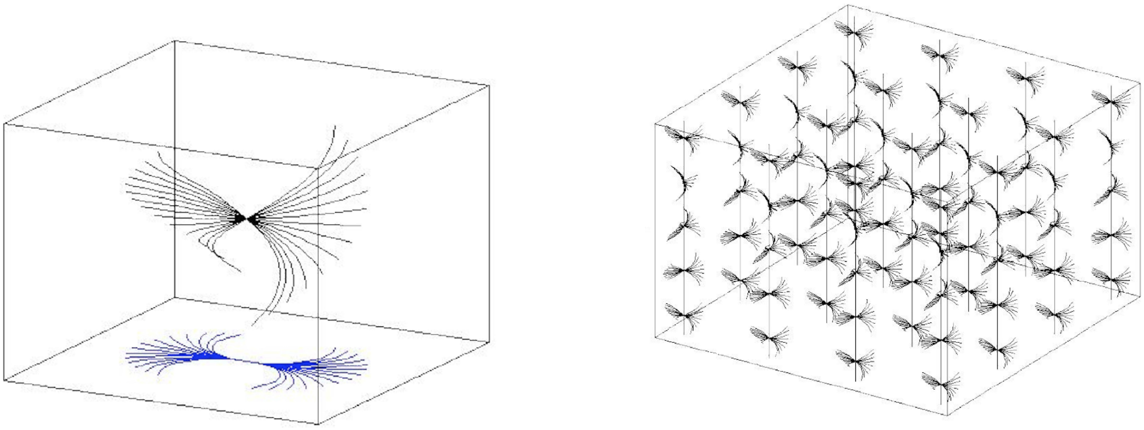


Figure 4.3: Left: the integral curves of the vector fields X_1 and X_2 in the (x, y, θ) space. In blue the projections of the integral curves on the xy plane. Right: the distribution of the integral curves, modeling the connectivity between points. Adapted from: [Sarti and Citti, 2011].

4.1.2.3 The Sub Riemannian structure

Generalizing Definition 1 considering the application to more than one feature, we will assume to have m vector fields in $R^n \times S^1$. The points of the space will be denoted $g = (x_1, \cdots, x_n, \theta)$. The vector fields will be denoted as:

$$X_1, \cdots X_m.$$

Clearly in the present setting $m = 2$ and $n = 2$. We call horizontal curves the integral curves of the vector fields $X_1, \cdots X_m$:

$$\dot{\gamma}(t) = \alpha_1\vec{X}_1(t) + \cdots + \alpha_m\vec{X}_m(t). \tag{4.1.12}$$

This condition defines a choice of planes, subset of the tangent plane at every point:

Definition 2. *We call horizontal plane at the point* $g = (x_1, \cdots, x_n, \theta)$ *the subspace of the tangent space, generated by* $X_1, \cdots, X_m$*:*

$$H = \{\alpha_1\vec{X}_1 + \cdots + \alpha_m\vec{X}_m : \alpha_1, \cdots, \alpha_m \in R\}$$

.

In $R^2 \times S^1$ with our choice of horizontal plane reduces to:

$$H = \{\alpha_1 \vec{X}_1 + \alpha_2 \vec{X}_2 : \alpha_1, \alpha_2 \in R\}$$

.

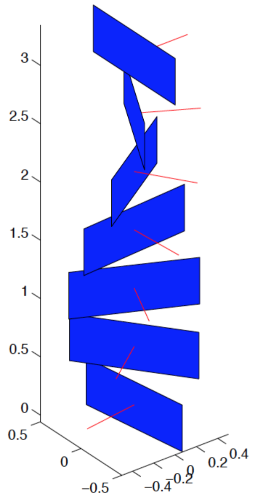


Figure 4.4: The contact planes at every point in $SE(2)$ and the orthogonal vector $\vec{X}_3$. Source: [Citti and Sarti, 2006].

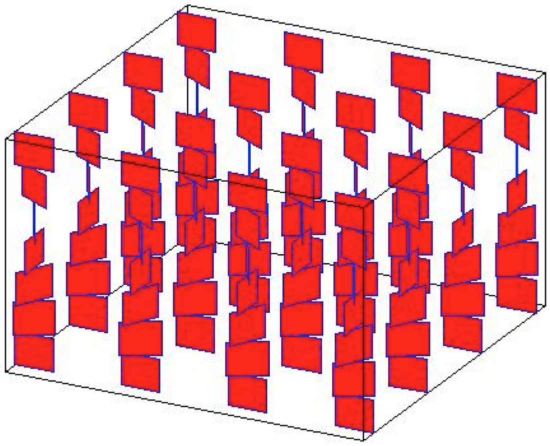


Figure 4.5: The horizontal tangent planes in each point of the rototranslation group is the span of the vectors $\vec{X}_1$, $\vec{X}_2$. Adapted from: [Sarti and Citti, 2015].

A norm can be defined on the elements of this plane:

$$||\alpha_1 \vec{X}_1 + \cdots + \alpha_m \vec{X}_m|| = \sqrt{\alpha_1^2 + \cdots + \alpha_m^2}. \qquad (4.1.13)$$

Consequently we can define the length of an admissible curve as the integral:

$$\int_0^1 ||\gamma'(t)||dt. \qquad (4.1.14)$$

These notions will lead to the definition of distance as the length of the shortest admissible path connecting two given points. However before giving this definition it is necessary to prove that couple of points can be indeed connected by admissible curve. This is a consequence of some properties of the horizontal planes. In this sense we will say that the horizontal planes (visualized in Figure 4.5) in each point of the Rototranslation group determine the differential structure of the space.

Let us now introduce first order differential operators with the same coefficients as the vector fields $\vec{X}_i$. In the specific 3D model we will have:

$$\begin{aligned} X_1 &= \cos(\theta)\partial_1 + \sin(\theta)\partial_2, \quad X_2 = \partial_2, \\ X_3 &= -\sin(\theta)\partial_1 + \cos(\theta)\partial_2. \end{aligned} \tag{4.1.15}$$

In general, if $\vec{X} = (v_1, \cdots, v_n)$ for suitable coefficients v_j we will call:

$$X = v_1\partial_1 + \cdots + v_n\partial_n. \tag{4.1.16}$$

We can define an operation of bracket between vector fields as following.

Definition 3. *Given two smooth vector fields X, Y we call bracket or commutator:*

$$[X, Y] = XY - YX. \tag{4.1.17}$$

It is important to note that $[Y, Z]$ is a fist order derivative from the Euclidean point of view, even though it is obtained as difference of second order derivative.

In the special 3D setting, a direct computation shows that:

$$X_3 = [X_2, X_1] \tag{4.1.18}$$

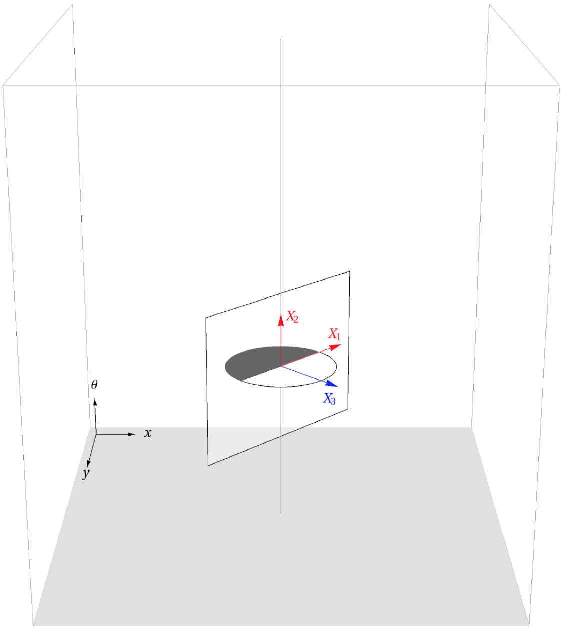


Figure 4.6: Schematic representation of a simple cell of V1 where the vectors $\vec{X}_1, \vec{X}_2, \vec{X}_3$ are indicated. Source: [Sanguinetti, 2011].

Note that we have now endowed the set of vector field with a new operation, which allows to define a Lie algebra.

Definition 4. *The Lie algebra generated by* $X_1, \cdots, X_m$ *is the set of all directional derivatives, represented as a linear combination of* $X_1, \cdots, X_m$ *and their commutators.*

In the 3D model, even if the Lie algebra has 2 generators it contains $X_3 = [X_2, X_1]$, so that it also contains the whole 3 dimensional tangent space at every point [Citti and Sarti, 2006] (see Figure 4.6).

We say that the Hörmander condition is satisfied if $X_1, \cdots, X_m$ and their commutators of any order span the Euclidean tangent space at every given point. In the present case X_1, X_2 and their commutator X_3 are linearly independent and span the tangent space to $\mathbb{R}^2 \times S^1$ at each point.

Hörmander condition is satisfied, then the connectivity condition, that is the possibility of connecting each couple of points with an admissible curve, also holds [Citti and Sarti, 2006] (see Figure 4.7).

Theorem 1. *(Chow theorem, see [Bellaïche, 1996]) If* $X_1, \cdots, X_m$ *satisfy the Hörmander condition at every point, for each couple of points* $g, \bar{g}$ *there exists an horizontal curve* γ *which connects them [Citti and Sarti, 2006].*

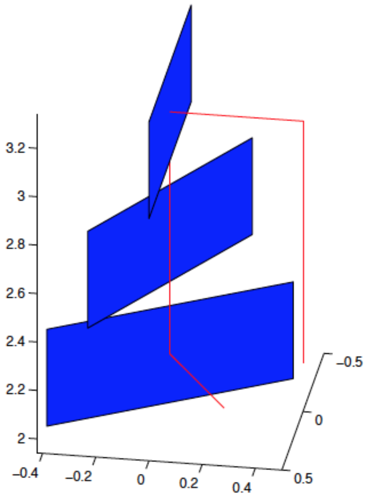


Figure 4.7: Piecewise constant integral curves of the structure. Any couple of points can be connected by a piecewise reguale path. Source: [Citti et al., 2015].

This Theorem 1 can be applied in $R^2 \times S^1$. As a consequence a distance between two points $(x, y, \theta), (\bar{x}, \bar{y}, \bar{\theta})$ can be defined as the length of the shortest path connecting them. The choice of a family of planes at every point which satisfies the Hörmander condition and the definition of a distance induced by the horizontal paths, defines a Sub-Riemannian structure on the space.

4.2 Stochastic models of cortical connectivity

4.2.1 Subelliptic differential operators and stochastic differential equations

As a consequence of the Hörmander condition, we can define second order operators in terms of the directional derivatives $X_1, \cdots, X_m$ instead of the standard partial derivatives. To this end we will define the subelliptic gradient as:

$$\nabla_R = (X_1, \cdots, X_m). \tag{4.2.1}$$

Under the assumption that the vector fields X_i are self adjoint, we will call subelliptic (or subriemannian) Laplacian:

$$\Delta_{SR} = X_1^2 + \cdots + X_m^2, \tag{4.2.2}$$

where X_1^2 denotes the second directional derivatives in the direction X_1. Analogously, we call subelliptic heat operator:

$$H_{SR} = \partial_t - \Delta_{SR}. \tag{4.2.3}$$

More generally, if $\alpha_1, \cdots, \alpha_m,, \sigma_1, \cdots, \sigma_m$ are constants we call Fokker Plank equation:

$$\begin{aligned} L_{SR,n} = {} & \partial_t + \alpha_1 X_1 + \cdots + \alpha_m X_m \\ & - \sigma_1^2 X_{11} - \cdots - \sigma_m^2 X_{mm}. \end{aligned} \tag{4.2.4}$$

If the vector fields X_i satisfy the Hörmander condition, then each of these operator have a smooth fundamental solution. This condition is strictly related to the connectivity result, since the fundamental solution with pole at a point g is smooth and positive on the set of points which can be connected to g with an admissible path.

It is well known that the fundamental solution can be obtained as a solution of the Langevin equation. The m-dimensional Langeving equation has the general form:

$$\begin{aligned} \gamma' = {} & \alpha_1 \vec{X}_1 + \cdots + \alpha_m \vec{X}_m \\ & + N(0, \sigma_1^2)\vec{X}_1 + \cdots + N(0, \sigma_m^2)\vec{X}_m \end{aligned} \tag{4.2.5}$$

where $N(0, {\sigma_i}^2)$ is a normally distributed variable with zero mean and variance equal to σ_i^2.

We refer to Oksendal [2013] for the following classical theorem:

Theorem 2. *If p is the transition probability of the stochastic process defined by the Langevin equation, from the initial state g_0 at time 0, then p is the fundamental solution of the Fokker Planck equation with pole at g_0.*

4.2.2 Overview of stochastical cortical models present in literature

A stochastical approach to continuation of boundaries was proposed by Mumford [1994]. In his model edges of natural images are represented as planar curves whose the curvature is expressed as white noise and the direction function as a one dimensional Brownian motion. In this way contours can be defined by the stochastic process:

$$\begin{cases} dx = cos(\theta)dt \\ dy = sin(\theta)dt \\ d\theta = \sigma dW(t) \end{cases} \tag{4.2.6}$$

where σ defines the amount of deviation of the particle from the straight path. When $\sigma = 0$ particles never deviate from straight paths, when $\sigma \to \infty$ the motion is completely random. In Figure 4.8 examples of different random paths (8, 16, 32) resulting from simulating the process in the system (4.2.6) for two different values of σ. When σ is higher the deviation of the particle from the straight path is more evident.

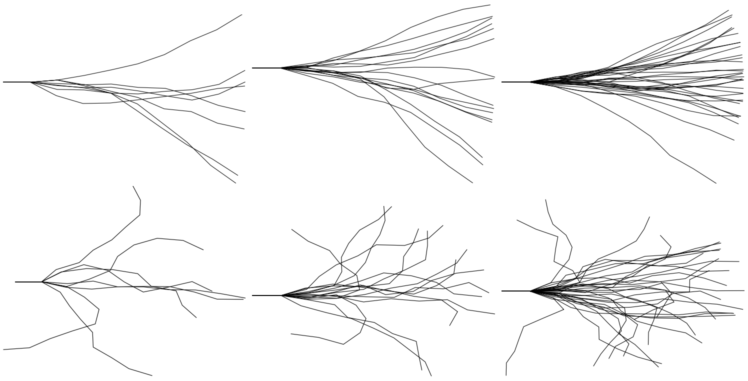


Figure 4.8: Examples of 8 (left), 16 (middle), 32 (right) random paths resulting from simulating the process in the system (4.2.6) for $\sigma = 0.1$ (top) and $\sigma = 0.5$ (bottom).

Called p the probability of transition $p(x, y, t)$ of the random paths, by Theorem 2 the deterministic partial differential equation modeling the evolution in the time of $p(x, y, t)$ is the Fokker-Planck equation:

$$\partial_t p = -cos(\theta)\frac{\partial}{\partial x}p - sin(\theta)\frac{\partial}{\partial y}p + \frac{\sigma 2}{2}\frac{\partial^2}{\partial \theta^2}p. \tag{4.2.7}$$

This is a probability distribution in the space $\mathbb{R}^2 \times S^1$ of positions $(x, y) \in \mathbb{R}^2$ and orientation $\theta \in S^1$ whose probability peaks follow elastica curves.

Following Mumford's ideas that contour continuation can be modelled by stochastic process Williams and Jacobs [1997b,a] used the stochastic processes to implement a mechanism of stochastic completion. In their approach of stochastic completion fields [Williams and Jacobs, 1997b,a, Thornber and Williams, 1996] they computed the likelihood that a completion joining two contour fragments passes through any given position and orientation within the image plane. This likelihood represents a measure of the saliency of a possible contour.

They also provided a stable finite difference scheme for solving the underlying Fokker-Planck equation (4.2.7) identified by Mumford. In particular, in Williams and Jacobs [1997b] were computed by large kernel convolution with filter generated by Monte Carlo simulation. In Figure 4.9 an example of the stochastic completion field applied to a Kanizsa example of illusory contours, the Kanizsa triangle.

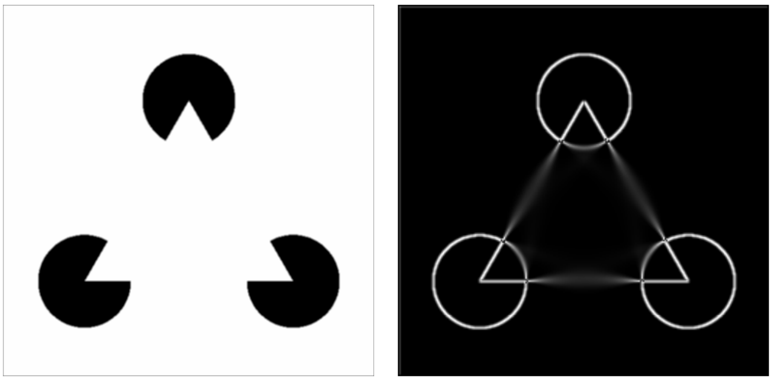


Figure 4.9: The Kanizsa triangle (left) and the stochastic completion fields applied to it (right). Source: [Williams and Jacobs, 1997a].

A generalization of the stochastic process of by Mumford, was proposed by August and Zucker [2003, 2000], who took in consideration also curvature, and will be described in detail in the next section.

4.2.3 Connectivity kernels as models of cortical connectivity

From the neurophysiological point of view, there is experimental evidence of the existence of connectivity between simple cells belonging to different hypercolumns. Hence connectivity can be summarized as preferentially linking neurons with co-circularly aligned receptive fields.

The propagation along the connectivity can be modeled as the stochastic counter part of the deterministic curves defined in equation (4.1.10) for the description of the output of simple cells. If we assume a deterministic component in direction X_1 (which describes the long range connectivity) and stochastic component along X_2 (the direction of intracolumnar connectivity), the equation can be written as follows:

$$(x', y', \theta') = \vec{X}_1 + N(0, \sigma^2)\vec{X}_2 \tag{4.2.8}$$

where $N(0, \sigma^2)$ is a normally distributed variable with zero mean and variance equal to σ^2. By Theorem 2 the probability density of this process, denoted by p, is the solution of the time dependent Fokker Planck operator:

$$L_{SR} = \partial_t - X_1 - \sigma^2 X_{22}. \tag{4.2.9}$$

In Cocci [2014] it has been proposed to integrate this kernel in time, in order to obtain a model of cortical connectivity. Integrating in time the density p_1, we obtain a time independent kernel:

$$\Gamma_1(x, y, \theta) = \int_0^{+\infty} p_1(x, y, \theta, t)dt \tag{4.2.10}$$

which is the fundamental solution of the Fokker Planck operator

$$FP_{SR} = X_1 + \sigma^2 X_{22}. \tag{4.2.11}$$

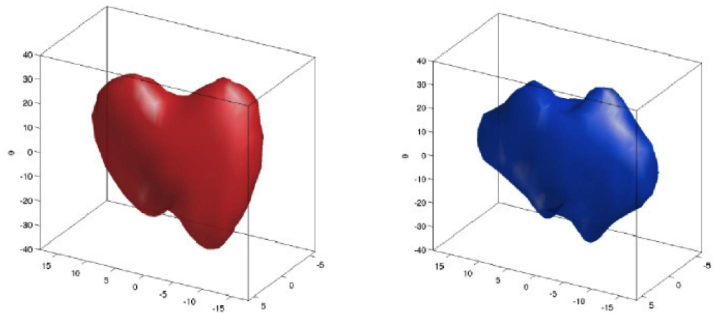
The kernel Γ_1 is strongly biased in direction X_1 and not symmetric; a new symmetric kernel can be obtained as following:

$$\begin{aligned} \omega_1((x, y, \theta), (x', y', \theta')) = {} & \frac{1}{2}\Gamma_1((x, y, \theta), (x', y', \theta')) \\ & + \frac{1}{2}\Gamma_1((x', y', \theta'), (x, y, \theta). \end{aligned} \tag{4.2.12}$$

In Figure 4.10 (a) it is visualized an isosurface of the simmetrized kernel ω_1, showing its typical twisted butterfly shape. The kernel ω_1 has been proposed in Sanguinetti et al. [2008] as a model of the statistical distribution of edge co-occurrence in natural images, as described in Sanguinetti et al. [2008]. The similarity between the two is proved both at a qualitative and at a quantitative level (see [Sanguinetti et al., 2010]) (see also Figure 4.10 (a) and (b)).

If we assume that intracolumnar and long range connections have comparable strength, the stochastic equation equation (4.2.8) reduces to:

$$(x', y', \theta') = N(0, \sigma_1^2)\vec{X}_1 + N(0, \sigma_2^2)\vec{X}_2 \tag{4.2.13}$$



(a) (b)

Figure 4.10: An isosurface of the connectivity kernel ω_1 obtained by symmetrization of the Fokker Planck fundamental solution equation (4.2.10) (a). The distribution of co-occurrence of edges in natural images (from [Sanguinetti et al., 2010]) (b).

where $N(0, {\sigma_i}^2)$ are normally distributed variables with zero mean and variance equal to σ_i^2. In this case the speed of propagation in directions X_1 and X_2 is comparable. The associated probability density is the fundamental solution of the Sub-Riemannian Heat equation [Jerison and Sánchez-Calle, 1986]. The integral in time of this probability density:

$$\Gamma_2(x, y, \theta) = \int_0^{+\infty} p_2(x, y, \theta, t)dt \qquad (4.2.14)$$

is the fundamental solution of the Sub-Riemannian Laplacian, defined in (4.2.2).

It is a symmetric kernel, so that we do not need to symmetrize it and we use it as a model of the connectivity kernel:

$$\omega_2((x, y, \theta), (x', y', \theta')) = \Gamma_2((x, y, \theta), (x', y', \theta')). \qquad (4.2.15)$$

In Figure 4.11 (a) it is shown an isosurface of the connectivity kernel ω_2.

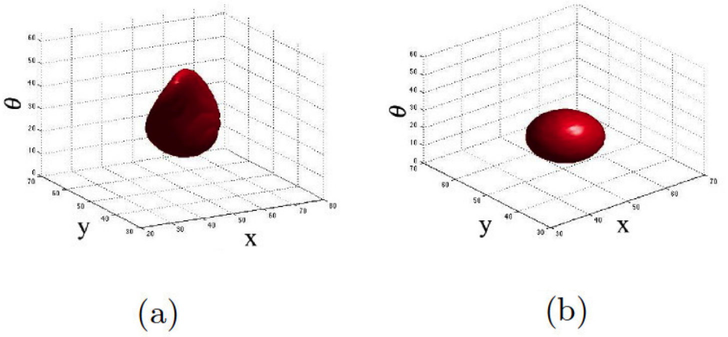
Let us finally note that the kernel introduced here have a strongly anysotropic behavior, if compared with the standard Riemannian kernel, fundamental solution of the operator:

$$L = \sigma_1^2 X_{11} + \sigma_2^2 X_{22} + \sigma_3^2 X_{33}.$$

As it is well known this kernel is associated to an isotropic stochastic equation:

$$(x', y', \theta') = N(0, \sigma_1^2)\vec{X}_1 + N(0, \sigma_2^2)\vec{X}_2 + N(0, \sigma_3^2)\vec{X}_3, \qquad (4.2.16)$$

where $N(0, {\sigma_i}^2)$ are normally distributed variables with zero mean and variance equal to σ_i^2. One of its level sets is represented in Figure 4.11 (b).



(a) (b)

Figure 4.11: Isosurface of the connectivity kernel ω_2 obtained from the fundamental solution Γ_2 of the Sub-Riemannian Laplacian equation equation (4.2.14) (a). An isosurface of the fundamental solution of the isotropic Laplacian equation (b).

4.3 Our contribution to the model

Even though the stochastic approach was already applied to the visual cortex, here we provide a new contribution to the model, clarifying how it is neurally implemented in the cortex, extending it in high generality and providing a solid numerical implementation.

4.3.1 Propagation, stochastic paths and connectivity kernels

The propagation of each visual signal in the cortical architecture will be described as a realization of stochastic sample functions. In particular if a tracer is injected in the cortex, each single injection can be described with stochastic paths. The probability density obtained as a combination of Fokker Plank is an integration of stochastic paths and describes the connectivity kernel.

4.3.2 General time independent kernels

The technique we have presented here is very general and can be applied to obtain fundamental solutions of general time independent operators.

If $X_1, \cdots X_m$ are differential operators in $R^n \times S^1$, $\alpha_1, \cdots, \alpha_m$ $\sigma_1, \cdots, \sigma_m$ are constants, we call:

$$\begin{aligned} FP_{SR,n} = {} & \alpha_1 X_1 + \cdots + \alpha_m X_m \\ & - \sigma_1^2 X_{11} - \cdots - \sigma_m^2 X_{mm}. \end{aligned} \tag{4.3.1}$$

the time independent Fokker Planck operator associated to the time dependent operator $L_{SR,n}$ defined in (4.2.4).

Proposition 2. *Let us denote* $g = (x_1, x_n, \theta)$ *the points of* $R^n \times S^1$. *If* p *is the fundamental solution of the operator* L_{SR} *defined in* (4.2.4) *and:*

$$\Gamma(g, g_0) = \int_0^{+\infty} p(g, g_0, t)dt,$$

then Γ *is the fundamental solution of the operator* (4.3.1).

4.3.3 Numerical approximations of the kernels

In this section we numerically approximate the connectivity kernels ω_i, defined in Section 4.2, with a technique partially inspired by [Williams and Jacobs, 1997b] .
We obtain the discrete fundamental solution Γ_1 of equation (4.2.10) by developing random paths from the numerical solution of the system (4.2.8), that can be approximated by:

$$\begin{cases} x_{s+\Delta s} - x_s = \Delta s \cos(\theta) \\ y_{s+\Delta s} - y_s = \Delta s \sin(\theta), \quad s \in 0, ..., H \\ \theta_{s+\Delta s} - \theta_s = \Delta s N(\sigma, 0) \end{cases} \tag{4.3.2}$$

where H is the number of steps of the random path and $N(\sigma, 0)$ is a generator of numbers taken from a normal distribution with mean 0 and variance σ. In that way, the kernel is numerically estimated with Markov Chain Monte Carlo methods (MCMC) [Robert and Casella, 2013]. Various realizations n of the stochastic path will be given solving this finite difference equation n times; the estimated kernel is obtained averaging their passages over discrete volume elements, as described in detail in Higham [2001], Sarti and Citti [2015]. In particular, we first fix a discretization step $\Delta s = 1$ without loss of generality; then we simulate n several discrete-times random paths, assigning a value between 0 and 1 corresponding to the number of paths that passed through it, divided by n. This provides a distribution over the cells that, for a large value of n, gives a discrete approximation of the connectivity kernel [Cocci et al., 2015]. Proceeding with the same methodology the numerical evaluation of fundamental solution Γ_2 of the hypoelliptic Laplacian (equation (4.2.14)) is obtained and the system (4.2.13) discretized:

$$\begin{cases} x_{s+\Delta s} - x_s = \Delta s N(\sigma_1, 0) \cos(\theta) \\ y_{s+\Delta s} - y_s = \Delta s N(\sigma_1, 0) \sin(\theta), \quad s \in 0, ..., H \\ \theta_{s+\Delta s} - \theta_s = \Delta s N(\sigma_3, 0) \end{cases} \tag{4.3.3}$$

where and σ_3 is the variance in the θ direction. The kernel represented in Figure 4.11 (a) is obtained by the numerical integration of that system and averaging as before the resulting paths.

Finally, the same technique can be potentially adapted to study the fundamental solution of general operators of the form (4.3.1).

4.4 The cortical activity equation

The most classical equation describing the cortical activity is the classical mean field equation of Ermentrout and Cowan [Ermentrout and Cowan, 1980] and Bressloff and Cowan [Bressloff et al., 2002, Bressloff and Cowan, 2003]. This equation describes the evolution of the cortical activity and depends on a connectivity kernel. In Sarti and Citti [2015] the relation between the stable states of these equation and perceptual units of the input has been established. The discrete output h of the simple cells, selects in the cortical space (x, y, θ) the set of active cells and the cortical connectivity, restricted on this set, defines a neural affinity matrix. The eigenvectors of this matrix describe the stationary states of the mean field equation hence the emergent perceptual units. The system will tend to the eigenvector associated to the highest eigenvalue, which corresponds to the most important object in that scene. Mathematically the approach is strongly linked to spectral analysis techniques for locality-preserving embeddings of large data sets [Coifman and Lafon, 2006, Belkin and Niyogi, 2003, Roweis and Saul, 2000], for data segregation and partitioning [Perona and Freeman, 1998, Meila and Shi, 2001, Shi and Malik, 2000], grouping process in real images [Weiss, 1999].

We have described in Section 4.1 that in presence of a visual stimulus cells aligned to its boundary give the maximal response. We will assume that a discrete number of cells N are maximally activated and we will denote them (x_i, y_i, θ_i) for $i = 1, ..., N$.

In Figure 4.12 we show as an example the cells responding to a Kanizsa figure, represented with their Gabor-like receptive profiles.

Following Sarti and Citti [2015] the cortical connectivity is restricted to this discrete set and reduces to a matrix A:

$$A_{i,j} = \omega((x_i, y_i, \theta_i), (x_j, y_j, \theta_j)). \tag{4.4.1}$$

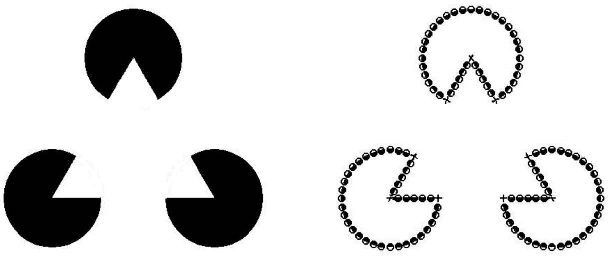


Figure 4.12: The Kanizsa triangle (left) and the maximally responding odd filters (right).

In this discrete setting the mean field equation for the cortical activity reduces to:

$$\frac{du}{dt} = -\lambda u(i) + s\Big(\sum_{j=1}^{N} A(i,j)u(j)\Big) \tag{4.4.2}$$

where s is a sigmoidal function and λ is a physiological parameter. The solution tends to its stationary states, which are the eigenvectors of the associated linearized equation:

$$\sum_{j=1}^{N} A_{i,j}u_j = \frac{\lambda}{s'(0)}u_i \tag{4.4.3}$$

Hence these are the emergent states of the cortical activity, that individuate the coherent perceptual unit in the scene and allow to segment it. This is why we will assign to the eigenvalues of the affinity matrix the meaning of a saliency index of the objects. Since we have defined three different kernels different affinity matrices will be defined. However all kernels are real and symmetric, so that the matrix A is a real symmetric matrix $A_{i,j} = A_{j,i}$. Their eigenvalues are real and the highest eigenvalue is defined. The associated principal eigenvectors emerge as symmetry breaking of the stationary solutions of mean fields equations and they pop up abruptly as emergent solutions. The first eigenvalue will correspond to the most salient object in the image.

4.4.1 Individuation of perceptual units

Since the three different kernels described in Section 4.2 assign different role to different direction of connectivity, the different affinity matrices and their spectrum will reflect these different behavior. Consequently the resulting data set partitioning will be stronger in the straight direction using the Fokker Planck ω_1 kernel, or will allow rotation using the ω_2 kernel (see also [Cocci et al., 2015] for a deeper analysis). Using the kernel ω_3

we expect an equal grouping capability in the collinear direction and in the ladder direction.

In Figure 4.13 we visualize the affinity matrix of the image presented in Figure 4.12. It is a square matrix with dimensions NxN, where *N* is exactly the number of active patches. It represents the affinity of each patch with respect to all the others.

The structure of the affinity matrix is composed by blocks and the principal ones corresponds to coherent objects. On the right we visualize the complete set of eigenvalues in a graph: (eigenvalue number, eigenvalues).

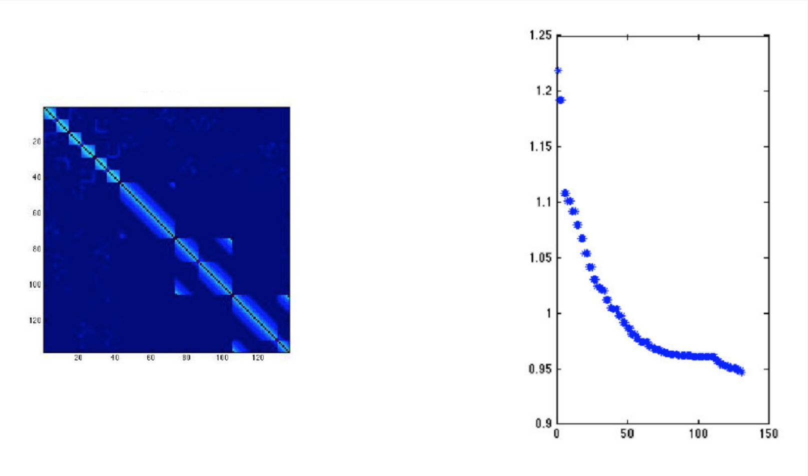


Figure 4.13: On the left it is visualized the affinity matrix of the image presented in Figure 4.12, that contains informations about the affinity of an active patch with respect to all the others. On the right the set of its sorted eigenvalues.

The first eigenvector can be recognized as the emergent perceptual unit, but here we also study the role of the other eigenvectors. They do not describe an ordered sequence of figures with different rank. However, their presence is important, above all when two eigenvalues have similar values. In this case, small deformation of the stimulus can induce a change in the order of the eigenvalues and produce a sudden emergence of the correspondent eigenvector with an abroupt change in the perceived image.

This is in good agreement with the perceptual characteristics of salient figures of temporal and spatial discontinuity, since they pop up abruptly from the background, while the background is perceived as indifferentiated [Merleau-Ponty and Smith, 1996]. Spectral approaches give reason to the discontinuous character of figure-ground articulation better than continuous models, who instead introduce a graduality in the perception of figure and background [Lorenceau and Alais, 2001].
To find the remaining objects in the image, the process is then repeated on the vector space orthogonal to *p*, the second and

the following eigenvectors can be found, until the associated eigenvalue is sufficiently small. In this way only n eigenvectors are selected, with $n < N$, this procedure reduces the dimensionality of the description. This procedure neurally reinterprets the process introduced by Perona and Freeman in Perona and Freeman [1998].

5 Quantitative kernel validation

In this chapter we will compare kernels previously described with neurophysiological data of horizontal connectivity [Angelucci et al., 2002, Bosking et al., 1997]. We also perform a quantitative validation of the kernel considering the experiment of Gilbert et al. [1996], showing the link between the connectivity kernel and cell's response. This work has been published in Favali et al. [2016b].

5.1 Validation of the model of connectivity

As described in Favali et al. [2016b], we will introduce here various substantial differences from the techniques in literature in the definition of a mathematical model of figure-ground segmentation. The Fokker Planck and the Laplacian kernel in the motion group are already largely used for the description of the connectivity, since they qualitatively fit the experimental data [Sarti and Citti, 2015]. We perform a quantitative fitting between the computed kernels and the experimental ones, in order to validate the model. We show that the cortical architecture is a realization of stochastic sample functions and how through this realization we can construct the connectivity kernel. We make a comparison between the fundamental solution of the Fokker Planck equation with experimental data of Bosking et al. [1997], Ben-Shahar and Zucker [2004], and Gilbert et al. [1996], showing how the stochastic paths are implemented in the neural network. In particular, we consider the distribution of a tracer through lateral connection modeling each injection with stochastic paths. The bouton distributions are realizations of a stochastic process, in particular of a random walk in $\mathbb{R}^2 x S^1$ space. We will show how the probability density obtained as a combination of Fokker Plank is an integration of stochastic paths. Moreover we propose to use also the Subelliptic Laplacian kernel, in order to account for the variability of connectivity patterns.

5.1.1 Stochastic paths and cortical connectivity

In this section, we describe the cortical architecture as a realization of stochastic sample functions, in particular we will see how the connectivity is associated to random paths. We will show that the position of presynaptic boutons in the images of Bosking et al. [1997] can be seen as the realization of stochastic paths via the anatomy. Every single random walk starts from the injection site of a tracer and gives the position of a set of boutons, as visualized in Figure 5.2 (a). Finally, the probability density, that is described as a kernel, is obtained as the integration of all the random paths. From the neural point of view this integration can be interpreted as the action of a columnar population and provides an estimation of the density of the boutons.

Let's consider an hypercolumn of the ice cube of visual cortex, it is composed by approximately 100 neurons. In the connectivity map in Figure 5.2 (a) we notice the presence of an average of 6 boutons. In this way, the number of possible connections in the visual cortex is 100^6 and in our model we use a number of paths compatible with this data. Now we make a comparison between the connectivity kernel previously defined and the experiments of Bosking et al. [1997] and Ben-Shahar and Zucker [2004].

In Figure 5.1 we can see the results of [Ben-Shahar and Zucker, 2004]. On the left the mean and standard deviation of the distribution of long-range connections of 7 injection sites considering the data of [Bosking et al., 1997], in the middle the expected median distribution for 7 cells from the curve model described in Ben-Shahar and Zucker [2004]. They noticed that the standard deviation is nonmonotonic, finding two local minima at approximately +30 and -30 degrees. Confirming their results, we show that our model implies a nonmonotonically changing variance as the orientation difference increases. In particular, on the right of Figure 5.1 it is visualized the mean and the standard deviation of 7 random paths, at a fixed orientation. We notice the presence of the nonmonotonicity of the standard deviation and that the two local minima at almost 30 degrees are preserved.

Moreover the fact that the mean and the variance of the model are similar to the experimental data suggests that the choice of the normal distribution allows to find physiological values. For these reasons, the connectivity represents the anatomical implementation of random paths.

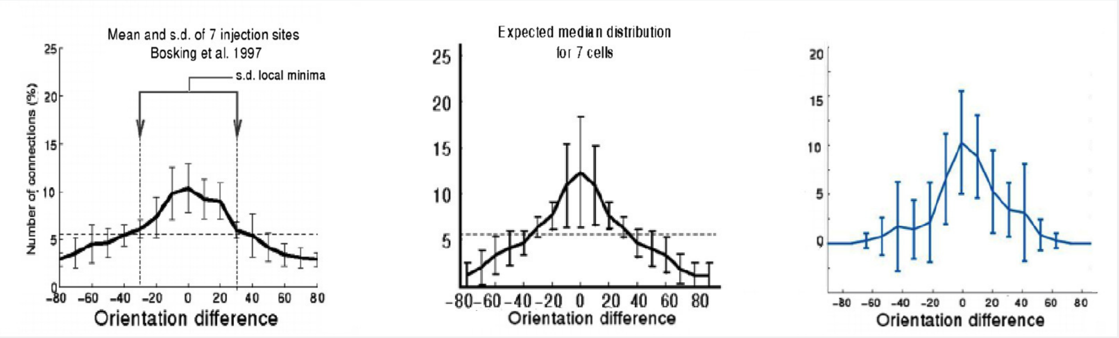


Figure 5.1: From left to right: the mean connection distribution computed in Ben-Shahar and Zucker [2004] from Bosking et al. [1997], the expected median distribution for seven cells and the results of our model considering 7 different random paths. It is evident that also in our model the standard deviation is nonmonotonic with two local minima at +30 and -30 degrees.

We will now study in which extent kernels ω_i, $i = 1, 2$ are models of connectivity. The kernel ω_3 is used for comparison and to show that an uniform Euclidean kernel does not capture the anysotropic structure of the cortex. Random paths that we compute through MCM are implemented in the functional architectures in terms of horizontal connectivity of a single cell. On the other hand the connectivity of an entire population of cells corresponds to the set of all single cells connectivities, then to the Fokker Planck fundamental solution.

A first qualitative comparison between the kernels ω_1, ω_2 and the connectivity pattern has been provided in Sarti and Citti [2015]. Here we follow the same framework, but we propose a more accurate, quantitative comparison.

As described in 3.2.2.3, the 3D cortical structure is implemented in the 2D cortical layer as a pinwheel structure (see Figure 5.2 (b)). The pinwheel structure has a large variability from one subject to one other, but within each species common statistical properties have been obtained. Cortico-cortical connectivity has been measured by Bosking et al. [1997] by injecting a tracer in a simple cell and recording the trajectory of the tracer, represented by black points in Figure 5.2 (a). Bosking found a large variability of injections, which however have common stochastic properties as the direction of propagation, a patchy structure with small blobs at approximately fixed distance and the decay of the density of tracer along the injection site.

We model each injection with stochastic paths solutions of equation (4.2.8). Then we evaluate the stochastic paths on the pinwheel structure.

Due to the stochastic nature of the problem, we do not compare pointwise the image of the tracer and the stochastic paths but we average them on the pinwheels. We partition both the images of the tracer and of the stochastic paths in M regions corresponding to the pinwheels:

$$I = \cup_i R_i \tag{5.1.1}$$

and for every R_i we compute the density of tracer DT_i and the density of the stochastic paths DP_i. The two vectors DT_i and DP_i are then rescaled in such a way to have unitary L^2-norm and the mean square error is computed:

$$E = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(DP_i - DT_i\right)^2} \tag{5.1.2}$$

The free parameters of the model are the value of the standard deviation, the number of paths, the number of steps, appearing in equation (4.2.8) and in the system (4.3.2). The best fit between the experimental and simulated distributions has been accomplished by minimizing the mean square error by varying these parameters.

Due to the different role of the directions X_1 and X_2 in the definition of these kernels, the Sub-Riemannian Laplacian paths and the Fokker Planck paths have different structure.

The Subriemannian Laplacian allows diffusion in direction X_2, allows the changement of the angle and it can be used to describe short range connectivity. Hence it is responsible for the central blob, in a neighborhood of the injection points (see Figure 5.2 (c)). The Fokker Planck kernel produces an elogated, patchy structure and seems responsible for the long range connection (see Figure 5.2 (d)). We apply our quantitative fit only to the long range connectivity, hence discarding the tracer in a neighborhood of the injection. For this reason the Sub-Riemannian Laplacian is not involved in the validation of the model.

The method is first applied to fit the image of the tracer taken by Bosking [Bosking et al., 1997] (see Figure 5.2 (a)). All the kernels are evaluated on the pinwheels provided in the same paper (see figure 5.2 (b)), to obtain a patchy structure. In order to apply the formula (5.1.1), we cover both the image of the tracer and the Fokker Planck with a regular distribution of rectangles, with edges equal to the mean distances between pinwheels (see Figure 5.2 (c),(d)) (clearly we do not cover the central zone, where we can not fit the Fokker Planck kernel).

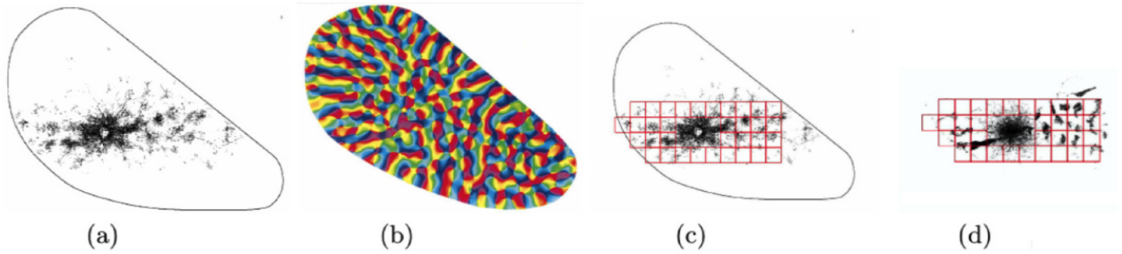


Figure 5.2: The connectivity map measured by Bosking et al. [1997] (a), the pinwheel structure used for the estimate (b), the tracer partitioned according to rectangles with sides equal to the distance between pinwheels (c) and the best fit results (d).

The resulting error value is $E < 8\%$, showing that the model accurately represents the experimental distribution.

A similar procedure has been applied to the image of the tracer provided in Angelucci et al. [2002] (see Figure 5.3 (a)). The result of Angelucci is obtained with various injections in a neighborhood of a pinwheel, so that all orientations are present, and the tracer propagates in all directions. In this case we do not have natural pinwheels, hence we use artificial pinwheels, obtained with the algorithm presented in Barbieri et al. [2012] (see Figure 5.3 (b)), with the constraint that the mean distance between the artificial pinwheels is equal to the mean distance between the blobs produced by the tracer. Here we consider Fokker Planck paths with all directions, to obtain the apparent isotropic diffusion. Also in this case we cover with rectangles and perform a best fit and the minimum error value is $E < 8\%$, (see figure 5.3 (c), (d).

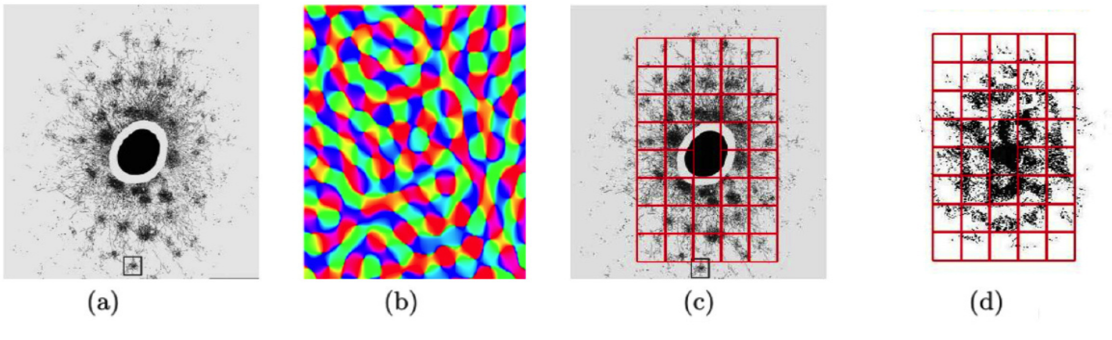


Figure 5.3: The connectivity map measured by Angelucci et al. [2002] (a), the pinwheel structure used for the estimate (b), the tracer partitioned according to rectangles with sides equal to the distance between pinwheels (c) and the best fit results (d).

In his paper Bosking et al. [1997] showed a famous image, with the tracer superimposed to the piwheel structure (see Figure 5.4 (a)). In particular in this case we have the tracer and the pinwheel of the same animal. This allows to go below the scale of the pinwheel and we correctly recover the orientation with the pinwheel (see Figure 5.4 (b)). The estimated kernel is again a combination of Fokker Plank. As before, we focus on orientations, hence we only model the long range part of the image, discarding the central blob. The evaluation of the error is made with squared regions at a scale smaller that the pinwheel and the error goes below $E < 9\%$.

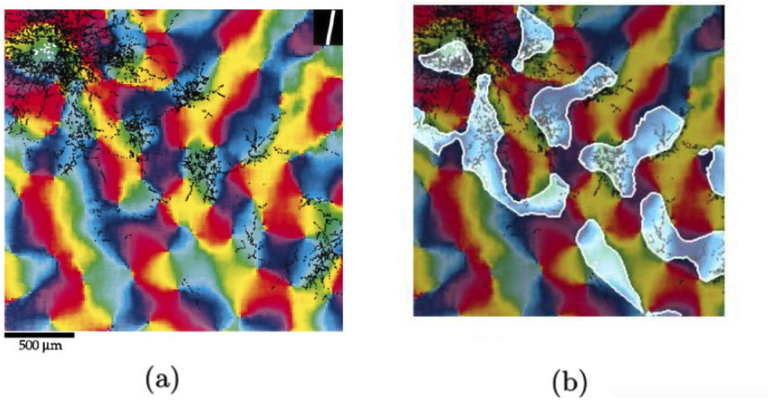


Figure 5.4: In (a) the tracer superimposed to the piwheel structure found by Bosking et al. [1997] and in (b) the isocontours obtained from a combination of Fokker Planck.

5.1.2 Perceptual facilitation and density kernels

In order to obtain a stable and deterministic estimate of this stochastic model, we use the density kernel, which is a regular deterministic function, coding the main properties of the process. We perform in Favali et al. [2016b] a quantitative validation of these regular kernel comparing to an experiment of Gilbert et al. [1996].

This work studies the capability of cells to integrate information out of the single receptive field of the cells. This integration process is due to the long-range horizontal connections, hence it can be used to validate our model of long range connectivity. As we have recognized in the previous section it is the Fokker Planck kernel which can be considered as a model for long range connectivity, hence we use here this kernel.

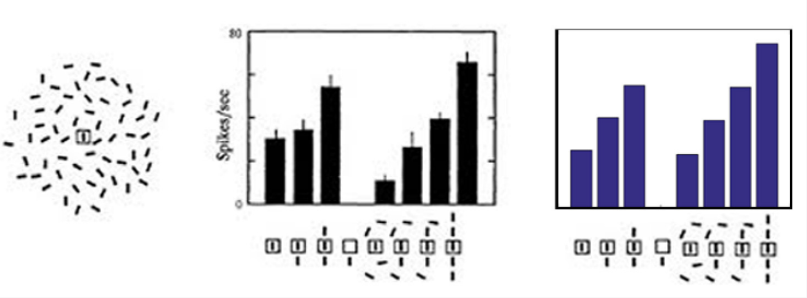


Figure 5.5: The experiment of Gilbert et al. [1996], with the stimulus composed by randomly placed and oriented lines (left) and the black histogram of cell's response (middle). On the right the histogram evaluated from the probability density in response to the same distribution of lines.

In Figure 5.5 (left) it is shown the results of Gilbert et al. [1996], where it is visualized the cell's response to randomly placed and oriented lines in a black histogram (middle). A vertical line is present in the receptive field of a cell selective to this orientation and the intensity of its response is represented in the first column of the histograms. If the stimulus is surrounded by random elements aligned with the first one, the cell's response increases (respectively in the second, third and the last column of the histograms). When the other random elements are not aligned with the fixed one (as in the fifth, sixth, seventh columns), the cell's response decreases because it reflects an inhibitory effect.

On the right in the blue histogram we evaluate the probability density modelled by the kernel in equation (4.2.12) in presence of the same configuration of elements. The same trend is obtained considering the probability density distribution, as visualized in Figure 5.5 (right). In order to consider the inhibitory effect we evaluate the kernel with 0 mean. A quantitative analysis of the differences between them have been evaluated considering the mean square error between the two normalized histograms. The error of 8% underlines how this connectivity kernel well represents neural connections.

6 Generalization of the model of cortical connectivity to $R^n \times S^1$

In Chapter 4 we presented a cortical model in $R^2 \times S^1$, and recognized that the theoretical instruments, which support the model, can be extended to a general space $R^n \times S^1$. This remark has an interest which goes beyond the theoretical content of the theorems we stated. Indeed the visual cortex has a modular structure and its cells are capable to extract other features, as contrast, curvature, scale. This is why we will now include in the model of cortical connectivity and of individuation of perceptual units other features. We present how to consider in the model of connectivity the feature of intensity [Favali et al., 2016a] and an extension to a 5 dimensional kernel in the lifted space of positions, orientations, intensities and curvatures [Abbasi-Sureshjani et al., 2016a].

In chapter 7 we will present applications of these models to the problem of individuation of perceptual units in illusory figures and to the analysis of retinal vessel, for which these features are relevant.

6.1 4-D kernel - intensity integration

In order to include the intensity term in the model of cortical connectivity, we use the Euclidean distance between the intensities of two corresponding points. If $f(x,y)$ represents the image intensity at position (x,y) the stimulus is lifted to the extended 4-dimensional feature space:

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

An admissible curve in this space is defined as the solution of the following differential equation:

$$\begin{aligned} \gamma'(s) &= (k_1(s)\vec{X}_1 + k_2(s)\vec{X}_2 + k_4(s)\vec{X}_4)(\gamma(s)) \\ \gamma(0) &= (x_1, y_1, \theta_1, f_1), \quad \gamma(1) = (x_2, y_2, \theta_2, f_2) \end{aligned} \tag{6.1.1}$$

where the vector fields are:

$$\vec{X}_1 = (\cos\theta, \sin\theta, 0, 0), \quad \vec{X}_2 = (0, 0, 1, 0),$$

$$\vec{X}_4 = (0, 0, 0, 1) \tag{6.1.2}$$

and the coefficients k_1 and k_2 represent a distance in the (x, y, θ) domain and k_4 is an Euclidean distance. Starting from these vector fields we can model the cortical connectivity, that can be described as the probability of connecting two points in the cortex, directly applying the theory developed in Section 3.

The stochastic counterpart of the curves in equation (6.1.1) can be represented as:

$$(x', y', \theta', f') = \vec{X}_1 + N(0, \sigma_1^2)\vec{X}_2 + N(0, \sigma_2^2)\vec{X}_4 \tag{6.1.3}$$

where $N(0, \sigma_1^2)$, $N(0, \sigma_2^2)$ are normally distributed variable with zero mean and variance equal to σ_i^2. It is a particular case of equation (4.2.5), so that Theorem 2 and Proposition 2 can be applied and we could use our general theory to find the fundamental solution of the operators $L_{SR,4}$ and $FP_{SR,4}$ in $R^3 \times S^1$ and the 4D symmetrized kernel ω_4 in $R^3 \times S^1$. Note that these operators are:

$$L_{SR,4} = \partial_t + X_1 - \sigma_2^2 X_{22} - \sigma_4^2 X_{44} \tag{6.1.4}$$

and:

$$FP_{SR,4} = X_1 - \sigma_2^2 X_{22} - \sigma_4^2 X_{44}. \tag{6.1.5}$$

Let us explicitly note that the vector fields X_1, X_2, X_3 from one side and X_4 depend on completely different variables. This allows to simplify the study when considering the kernel associated to the operator $L_{SR,4}$.

Its fundamental solution satisfies the following property:

Proposition 3. *The fundamental solution p_4 of the operator $L_{SR,4}$ can be explicitly written as:*

$$p_4 = p_1 p_f,$$

where p_1 is the fundamental solution of the operator L_{SR} defined in (4.2.9) *and p_f is the fundamental solution of the operator $\partial_t - X_{44}$, which is a Gaussian bell.*

Proof. For the sake of simplicity we make the computation only far from the pole:

$$\partial_t(p_4) = \partial_t(p_1 p_f) = \partial_t(p_1)p_f + p_1\partial_t(p_f) =$$

$$= (-X_1 + X_{22})p_1 p_f + p_1 X_{44} p_f = (-X_1 + X_{22} + X_{44})(p_1 p_f).$$

Hence $L_{SR,4}(p_1 p_f) = 0$ far from the pole, which proves the assertion. □

As a consequence, applying Proposition 2 the fundamental solution Γ_4 of the operator $FP_{SR,4}$ can be computed as:

$$\Gamma_4 = \int_0^\infty p_4(x, y, \theta, f, t)dt,$$

and the associated connectivity kernel ω_4 can be obtained by symmetrization.

In order to simplify, we will approximate the connectivity kernel ω_4 as the product of the connectivity kernel ω_1 obtained by symmetrization of the 3D Fokker Planck fundamental solution in equation (4.2.12) with a Gaussian bell in the variables f:

$$\omega_f(f_i, f_j) = e^{(-\frac{1}{2}(\frac{f_i - f_j}{\sigma_2}))^2}. \tag{6.1.6}$$

The final connectivity kernel can be written as the product (as these are probabilities) of the two components:

$$\begin{aligned} &\omega_4((x_i, y_i, \theta_i, f_i), (x_j, y_j, \theta_j, f_j)) \\ &= \omega_1((x_i, y_i, \theta_i), (x_j, y_j, \theta_j))\omega_f(f_i, f_j). \end{aligned} \tag{6.1.7}$$

Starting from the connectivity kernel defined previously, it is possible to extract perceptual units from images by means of spectral analysis of suitable affinity matrices. The eigenvectors with the highest eigenvalues are linked to the most salient objects in the scene [Perona and Freeman, 1998]. The connectivity is represented by a real symmetric matrix $A_{i,j}$:

$$A_{i,j} = \omega_4((x_i, y_i, \theta_i, f_i), (x_j, y_j, \theta_j, f_j)) \tag{6.1.8}$$

that contains the connectivity information between all the lifted points. The eigenvectors of the affinity matrix are interpreted as perceptual units [Sarti and Citti, 2015, Favali et al., 2016b].

This approximation allows us to use the kernel ω_1, already numerically computed in Section 4.3.3. The intensity-based kernel ω_f, the final connectivity kernel ω_4 and the affinity matrix A, were calculated using equations (6.1.6), (6.1.7), (6.1.8) respectively.

6.2 5-D kernel - curvature integration

6.2.1 Review of Curvature models

According to the experiments of Dobbins et al. [1987, 1989], the primary visual cortex not only includes the orientation selective cells, but it is also sensitive to the curvature at each point. According to this conjecture it is the role of the end-stopping cell to detect curvature. The role of curvature within the problem of edge organization in images was also addressed by August and Zucker [2003, 2000]. Their main idea was to use the positions orientations space and Mumford's Fokker-Planck equation defined in (4.2.7) for the propagation of the direction information. Including the curvature, leads to the representation of the image in the space $\mathbb{R}^2 \times S^1 \times \mathbb{R}$. It is worthwhile to note that this 4D space does not coincide with the 4D one studied in the previous section, since the role of the variables is not the same. When the curvature of the structures present in the image is high, this method allows to obtain better results as visualized in Figure 6.1. In Ben-Shahar and Zucker [2004] is given a justification based on biological evidence for the consideration of the term of curvature. This framework leaves the curve model as a free parameter, allowing to introduce a new stochastic model for the detection of curvature to better capture the shape of the curves.

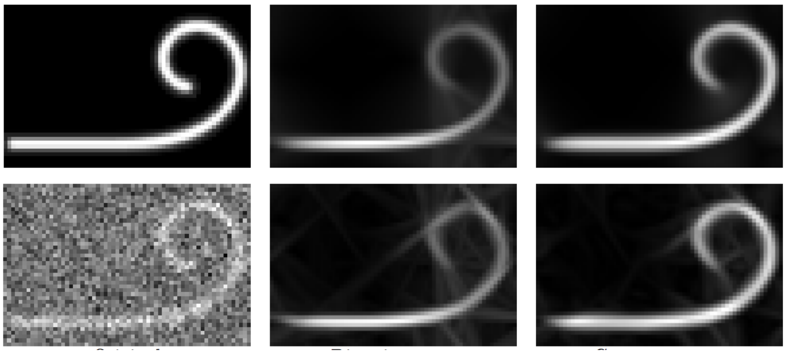


Figure 6.1: Filtering an Euler spiral without noise (top) and with noise (bottom). The original images (left); the result after filtering using the curve indicator random field based on Mumford's direction-based Markov process (center) and using curvature-based Markov process of August and Zucker [2003] (right). Source: [August and Zucker, 2003].

In computer vision and more specifically in retinal image analysis, several methods have been introduced for measuring the curvatures of curvilinear structures (i.e. blood vessels) [Kalitzeos et al., 2013]. The classical methods that measure the curvatures locally need an initial segmentation, centreline extraction, and separation of segments located between junctions. It is then followed by fitting curves to the segments and by curvature measurement using equation (6.2.3) [Hart et al., 1999, Annunziata et al., 2016, Wilson et al., 2008]. The drawback of all these methods is their dependency to initial preprocessing and segmentation steps, which may contain errors and missing information. More importantly, the curvature information is not available for junction points because it is not possible to fit a curve to these points where more than one elongated structure meets.

To solve these problems, Bekkers et al. [2015] proposed a local curvature measurement technique by locally fitting exponential curves [Duits et al., 2016] to the lifted image in $SE(2)$. The exponential curves in $SE(2)$ are interpreted as straight lines considering the curved geometry of $SE(2)$ and they have constant tangent vectors relative to the rotating frame $\{\vec{X}_1, \vec{X}_2, \vec{X}_3\}$. The tangent vectors of the exponential curve that best fit the data in the lifted image are obtained by eigensystem analysis of the Gaussian Hessian (expressed in the rotating frame). Then they directly define the curvature value of their spatial projections [Franken and Duits, 2009]. This approach makes it possible to assign to each location and orientation in the lifted image a curvature value, without needing explicit curve parameterizations. Such curvature maps (on SE(2)) can be projected on the plane whereby only one value of curvature value is assigned to each spatial location in the image. Finally, these 2D curvature maps can be filtered in a later stage by a vessel confidence map (as a Laplacian ridge detector) [Franken and Duits, 2009] or any other vessel enhancement methods.

6.2.2 The neurogeometry of the curvature space

A regular curve in the two-dimensional plane can always parametrized by arch length, so that then the tangent vector becomes unitary and can be represented in the form:

$$(\dot{x}(t), \dot{y}(t)) = (cos(\theta(t)), sin(\theta(t))). \tag{6.2.1}$$

By differentiating x and y once more in equation (6.2.1), we then have:

$$(\ddot{x}, \ddot{y}) = (-\sin(\theta(t))\dot{\theta}(t), \cos(\theta(t))\dot{\theta}(t)). \tag{6.2.2}$$

So the curvature can be computed as:

$$\kappa = \dot{\theta} = \frac{\dot{x}\ddot{y} - \ddot{x}\dot{y}}{(\dot{x}^2 + \dot{y}^2)^{3/2}}. \tag{6.2.3}$$

We assume that the two-dimensional curve in cortical plane ($\mathbb{R}^2$) is lifted to a 5D space of positions, orientations, intensity and curvature ($\mathbb{R}^2 \times S^1 \times \mathbb{R} \times \mathbb{R}$). Thus, the lifted curve may be written as:

$$\gamma_{2D} = (x(t), y(t)) \rightarrow \gamma(t) = (x(t), y(t), \theta(t), f(t), \kappa(t)). \tag{6.2.4}$$

According to equations (6.2.1) and (6.2.3), we will have:

$$\dot{\gamma}(t) = (\dot{x}(t), \dot{y}(t), \dot{\theta}(t), \dot{f}(t), \dot{\kappa}(t)) \tag{6.2.5}$$

where $(\dot{x}(t), \dot{y}(t)) = (\cos(\theta), \sin(\theta))$ and $\dot{\theta}(t) = \kappa(t)$. Hence:

$$\dot{\gamma}(t) = (\cos(\theta), \sin(\theta), \kappa(t), \dot{f}(t), \dot{\kappa}(t)). \tag{6.2.6}$$

By defining new vectors in the 5D space as:

$$\begin{aligned} \vec{Y}_1 =& (\cos(\vec{\theta}), \sin(\vec{\theta}), \kappa, 0, 0) \\ \vec{Y}_5 =& (0, 0, 0, 0, 1) \qquad \vec{Y}_4 = (0, 0, 0, 1, 0) \end{aligned} \tag{6.2.7}$$

we are able to write $\dot{\gamma}(t)$ in terms of these vectors:

$$\dot{\gamma}(t) = \vec{Y}_1(t) + \dot{f}(t)\vec{Y}_4 + \dot{\kappa}(t)\vec{Y}_5(t). \tag{6.2.8}$$

Let us explicitly note that this expression is a generalization of the lifted curve used by Citti and Sarti [2006] in equation (4.1.10). In general, we will call horizontal curve the solution of the following differential equation:

$$\begin{aligned} \dot{\gamma}(t) &= (\alpha_1(t)\vec{Y}_1(t) + \alpha_5(t)\vec{Y}_5(t) + \alpha_4(t)\vec{Y}_4(t))(\gamma(t)) \\ \gamma(0) &= (x_0, y_0, \theta_0, f_0, \kappa_0), \gamma(1) = (x_1, y_1, \theta_1, f_1, \kappa_1). \end{aligned} \tag{6.2.9}$$

The horizontal distribution of planes is now Span$\{\vec{Y}_1, \vec{Y}_5, \vec{Y}_4\}$ and the Lie algebra is generated by these three vector fields. For the sake of simplicity, we will call $\vec{Y}_2 = \vec{X}_2$, $\vec{Y}_3 = \vec{X}_3$. Following the same notation as in (4.1.16) we denote Y_i the directional derivatives in the direction of the vectors $\vec{Y}_i$:

$$\begin{aligned} Y_1 =& \cos(\theta)\partial_x + \sin(\theta)\partial_y + \kappa\partial_\theta \\ Y_2 =& \partial_\theta \quad Y_3 = -\sin(\theta)\partial_x + \cos(\theta)\partial_y \\ Y_5 =& \partial_k \quad Y_4 = \partial_f \end{aligned} \tag{6.2.10}$$

Therefore, the commutators of these vectors are as follows:

$$\begin{aligned} [Y_1, Y_5] &= -\partial_\theta = -X_2 = -Y_2 \\ [[Y_1, Y_5], Y_1] &= \sin(\theta)\partial_x - \cos(\theta)\partial_y = -X_3 = -Y_3 \\ [[Y_1, Y_5], Y_5] &= 0. \\ [Y_i, Y_4] &= 0 \text{ for every } i. \end{aligned} \tag{6.2.11}$$

In particular the Hörmander condition (see Theorem 1) is satisfied by the vector fields Y_1, Y_4, Y_5. Figure 6.2 represents the fan of integral curves of equation (6.2.8) in the $SE(2)$ group for similar number of κ values changing with two different $\dot{\kappa}$ rates. This figure shows how the shape of curves is dependent on the curvature value (κ) and its rate of change ($\dot{\kappa}$). The projection of both curves on 2D cortical plane represents a good model of the association fields introduced by Field et al. [1993] for modeling the cortical connectivity. The association fields (or connectivity patterns) are considered as the basis for the creation of connected boundaries in visual perception, imposing the Gestalt law of good continuation [Wagemans et al., 2012, Wertheimer, 1938].

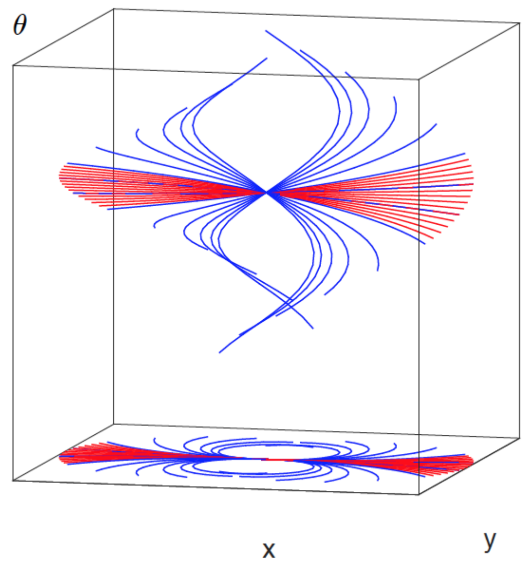


Figure 6.2: The fan of integral curves of equation (6.2.8), visualized in $\mathbb{R}^2 \times S^1$ and their projection on the xy plane. The blue curves have a higher rate of curvature change ($\dot{\kappa}$).

The cortical connectivity can also be modeled by a stochastic counterpart of equation (6.2.8), which follows in the general framework stated in equation (4.2.5). The Markov process that results from the Brownian motion with randomly curved paths has been introduced by August and Zucker [2003]. The process is represented by the following differential equations:

$$\gamma' = \alpha_1 \vec{Y}_1 + N(0, \sigma_1^2)\vec{Y}_4 + N(0, \sigma_2^2)\vec{Y}_5 \qquad (6.2.12)$$

where $N(0, {\sigma_i}^2)$ is a normally distributed variable with zero mean and variance equal to σ_i^2. If p_5 denotes the probability density to find a particle at a point (x, y) with a certain direction θ, intensity f and curvature κ, at a specific time t, then the Fokker-Planck equation describing the diffusion of the particle density will be according to Theorem 2:

$$\partial_t p_5 = \frac{\sigma_5^2}{2} Y_{55} p_5 + \frac{\sigma_4^2}{2} Y_{44} p_5 - Y_1 p_5 \tag{6.2.13}$$

so that $Y_{44} = \partial^2/\partial f^2$ and $Y_{55} = \partial^2/\partial \kappa^2$. This partial differential equation means that a particle at a point $(x, y, \theta, f, \kappa)$ transports in the direction of $(\cos(\theta), \sin(\theta), \kappa, 0, 0)$ in the 5D space. There is no transport in the f or κ direction, but the diffusion in the κ direction indicates the rate of transport in the θ direction. The diffusion in direction f is independent from the other variables.

Integrating in time, we obtain thanks to Proposition 2 the fundamental solution Γ_5 of the following equation:

$$\begin{aligned} &\frac{\sigma_4^2}{2} Y_{44} \Gamma_5((x, y, \theta, f, \kappa), (x', y', \theta', f', \kappa')) \\ +&\frac{\sigma_5^2}{2} Y_{55} \Gamma_5((x, y, \theta, f, \kappa), (x', y', \theta', f', \kappa')) - \\ &Y_1 \Gamma_5((x, y, \theta, f, \kappa), (x', y', \theta', f', \kappa')) = \\ &\delta(x, y, \theta, f, \kappa). \end{aligned} \tag{6.2.14}$$

In order to compute an estimate of the fundamental solution we first restrict as in the previous section to a fundamental solution Γ_k of the Fokker Planck operator independent of f:

$$\frac{\sigma^2}{2} Y_{55} + \frac{\sigma^2}{2} Y_{44} - Y_1.$$

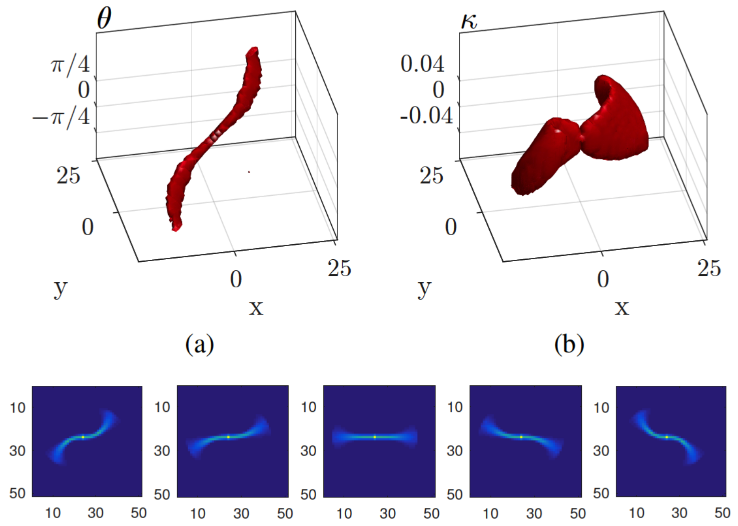
We recall that this is possible because the vector field Y_4 is independent of all the other variables. Let us explicitly note that we can not perform the same simple restriction to the variable k, since the vectors Y_1 and Y_5 do not commmute. The restriction in this case is more delicate. Then we restrict the fundamental solution to a 3D kernel, with κ fixed:

$$\Gamma'_\kappa((x, y, \theta), (x', y', \theta')) = \Gamma_k((x, y, \theta, k), (x', y', \theta', k))$$

, symmetrized and multiplied by an exponential term which considers the closeness between two points located in the different intensity planes, times an exponential term in the curvature planes. Hence, this new connectivity kernel is presented as:

$$\begin{aligned} &\omega_5((x, y, \theta, f, \kappa), (x', y', \theta', f', \kappa')) = \\ &e^{-\frac{(\kappa-\kappa')^2}{\sigma_\kappa^2}} \times e^{-\frac{(f-f')^2}{\sigma_{int}^2}} \times \\ &\frac{1}{2}\left(\Gamma'_\kappa((x, y, \theta), (x', y', \theta')) + \Gamma'_{\kappa'}((x', y', \theta'), (x, y, \theta))\right) \end{aligned} \tag{6.2.15}$$

Figure 6.3 represents two level sets of the 5D kernel by fixing κ (a) and θ (b) dimensions. The 2D projection on R^2 by summation over all orientations of five different 4D stochastic kernels having different curvature (κ) values is also presented by Figure 6.3 (c). The intensity term is kept constant for all figures. As seen in these figures, by increasing the absolute value of curvature, the shape of the kernel also changes and it deviates from the elongated shape.



(a) (b)

(c)

Figure 6.3: Visualizations of the 5D stochastic kernels in 3D and 2D. (a) and (b) the iso-surfaces of the kernel while keeping κ and θ fixed respectively; (c) the 2D projection of the kernel over all orientations for several curvature values: $\kappa = \{-0.08, -0.04, 0, 0.04, 0.08\}$ from left to right. Intensity is constant for all figures.

This new kernel can be used for defining the affinity matrix as:

$$A_{i,j} = \omega_5((x_i, y_i, \theta_i, f_i, \kappa_i), (x'_j, y'_j, \theta'_j, f'_j, \kappa'_j)). \qquad (6.2.16)$$

By this definition, considering the fact that this matrix includes information about the correct grouping, this problem has been presented in terms of dimensionality reduction of this matrix [Favali et al., 2016b, Sarti and Citti, 2015], often done by eigensystem analysis.

6.2.3 Numerical approximation of the 5-D kernel

The kernel is numerically estimated using the general Markov Chain Monte Carlo method technique recalled in section 4.3.3. We say a few words on how to adapt it to the 5D case for reader convenience. The system in equation (6.2.12) can be approximated by:

$$\begin{cases} x_{s+\Delta s} - x_s = \Delta s \cos(\theta) \\ y_{s+\Delta s} - y_s = \Delta s \sin(\theta), \quad s \in 1, ..., H \\ \theta_{s+\Delta s} - \theta_s = \Delta s \kappa \\ \kappa_{s+\Delta s} - \kappa_s = \Delta s N(0, \sigma_5) \end{cases} \tag{6.2.17}$$

where H is the number of steps of the random path and $N(0, \sigma_5)$ is a generator of numbers taken from a normal distribution with mean 0 and standard deviation of σ_κ. The stochastic path is obtained from the estimate of the kernel as the average of their passages over discrete volume elements, solving this finite difference equation n times [Sarti and Citti, 2015]. The affinity matrix described in equation (6.2.16) is evaluated from this kernel.

6.2.4 Product of normalized affinity matrices

We have introduced a quite general setting which generates an affinity matrix starting from a set of vector fields $X_1, \cdots, X_m$, satisfying an Hörmander condition in $R^n \times S^1$. In general the affinity matrix A does not have a probabilistic meaning. It can be normalized by columns to recover this meaning. Defining the diagonal matrix D as the sum $d_i = \sum_j a_{ij}$, the normalized affinity matrix is obtained as: $P = D^{-1}A$ where P represents the transition probability. In case different normalized affinity matrices A_1, A_2, are used to describe the measured connectivity in different cortical areas, their product $A_1 A_2$ describes the joint action of the two probability measures.

7 Individuation of perceptual units

We have presented in Chapter 4 and 5 the cortical connectivity kernels. Here we describe the results of simulations presented in Favali et al. [2016b]: we will identify perceptual units in different Kanizsa figures, highlighting the role of polarity, discussing and comparing the behavior of the different kernels.

We will also apply the presented method to the analysis of retinal images [Abbasi-Sureshjani et al., 2016a, Favali et al., 2016a], to afford the problem of grouping during the tracking of blood vessel, highlighting the role of the feature of intensity and curvature. These works have been developed in collaboration with the Eindhoven Univesity of Technology (TU/e).

7.1 Phenomenological experiments: emergence of percepts in illusory images

In the following experiments some numerical simulations will be performed in order to test the reliability of the method presented in Section 4.2 for performing grouping and detection of perceptual units in images. The kernel considered here only depends on orientation. Hence it can be applied to detect the saliency of geometrical figures, which can be very well described using this feature.

7.1.1 Proposed technique

The purpose is to select the perceptual units from visual images, using the following algorithm:

1) define the affinity matrix $A_{i,j}$ from the connectivity kernel;

2) solve the eigenvalue problem $A_{i,j}u_i = \lambda_i u_i$, where the order of i is such that λ_i is decreasing;

3) find and project on the segments the eigenvector u_1 associated to the largest eigenvalue.

The parameters used are: 1000000 random paths with $\sigma = 0.15$ in the system (4.3.2), $\sigma_1 = 1.2$, $\sigma_3 = 0.11$ in the system (4.3.3), σ, $\rho = 0.15$ in the discretization of (4.2.16). The value of H is defined as follows: $H = \frac{1}{3} d_{max}$, where d_{max} is the maximum distance between the inducers of the stimulus. Similar parameters have been used for all the experiments.

7.1.2 The Field, Hayes and Hess experiment

In this section we consider some experiments similar to the ones of Field et al. [1993], where a subset of elements organized in a coherent way is presented out of a ground formed by a random distribution of elements. A first stimulus of this type is represented in Figure 7.1 (left). The connectivity among these elements is defined as in equations (4.2.10) and (4.2.14).

After the affinity matrix and its eigenvalues, the eigenvector corresponding to the highest eigenvalue is visualized in red. The results show that the stimulus is well segmented with the fundamental solutions of Fokker Planck and Sub-Riemannian Laplacian equations (Figure 7.1 (right)).

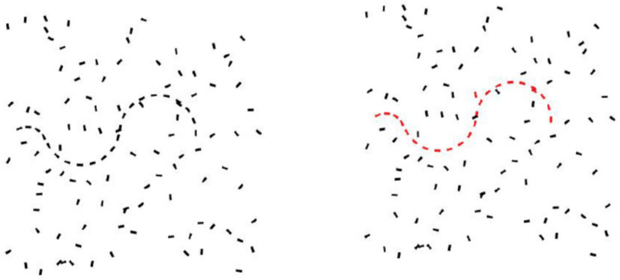


Figure 7.1: Example of stimulus (left) similar to the experiments of Field et al. [1993]. The stimulus containing a perceptual unit is segmented with Fokker Planck and Sub-Riemannian Laplacian (right), using the first eigenvector of the affinity matrix.

Now we consider a similar experiment proposed in Field et al. [1993], where the orientation of successive elements differs by 15, 30, 45, 60 and 90 degrees and the ability of the observer to detect the path was measured experimentally. It was proved that the path can be identified when the successive elements differ by 60 deg or less. With our method, we obtain similar results: if the angle between successive elements is less than 60 degree (Figure 7.2 (a), (b),(c)), the identification of the unit is correctly performed. With an angle equal to 60 degree (Figure 7.2 (d)) only a part of the curve is correctly detected: this can be interpreted as the increasing observer's difficulty to detect the path. Considering higher angles (Figure 7.2 (e)) the first eigenvector of the affinity matrix corresponds to random inducers, confirming the results.

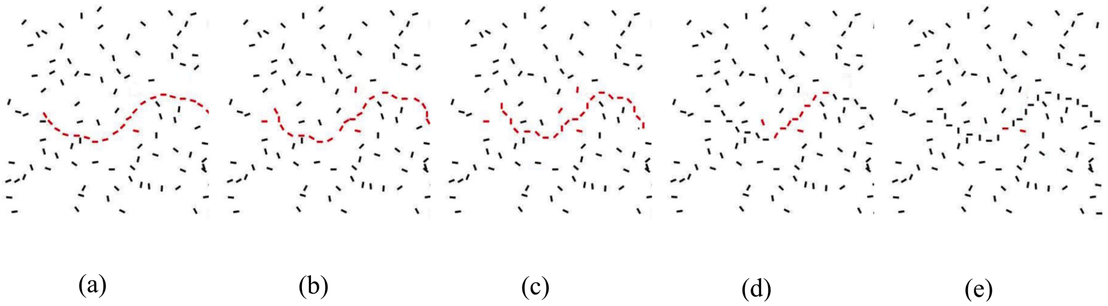


Figure 7.2: In red the first eigenvector of the affinity matrix considering images containing paths in which the orientation of successive elements differs by 15 (a), 30 (b), 45 (c), 60 (d) and 90 (f) degrees.

Finally we present an example where there are two units in the scene with roughly-equal salience, they have roughly-equal eigenvalues. In the first and in the second row of Figure 7.3 the stimuli are composed by a curve and a line in a background of random elements. In the stimulus (a) represented in the first row, the elements composing the curve are perfectly aligned and very nearby, so that this has the highest saliency and it represents the eigenvector associated to the first eigenvalue (as shown in red in Figure 7.3 (b)). The second eigenvalue in this case is sligtly smaller. After the computation of the first eigenvector, the stimulus is updated (Figure 7.3 (c)), the first eigenvector of the new affinity matrix is computed and it corresponds to the inducers of the line (Figure 7.3 (d)).

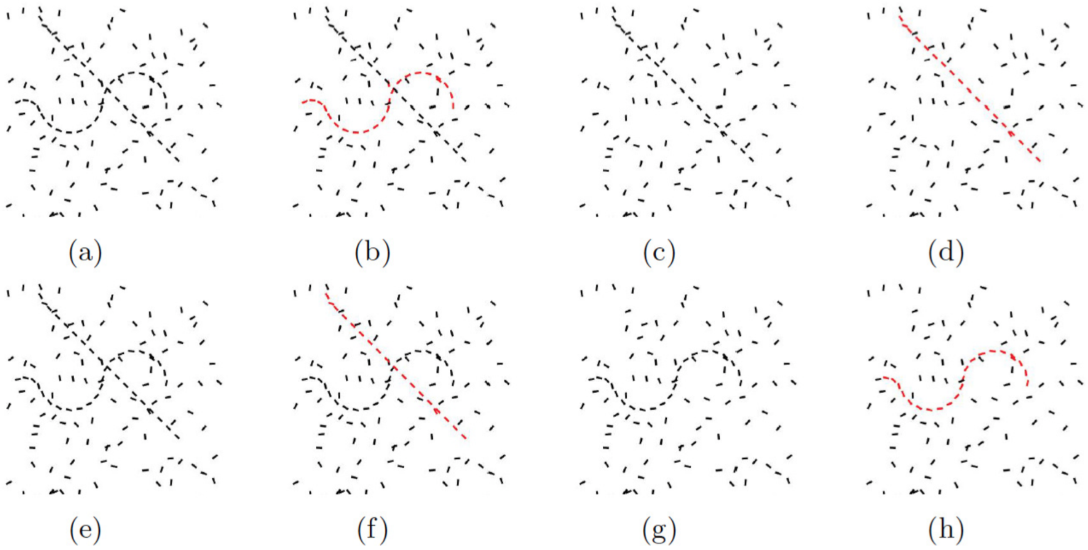


Figure 7.3: First and second rows. Examples (a,e) with two units in the scene, where a change in the angle leads to a change in the order of the eigenvalues (b,f), (c,g), (d,h).

In the second row (e) we slightly modify the stimuli, in particular the alignement of the element forming the curve (e.g. an angle of pi/18). As a consequence, the line becomes the most salient perceptual unit and the first eigenvector (Figure 7.3 (f)). The stimulus is updated (Figure 7.3 (g)) and the first eigenvector of the new affinity matrix corresponds to the inducers of the curve (Figure 7.3 (h)). It is notable that in this case a small changement of the eigenvalues corresponds to small changement of the eigenvectors, but the first eigenvalue swaps with the second one and consequently we obtain an abrupt change in the perceved object.

In the previous examples we have considered all contours with almost the same length. We show here that this length does not affect the faeture of saliency. In Figure 7.4 (a), (e) are visualized two perceptual units with different length. The results underline how the proximity of contours is stronger than lenght: the shortest units with nearer segments are the first perceptual units, associated to the most salient eigenvectors (Figure 7.4 (b), (f)). Then the stimuli are updated (Figure 7.4 (c), (g)) and the second eigenvectors are visualized in Figure 7.4 (d), (h).

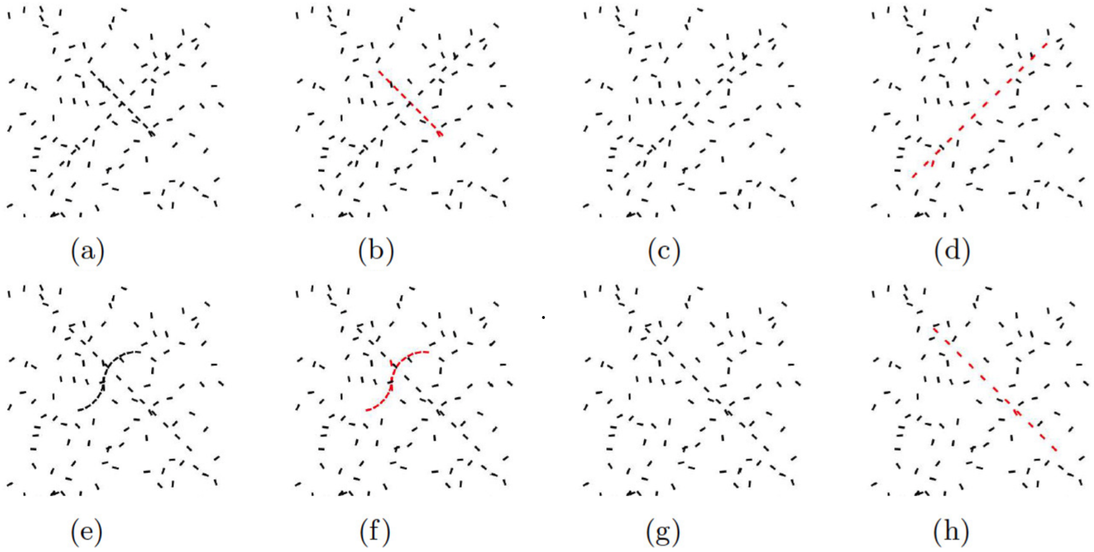


Figure 7.4: Examples (a,e) with two units in the scene with different length. In (b,f),(c,g) and (d,h) the results of simulation.

In this analysis different features can be considered. In particular also the distances between the segments play a central role. Let's consider for example the straight line in Figure 7.4 (a). If one or more segments is missing from the contour we could obtain a less accurate segmentation (a similar effect is noticed in the case of not-aligned segments). A similar analysis considering small or disconnected contours has been considered in Favali et al. [2016a], applied to the study of vessel connectivities.

7.1.3 The role of polarity

The term of polarity leads to insert in the model the feature of contrast: contours with the same orientation but opposite contrast are referred to opposite angles. For this reason we assume that the orientation θ takes values in $[0, 2\pi)$ when we consider the odd filters and in $[0, \pi)$ while studying the even ones.

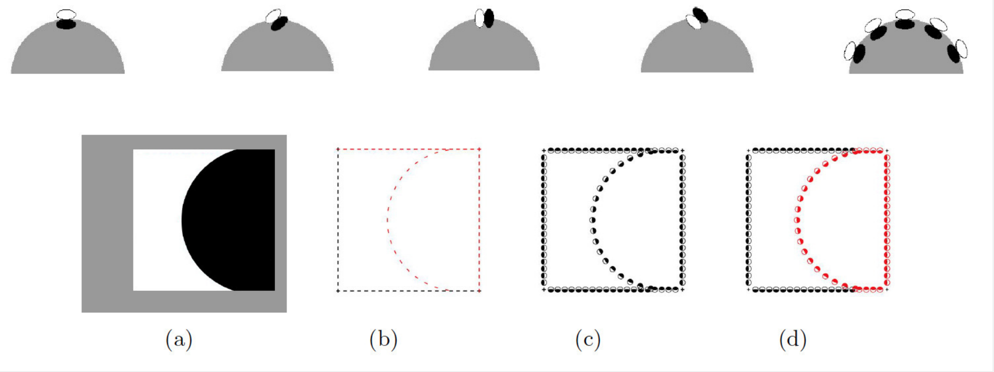


Figure 7.5: In the first row schematic description of the whole hypercolumn of odd simple cells centered in a point (x, y). The maximal activity is observed for the simple cell sensitive to the direction of the boundary of the visual stimulus. The set of maximally firing cells are visualized in the last image.
In the second row: a cartoon image (a), the first eigenvector of the affinity matrix without polarity (b), its representation with polarity dependent Gabor patches (b) and the corresponding first eigenvector (d).

The response of the odd filters in presence of a cartoon image is schematically represented in Figure 7.5. At every boundary point the maximally activated cell is the one with the same direction of the boundary. Then the maximally firing cells are aligned with the boundary (Figure 7.5, top right).

In order to clarify the role of polarity we consider an image in Figure 7.5 (a), that has been studied by [Kanizsa, 1980], in the contest of a study of convexity in perception. In this case, if we consider only orientation of the boundaries without polarity, we completely loose any contrast information and we obtain the grouping in Figure 7.5 (b). Here the upper edge of the square is grouped as an unique perceptual unit. On the other side, while inserting polarity, the Gabor patches on the upper edge boundary of the black or white region have opposite contrast and detect values of θ which differs of π (see Figure 7.5 (c)). In this way, there is no affinity between these patches, and the first eigenvector of the affinity matrix represented in red correctly detects the unit present in the image and corresponds to the inducers of the semicircle (see Figure 7.5 (d)). This underlines the important role of polarity in perceptual individuation and segmentation. We also note that the fist perceptual unit detected is the convex one, as predicted by the gestalt law (see [Kanizsa, 1980]).

7.1.4 The Kanizsa illusory figures

We consider here stimuli formed by Kanizsa figures, represented by oriented segments that simulate the output of simple cells. In Lee [2001] it is described that completion of Kanizsa figures takes place in V1.

We first consider the stimulus of Figure 7.6 (a). The connectivity among its elements will be analysed with the kernels defined in equations (4.2.10), (4.2.14).

The results of simulations with the fundamental solutions of Fokker Planck and Sub-Riemannian Laplacian equations are shown in Figure 7.6.

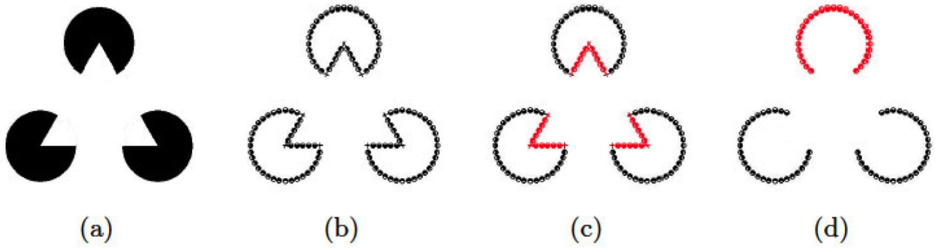

(a) (b) (c) (d)

Figure 7.6: The Kanizsa triangle (a) and the maximally responding odd filters (b). In (c) it is shown the first eigenvector of the affinity matrix, using the fundamental solutions of Fokker Planck (4.2.10,6.2.17) and Sub-Riemannian Laplacian equations (4.2.14,4.3.3). After this computation, the affinity matrix is updated removing the detected perceptual unit; the first eigenvector of the new matrix is visualized (d).

The first eigenvector is visualized in red and it corresponds to the inducers of the Kanizsa triangle (Figure 7.6 (c)). In this example, after the computation of the first eigenvector of the affinity matrix, this matrix is updated removing the identified perceptual unit and then the first eigenvector of the new matrix is computed (Figure 7.6 (d)): these simulations show that circles are associated to the less salient eigenvectors. In that way, the first eigenvalue can be considered as a quantitative measure of saliency, because it allows to segment the most important object in the scene and the results of simulations confirm the visual grouping.

When the affinity matrix is formed by different eigenvectors with almost the same eigenvalues, as in Figure 7.6 (d), it is not possibile to recognize a most salient object, because they all have the same influence. We choose here to show just one inducer in red. The other two have the same eigenvalue. That also happens, for example, when the inducers are not co-circularly aligned or they are rotated.

Now we consider as stimulus the Kanizsa square and then we change the angle between the inducers, so that the subjective contours become curved (Figure 7.7 (a), (b), (c), (d)). The fact that illusory figures are perceived depends on a limit curvature. Indeed we perceive a square in the first three cases, but not in the last one. The results of simulations with the fundamental solutions of Fokker Planck and Sub-Riemannian Laplacian equations confirm the visual grouping (Figure 7.7 (a), (b), (c), (d), second row): when the angle between the inducers is not too high (cases (a), (b), (c)) the first eigenvector corresponds to the inducers that form the square, otherwise (case (d)) the pacman becomes the most salient objects in the image. In this case, we obtain 4 eigenvectors with almost the same eigenvalue.

Now we consider a Kanizsa bar (Figure 7.7 (e)), that is perceived only if the inducers are aligned. Also in that case, the result of simulation confirms the visual perception if we use the fundamental solutions of the Fokker-Planck and the Sub-Riemannian Laplacian equations (Figure 7.7 (e), second row).

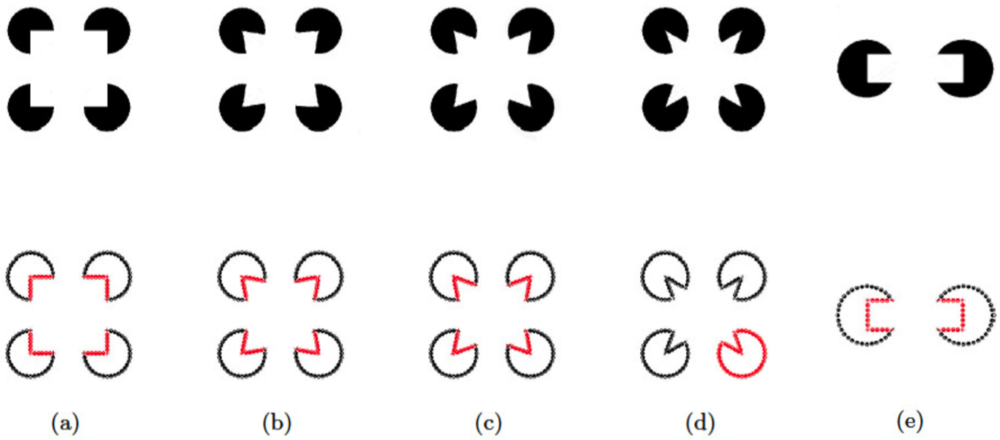


Figure 7.7: Examples of stimulus with aligned and not-aligned inducers. The first eigenvectors of the affinity matrix using the fundamental solutions of Fokker Planck and Sub-Riemannian Laplacian are visualized in red (second row).

When the inducers are not aligned, all the kernels confirm the visual perception, showing two different perceptual units (Figure 7.8 (a)).

Considering a stimulus composed of rotated or not-aligned inducers, as in Figure 7.8 (b), (c) it is not possible to perceive it and the results of simulations, using all the connectivity kernels described, confirm the visual grouping. In that case, the affinity matrix is decomposed in 3 eigenvectors with almost the same eigenvalues, which represent the 3 perceptual units in the scene.

7.1.5 Sub-Riemannian Fokker Planck versus Sub-Riemannian Laplacian

The two kernels we are going to analyze are not mutually exclusive and they can be implemented in different cells. The

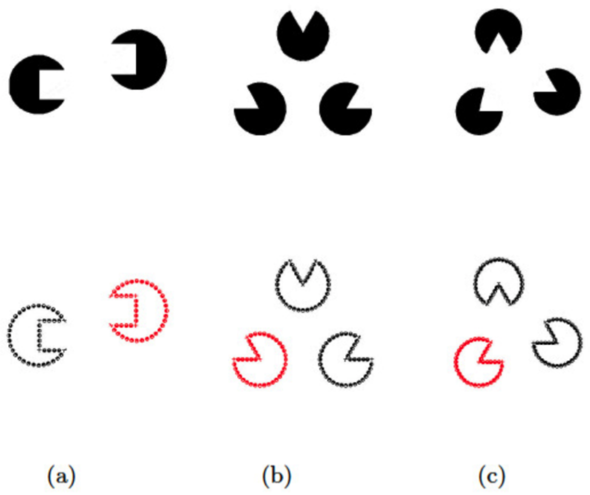


Figure 7.8: Stimulus with rotated (a) and not-aligned (b), (c) inducers. The first eigenvectors of the affinity matrix using the fundamental solutions of Fokker Planck and Sub-Riemannian Laplacian are visualized in red (second row).

presence of different population of cells in relation with mathematical models has been also studied in Ben-Shahar and Zucker [2004]. We have outlined in Sections 4.2.3 and 5.1.1 that the Fokker Planck kernel accounts for long range connectivity, while Sub-Riemannian Laplacian for short range. In the previous examples we obtain good results with both kernels, but this difference emerges while we suitable change the parameters. In Figure 7.9 we compare the action of these two kernels.

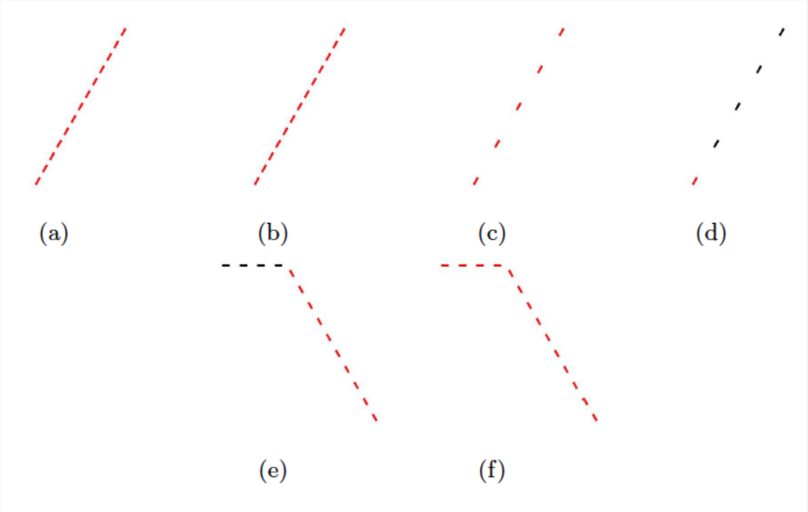


Figure 7.9: In the first row a few aligned segments, which are correctly grouped by the Fokker Planck and the Sub-Riemannian Laplacian (a), (b). When we separate the inducers, the perceptual unit is correctly detected using the Fokker Planck kernel (c), while the Sub-Riemannian Laplacian is not able to perform the grouping (d). In the second row we consider an angle. In this case the Fokker Planck is unable to perform the grouping (e), while the Sub-Riemannian Laplacian can correctly perform the grouping (f).

In the first row we see some segments, which form an unique perceptual unit. If they are not too far, the grouping is correctly performed both by the Fokker Planck and the Sub-Riemannian Laplacian (Figure 7.9 (a),(b)). When we separate the inducers, the perceptual unit is correctly detected by the Fokker Planck kernel (Figure 7.9 (c)), while the Sub-Riemannian Laplacian is not able to perform the grouping (Figure 7.9 (d)). This confirms that the Fokker Planck kernel is responsible for long range connectivity. In the second row we consider an example with segments disposed to form an angle. When the angle is sufficiently high, the Fokker Planck becomes unable to perform the grouping (Figure 7.9 (e)), while the Sub-Riemannian Laplacian, correctly performs the grouping of the perceptual unit (Figure 7.9 (f)). This confirms that the Sub-Riemannian Laplacian can be used as a model of short range connectivity.

7.1.6 Sub-Riemannian versus Riemannian kernels

In order to further validate the Sub-Riemannian model we show that the model applied with the isotropic Laplacian kernel does not perform correctly. As shown in Figure 7.10 (first row) the visual perception is not correctly modeled: the first eigenvectors coincide with one of the inducers and the squares are not recognized. That also happens for the stimulus of Figure 10 (a) and when the inducers are not co-circularly aligned or they are rotated.

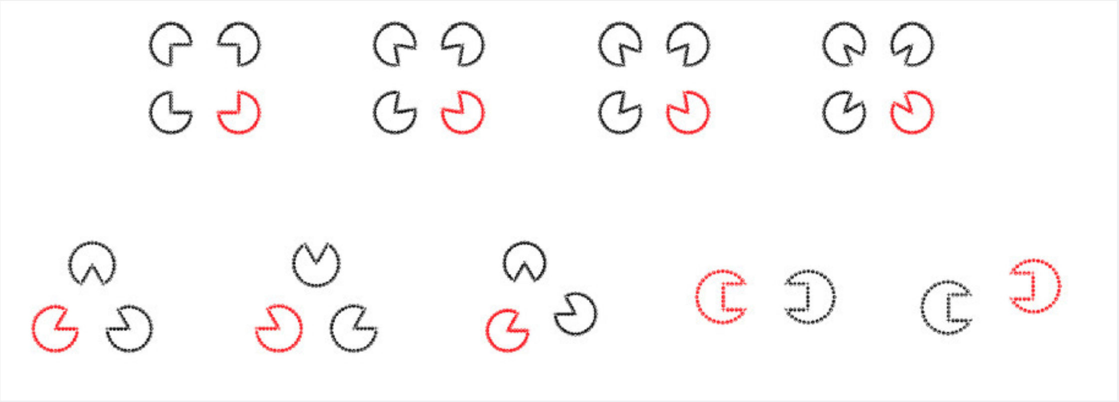


Figure 7.10: Stimulus of Figure 7.8. The results do not fit the visual perception if we use the isotropic Laplacian equation (4.2.16) and confirm the necessity to use a Sub-Riemannian kernel to model the cortical connectivity.

7.1.7 Discussion

We have presented a neurally based model for figure-ground segmentation using spectral methods, where segmentation has been performed by computing eigenvectors of affinity matrices. Different connectivity kernels that are compatible with the functional architecture of the primary visual cortex have been used. We have modelled them as fundamental solution of Fokker-Planck, Sub-Riemannian Laplacian and isotropic Laplacian equations and compared their properties. With this model we have identified perceptual units of different Kanizsa figures, showing that this can be considered a good quantitative model for the constitution of perceptual units equipped by their saliency.

We have also shown that the fundamental solutions of Fokker-Planck and Sub-Riemannian Laplacian equations are good models for the good continuation law, while the isotropic Laplacian equation is less representative for this gestalt law. However it retrieves information about ladder parallelism, a feature that can be analysed in the future. All the three kernels are able to accomplish boundary completion with a preference for the operators Fokker Planck and the Sub-Riemannian Laplacian. The proposed mathematical model is then able to integrate local and global gestalt laws as a process implemented in the functional architecture of the visual cortex. The kernel considered here only depends on orientation. Hence it can be applied to detect the saliency of geometrical figures which can be very well described using this feature.

Similar method can be applied to natural images if their main features are related to orientations, as presented in the Section 7.2 and 7.3. However for general images we can not rely on this simple geometric method, since different cortical areas can be involved in the definition of the saliency, with a modulatory effect on the connectivity of V1.

7.2 Application to medical images: emergence of percepts in retinal images

Retinal images provide early signs of diabetic retinopathy, glaucoma and hypertension. These signs can be investigated based on microaneurysms or smaller vessels. The diagnostic biomarkers are the change of vessel widths and angles especially at junctions, which are investigated using the vessel segmentation or tracking. Vessel paths may also be interrupted; crossings and bifurcations may be disconnected.

Here we apply the method presented in Section 6.1 based on the geometry of the primary visual cortex (V1) to study these difficulties. The proposed method represents an engineering application of segmenting and representing blood vessels inspired by the modeling of the visual cortex. This shows how these models can be applied to the analysis of medical images and how these two fields can be reciprocally used to better understand and reinforce each other. We have analysed the specific problems at junctions with a connectivity kernel obtained as the fundamental solution of the Fokker-Planck equation, which is usually used to represent the geometrical structure of multi-orientation cortical connectivity. By using the spectral clustering on a large local affinity matrix constructed by both the connectivity kernel and the feature of intensity, the vessels are identified successfully in a hierarchical topology each representing an individual perceptual unit.

7.2.1 Vessel extraction and its difficulties

The vasculature can be extracted by means of either pixel classification or vessel tracking. Several segmentation and tracking methods have been proposed in the literature [Fraz et al., 2012, Bühler et al., 2004, Felkel et al., 2001]. In pixel classification approaches image pixels are labeled either as vessel or non-vessel pixels. Therefore, a vessel likelihood (soft segmentation) or binary map (hard segmentation) is created for the retinal image. Although the vessel locations are estimated in these approaches, they do not provide any information about vessel connectivities. On the contrary, in tracking based approaches, several seed points are selected and the best connecting paths between them are found [Al-Diri et al., 2009, Chutatape et al., 1998, Poon et al., 2007, Quek and Kirbas, 2001, Xu et al., 2011, Can et al., 1999, De et al., 2014, 2013, Delibasis et al., 2010, González et al., 2010]. The main benefit of vessel tracking approaches is that they work at the level of a single vessel rather than a single pixel and they try to find the best path that matches the vessel profile. Therefore, the information extracted from each vessel segment (e.g. diameter and tortuosity) is more accurate and reliable.

There are several difficulties for both vessel segmentation and tracking approaches. Depending on imaging technology and conditions, these images could be affected by noise in several degrees. Moreover, non-uniform luminosity, drift in image intensity, low contrast regions and also central vessel reflex make the vessel detection and tracking complicated. Several image enhancement, normalization and denoising techniques have been developed to tackle these complications (e.g. [Abbasi-Sureshjani et al., 2015, Foracchia et al., 2005, Narasimha-Iyer et al., 2008]).

The tracking methods are often performed exploiting the skeleton of the segmented images. Thus, non-perfect segmentation or wrong skeleton extraction results in topological tracing errors e.g. disconnections and non-complete subtrees as discussed in several methods proposed in the literature [Joshi et al., 2011, Al-Diri et al., 2009, De et al., 2013, 2014, González et al., 2010]. Typical non-perfections include missing small vessels, wrongly merged parallel vessels, disconnected or broken up vessel segments and the presence of spur branches in thinning. Moreover, the greater difficulty arises at junctions and crossovers: small arteriovenous crossing angles, complex junctions when several junctions are close together, or presence of a bifurcation next to a crossing makes the centreline extraction and tracing challenging. These difficulties are mentioned as the tracking limitations in the literature. Some of these challenging cases are depicted in Figure 7.11 with their corresponding artery/vein ground truth labels. Arteries and veins are annotated in red and blue colors respectively. The green color represents the crossing and the types of the white vessels are not known.

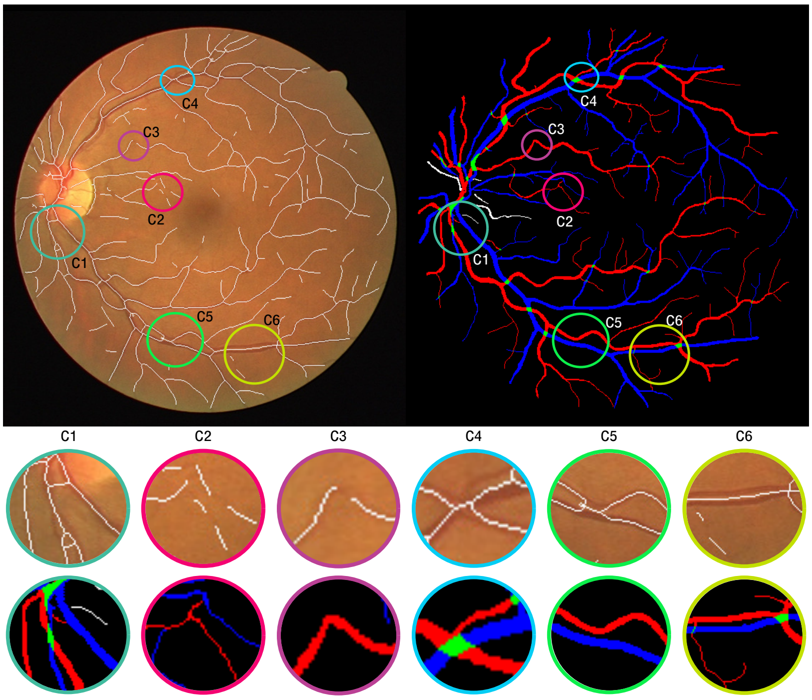


Figure 7.11: A sample image from the DRIVE dataset [Staal et al., 2004] (top left) and its corresponding artery/vein ground truth from the RITE dataset [Hu et al., 2013] (top right). Several difficult cases are shown. C1: complex junction (bifurcation and crossing with narrow crossing angle); C2: interrupted lines and missing small vessels; C3: high curvature vessel; C4: complex junction; C5: two nearby parallel vessels; C6: missing small vessel, merged parallel vessels and interrupted segment.

Our method, which is not dependent on centerline extraction, is based on the fact that in arteriovenous crossings there is a continuity in orientation and intensity of the artery and vein, respectively, i.e., the local variation of orientation and intensity of individual vessels is very low. The proposed method models the connectivity as the fundamental solution of the Fokker–Planck equation, which matches the statistical distribution of edge co-occurrence in natural images and is a good model of the cortical connectivity [Sanguinetti et al., 2008].

7.2.2 Spectral Analysis

The goal of clustering is to divide the data points into several groups such that points in the same group are similar and points in different groups are dissimilar to each other. The cognitive task of visual grouping can be considered as a form of clustering, with which it is possible to separate points in different groups according to their similarities. In order to perform visual grouping, we will use the spectral clustering algorithm. Traditional clustering algorithms, such as K-means, are not able to resolve this problem [Ng et al., 2002]. In recent years, different techniques have been presented to overcome the performance of the traditional algorithms, in particular spectral analysis techniques. It is widely known that these techniques can be used for data partitioning and image segmentation [Shi and Malik, 2000, Perona and Freeman, 1998, Weiss, 1999, Meila and Shi, 2001] and they outperform the traditional approaches. Above that, they are simple to implement and can be solved efficiently by standard linear algebra methods [Von Luxburg, 2007]. In the next section we will describe the spectral clustering algorithm used in the numerical simulations.

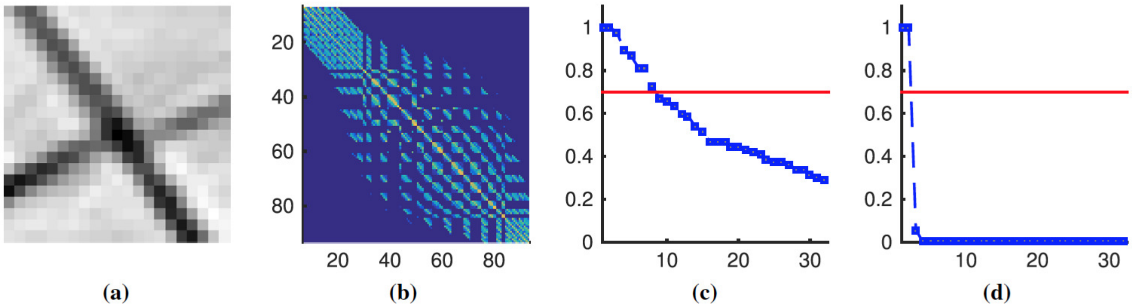
7.2.2.1 Spectral Clustering Technique

Different algorithms based on the theory of graphs have been proposed to perform clustering. In [Perona and Freeman, 1998] it has been shown how the edge weights $\{a_{ij}\}_{i,j=1,\ldots n}$ of a weighted graph describe an affinity matrix A. This matrix contains information about the correct segmentation and will identify perceptual units in the scene, where the salient objects will correspond to the eigenvectors with the highest eigenvalues. Even though it works successfully in many examples, in Weiss [1999] it has been demonstrated that this algorithm also can lead to clustering errors. In Von Luxburg [2007] and Meila and Shi [2001] the algorithm is improved considering the normalized affinity matrix. In particular we will use the normalization described in Meila and Shi [2001]. Defining the diagonal matrix D as formed by the sum of the edge weights (representing the degrees of the nodes, $d_i = \sum_{j=1}^{n} a_{ij}$), the normalized affinity matrix is obtained as:

$$P = D^{-1}A. \tag{7.2.1}$$

This stochastic matrix P represents the transition probability of a random walk in a graph. It has real eigenvalues $\{\lambda_j\}_{j=1,\ldots n}$ where $0 \leq \lambda_j \leq 1$, and its eigenvectors $\{u_i\}_{i=1,\ldots K}$, related to the K largest eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_K$, represent a solution of the clustering problem [Von Luxburg, 2007]. The value of K determines the number of eigenvalues and eigenvectors considered informative.

The important step is selecting the best value of K, which can be done by defining an a-priori significance threshold ϵ for the decreasingly ordered eigenvalues λ_i, so that $\lambda_i > 1 - \epsilon, \forall\ 1 \leq i \leq K$. However, selecting the best ϵ value is not always trivial and the clustering results get very sensitive to this parameter in many cases. Hence, considering the diffusion map approach of Coifman and Lafon [2006] and following the idea of Cocci et al. [2015], using an auxiliary diffusion parameter (τ, big positive integer value) to obtain the exponentiated spectrum $\{\lambda_i^\tau\}_{i=1,...n}$, the gap between exponentiated eigenvalues increases and sensitivity to the threshold value decreases very much. Using this new spectrum, yields to the stochastic matrix P^τ, that represents the transition matrix of a random walk in defined τ steps. The difference between thresholding the eigenvalues directly or the exponentiated spectrum is shown in an example in Figure 7.12.



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

Figure 7.12: A sample image patch at a crossing (a), its affinity matrix (A) built upon a connectivity measure (b), the eigenvalues ($\lambda_i, i = 1, ...n$) of the normalized affinity matrix (P) and the threshold value ($1 - \epsilon = 0.7$), represented in red (c). The exponentiated spectrum ($\lambda_i^\tau, i = 1, ...n$) with $\tau = 150$ and threshold value of 0.7 in red (d).

As seen in this figure, selecting the best discriminative threshold value for the eigenvectors (Figure 7.12 (c)) is not easy, while with the exponentiated spectrum (Figure 7.12 (d)) the threshold value can be selected in a wide range (e.g. $0.05 \leq 1 - \epsilon \leq 0.9$). The value of τ needs to be selected as a large positive integer number (e.g. 150).

After selecting the value of K, the number of clusters is automatically determined using Algorithm **??**.

Possible neural implementations of the algorithm are discussed in Cocci et al. [2015]. Particularly, in Bressloff et al. [2002], Faugeras et al. [2009] an implementation of the spectral analysis is described as a mean-field neural computation. Principal eigenvectors emerge as symmetry breaking of the stationary solutions of mean field equations. In addition, in Sarti

Algorithm 1 Spectral clustering algorithm

```
1: Define the affinity matrix $A_{i,j}$ from the connectivity kernel.
2: Evaluate the normalized affinity matrix: $P = D^{-1}A$.
3: Solve the eigenvalue problem $Pu_i = \lambda_i u_i$, where the order of $i$ is such
   that $\lambda_i$ is decreasing.
4: Define the thresholds $\epsilon$, $\tau$ and evaluate the largest integer K such that
   $\lambda_i^\tau > 1 - \epsilon, i = 1, \ldots, K$.
5: Let U be the matrix containing the vectors $u_1, \ldots, u_K$ as columns.
6: Define the clusters k = $\arg\max_j \{u_j(i)\}$ with $j \in \{1, \ldots, K\}$ and $i =
   1, \ldots, n$.
7: Find and remove the clusters that contain less than a minimum cluster
   size elements.
```

??.

and Citti [2015] it is shown that in the presence of a visual stimulus the emerging eigenvectors are linked to visual perceptual units, obtained from a spectral clustering on excited connectivity kernels. In the next section the application of this algorithm in obtaining the vessel clustering in retinal images will be presented.

7.2.3 Proposed Technique

In order to prove the reliability of the method in retrieving the connectivity information in 2D retinal images, several challenging and problematic image patches around junctions were selected. First step before detecting the junctions and selecting the image patches around them, is to apply preconditioning on the green channel (I) of a color fundus retinal image. The green channel provides a higher contrast between vessels and background and it is widely used in retinal image analysis. The preconditioning includes: *a*) removing the non-uniform luminosity and contrast variability using the method proposed by Foracchia et al. [2005]; *b*) removing the high frequency contents; and *c*) denoising using the non-linear enhancement in SE(2) as proposed by Abbasi-Sureshjani et al. [2015] . A sample color image before and after preconditioning (I_{enh}) are shown in Figure 7.13 (a) and (b) respectively.

In next step, soft (I_{soft}) and hard (I_{hard}) segmentations are obtained using the BIMSO (biologically-inspired multi-scale and multi-orientation) method for segmenting I_{enh} as proposed by [Abbasi-Sureshjani et al., 2015]. These images are shown in Figure 7.13 (c) and (d) respectively. The hard segmentation is used for detecting the junctions and selecting several patches with different sizes around them; while soft segmentation is used later in connectivity analysis.

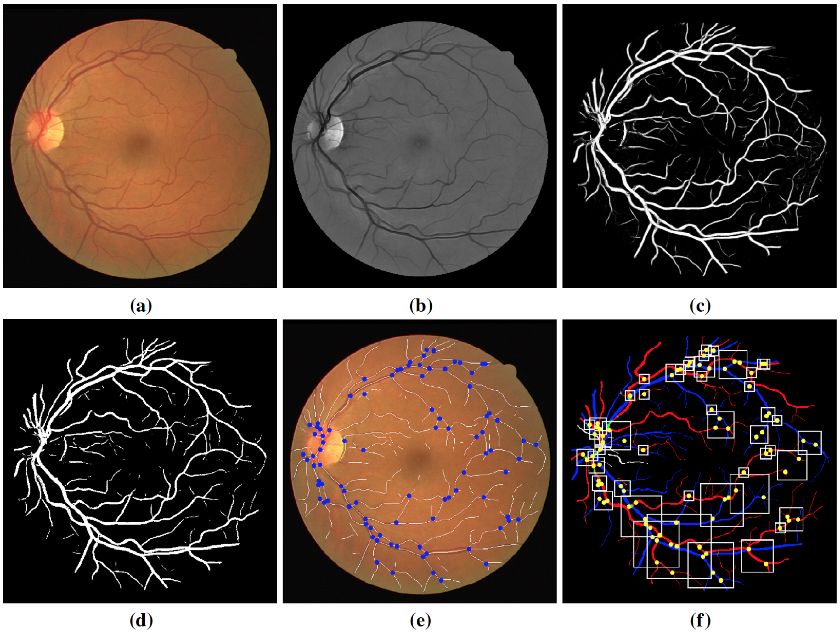


Figure 7.13: The different steps applied for selecting several image patches around junctions, (a) original RGB image, (b) enhanced image (I_{enh}), (c) soft segmentation (I_{soft}), (d) hard segmentation (I_{hard}), (e) detected junctions and the skeleton of the segmentation overlaid on color image, (f) selected patches overlaid on artery/vein ground truth.

In order to find the junction locations automatically, the skeleton of I_{hard} is produced using the morphological skeletonization technique. Then the method proposed by Olsen et al. [2011] is applied on this skeleton and the junction locations are determined as shown in Figure 7.13 (e). Using the determined locations, several image patches with similar sizes (s = 10 pixels) are selected at first stage. However, as seen in Figure 7.13 (e), some of the junctions are very close to each other and their distances are smaller than $s/2$. For these junctions, a new patch including both nearby junctions (with a size equal to three times the distance between them) is considered, and its centre is used for finding the distance of this new patch with the other ones. These steps are repeated until no more merging is possible or the patch size reaches the maximum possible size (we assumed 100 as the maximum possible value). Thus, all nearby junctions are grouped in order to decrease the number of patches that overlap in a great extent. This results in having different patch sizes ($0 \leq s_{p_i} \leq 100, 1 \leq i \leq m$) that could include more than one junction all over the image. Figure 7.13 (f) shows the junction locations and the corresponding selected patches overlaying on artery/vein ground truth.

In order to analyze the vessel connectivities for each image patch (I_{p_i}), we need to extract the location (x, y), orientation (θ) and intensity $(f(x, y))$ of vessel pixels in these patches. Hence for each group of junctions (i) with the size s_i, two patches from

I_{enh} and I_{soft} are selected, called I_{enh,p_i} and I_{soft,p_i} respectively. Then I_{soft,p_i} is thresholded locally to obtain a new hard segmented image patch (called I_{hard,p_i}). This new segmented image patch is different from selecting the corresponding patch from I_{hard}, because I_{hard} was obtained by thresholding the entire I_{soft} using one global threshold value, but this is not appropriate at all regions. If there are regions with very small vessels with a low contrast (often they get a very low probability of being vessel pixels), they are normally removed in the global thresholding approach. Accordingly, wrong thresholding leads into wrong tracking results e.g. C1, C2, C6 in Figure 7.11 are some instance patches with missing small vessels. In this work, we selected one threshold value for each patch specifically using Otsu's method [Otsu, 1975], to keep more information and cover a wider range of vessel pixels. Consequently, thicker vessels will be created in I_{hard,p_i} and the results will be more accurate.

By knowing the vessel locations (x, y) other information could be extracted for these locations using I_{enh,p_i}. So $f(x, y)$ equals the intensity value in I_{enh,p_i} at location (x, y). Moreover, by lifting I_{enh,p_i} using cake wavelets, at each location the angle corresponding to the maximum of the negative orientation response (real part) in the lifted domain is considered as the dominant orientation (θ_d) as equation (7.2.2):

$$\theta_d = \underset{\theta \in [0,\pi]}{\arg\max}\, Re(-U_f(x, y, \theta)) \tag{7.2.2}$$

The negative response is considered because the blood vessels in retinal images are darker than background. Next step is approximating the connectivity kernels as shown in Section 6.1. The first kernel (ω_1), was calculated numerically, so the fundamental solution Γ_1 was estimated using the Markov Chain Monte Carlo method [Robert and Casella, 2013]. This finite difference equation is solved for n (typically 10^5) times, so n paths are created. Then the estimated kernel is obtained by averaging all the solutions [Higham, 2001, Sarti and Citti, 2015]. An overview of different possible numerical methods to compute the kernel is explained in Zhang et al. [2016], where comparisons are done with the exact solutions derived in Duits and Franken [2010a], Duits and Van Almsick [2008]. From these comparisons it follows that the stochastic Monte-Carlo implementation is a fair and accurate method. The intensity-based kernel (ω_2), the final connectivity kernel (ω_f) and the affinity matrix (A), were calculated using equations (6.1.6), (6.1.7) and

(6.1.8) respectively. Finally, by applying the proposed spectral clustering step in Section 7.2.2, the final perceptual units (individual vessels) were obtained for each patch.

The above-mentioned steps for a sample crossing in a 21 × 21 image patch are presented in Figure 7.14.

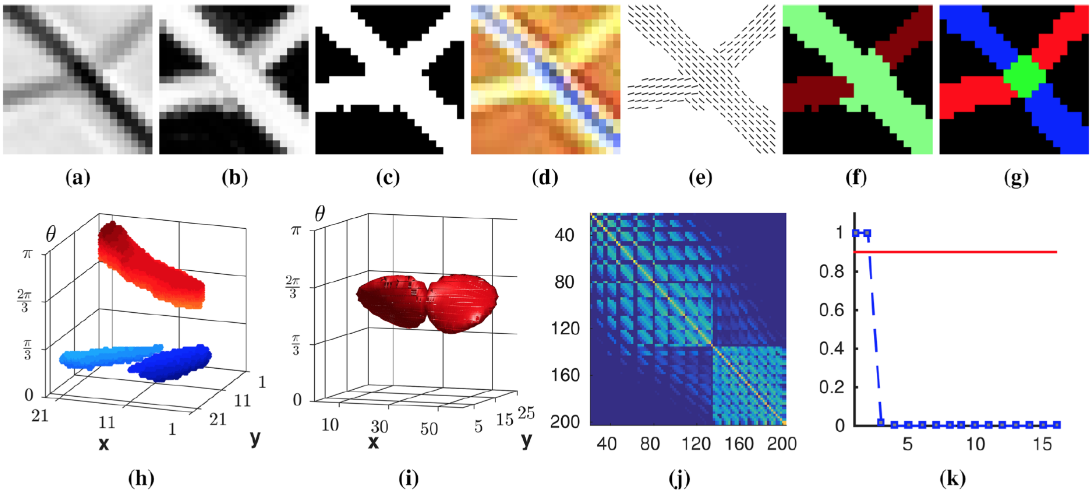


Figure 7.14: A sample 21 × 21 image patch at a crossing: (a) I_{enh,p_i}, (b) I_{soft,p_i}, (c) I_{hard,p_i}, (d) the differences in intensities are shown in color, (e) each oriented line represents the orientation at its position, (f) final perceptual units shown in different colors (g) the ground truth artery and vein labels, (h) lifted image in $SE(2)$, (i) connectivity kernel (ω_2), (j) affinity matrix (A) obtained using both orientation and intensity information, (k) thresholding the eigenvalues of the normalized affinity matrix.

After enhancing the image (Figure 7.14 (a)), obtaining soft segmentation (Figure 7.14 (b)) and thresholding it locally (Figure 7.14 (c)), the vessel locations, intensity and orientation have been extracted. As shown in Figure 7.14 (d) arteries and veins have different intensities and this difference helps in discriminating between them. Though, orientation information is the most discriminative one. The lifted image in $SE(2)$ using the π-periodic cake wavelets in 24 different orientations is shown in Figure 7.14 (h). The disentanglement of two crossing vessels at the junction point can be seen clearly in this figure. The dominant orientations (θ_d) for the vessel pixels are also depicted in Figure 7.14 (e), using line segments oriented according to the corresponding orientation at each pixel.

In the next step, this contextual information (intensity and orientation) is used for calculating the connectivity kernel (Figure 7.14 (i)) and the affinity matrix (Figure 7.14 (j)). For this numerical simulation, H, n, σ and σ_2 have been set to 7, 100000, 0.05 and 0.1 respectively. Next, by applying the spectral clustering on the normalized affinity matrix using ϵ and τ as 0.1 and 150, only two eigenvalues above the threshold will remain (Figure 7.14 (k)). This means that there are two main salient perceptual units in this image as it was expected. These two units are color coded in Figure 7.14 (f). The corresponding artery and vein labels are also depicted in Figure 7.14 (g) which approve the correctness of the obtained clustering results.

7.2.4 Validation

To validate the method, the proposed steps were applied on several image patches of the DRIVE [Staal et al., 2004] dataset. This public dataset contains 40 color images with a resolution of 565×584 ($\sim\ 25\mu m/px$) and a 45° field of view. The selected patches from each image were manually categorized into the following groups: simple crossing (category A), simple bifurcation (category B), nearby parallel vessels with bifurcation (category C), bifurcation next to a crossing (category D) and multiple bifurcations (category E), and each category narrowed down to 20 image patches. These patches have different complexities, number of junctions and sizes and they could contain broken lines, missing small vessels and vessels with high curvature. The parameters used in the numerical simulation of the affinity matrix and spectral clustering step (including σ, H, n, σ, σ_2, ϵ and τ) are chosen for each patch differently, with the aim of achieving the optimal results for each case. Automatic parameter selection remains a challenging task and will be investigated in future work.

Some sample figures of these cases are depicted in Figure 7.15. For each example, the original gray scale enhanced image, hard segmentation (locally thresholded), orientation and intensity information, and finally the clustering result together with artery/vein labels are depicted (Figure 7.15 (a)-(f) respectively). Although the complexity of these patches is quite different in all cases, the salient groups are detected successfully. All the vessel pixels grouped as one unit have similarity in their orientations and intensities, and they follow the law of good continuation. Therefore, at each bifurcation or crossover point, two groups have been detected.

In this figure, G1 is a good example of a crossing with a small angle. The method not only differentiates well between vessels crossing each other even with a small crossing angle, but it also determines the order of vessels, being at the bottom or passing over in crossover regions. The image patch in G2 is a good example showing the strength of the method in detecting small vessels. The detected small vessel in this image is even not annotated in the artery/vein ground truth. However, this detection is highly dependent on the soft segmented image and the threshold value used for obtaining the hard segmentation.

If the small vessel is not detected in the soft segmentation or if a hight threshold value is selected, then it also will not be available in the final result.

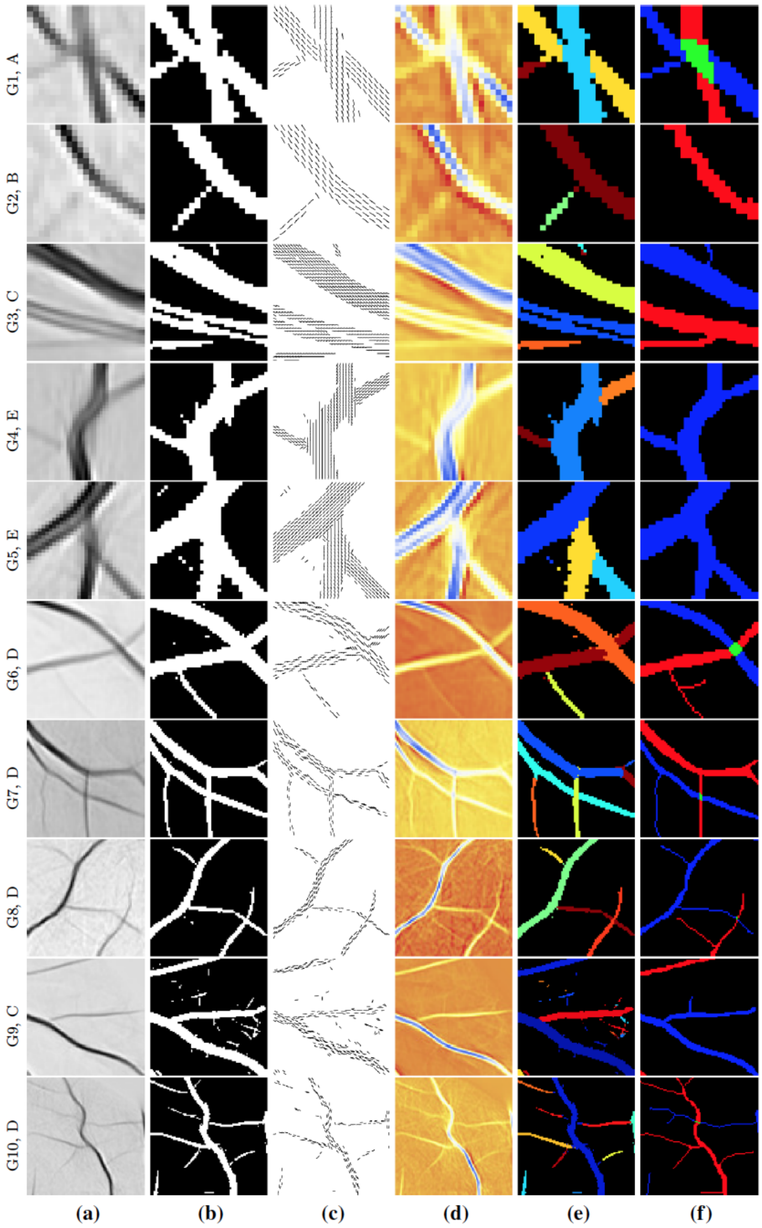


Figure 7.15: Sample image patches selected from the DRIVE dataset. Columns from left to right present the image patch at the green channel, segmented image, extracted orientation and intensity, clustering result and the artery/vein labels.

Other cases in this figure are good representations of the robustness of the method against the presence of a central vessel reflex (as in G3), interrupted lines (as in G10) or even noise (as in G9). In G9, noisy pixels are detected as individual units which are not similar to the other groups. They can be differentiated from others based on their sizes. If there are very few pixels in one group, then it can be considered as noise and removed. There are also several cases with complex junctions in this figure. Presence of multiple bifurcations in one image patch, or presence of several bifurcations close to the crossing points does not lead to wrong grouping results (as seen in G5, G6, G7, G8 and G10).

The parameters used during the numerical simulations of the image patches shown in Figure 7.15 and their corresponding sizes are presented in Table 7.1. For all experiments the values of n, ϵ and τ were set to 100000, 0.1 and 150 respectively and they remained constant.

Name	size	H	σ	σ_2
G1	21×21	7	0.02	0.3
G2	21×21	8	0.03	0.3
G3	41×41	10	0.03	0.1
G4	39×39	9	0.03	0.3
G5	33×33	8	0.03	0.3
G6	51×51	20	0.03	0.3
G7	71×71	17	0.07	0.3
G8	73×73	24	0.03	0.3
G9	89×89	30	0.03	0.3
G10	97×97	24	0.03	0.3

Table 7.1: The parameters used in numerical simulation of the image patches shown in Figure 7.15 and their corresponding sizes

The key parameters which are very effective in the final results are H, σ and σ_2. H and σ determine the shape of the kernel. Based on the experiments, the appropriate value for the number of steps of the random path generation is approximately $1/3$ of the image width. Selecting this parameter correctly is very important in connecting the interrupted lines. The parameters σ and σ_2 which determine the propagation variance in the θ direction and the effect of the intensity-based similarity term do not have a large sensitivity to variation. To quantify this, the mean and variance of these two parameters for each of the above-mentioned categories are calculated and presented in Table 7.2. Since the selected patches have varying sizes and H is dependent on that, this parameter is not presented in this table. Moreover, to evaluate the performance of the method, we introduced the correct detection rate (*CDR*) as the percentage of correctly grouped image patches for each category. These values are presented in Table 7.2. By considering higher number of image patches per category the *CDR* values will be more realistic.

7.2.5 Discussion

The proposed method allows finding accurate junction positions, which is the position where two groups meet or cross

2*Category	2*CDR%	σ		σ_2	
		mean	variance	mean	variance
A	85	0.032	0.0001	0.28	0.0039
B	95	0.033	$\simeq 0$	0.3	$\simeq 0$
C	85	0.0269	$\simeq 0$	0.22	0.01
D	75	0.035	0.00013	0.248	0.0125
E	95	0.03	$\simeq 0$	0.3	$\simeq 0$

Table 7.2: The correct detection rate and the mean and variance of σ and σ_2 used in numerical simulation for each category

each other. The main application of these connectivity analyses would be in modeling the retinal vasculature as a set of tree networks. Our method represents some limitations at blood vessels with high curvature: if there are some high curvature vessels, then depending on their curvature increasing σ might help in preserving the continuity of the vessel. As an example, G4 in Figure 7.15 is relatively more curved compared to the other cases, but the clustering works perfectly in this case. However, for some cases it does not solve the problem totally, and other kernels need to be considered for preserving the continuity. An example 49×49 image patch with a highly curved vessel is shown in Figure 7.16, where the method fails in clustering the vessels correctly. The parameters used for this case are $H = 16$, $\sigma = 0.03$ and $\sigma_2 = 0.3$

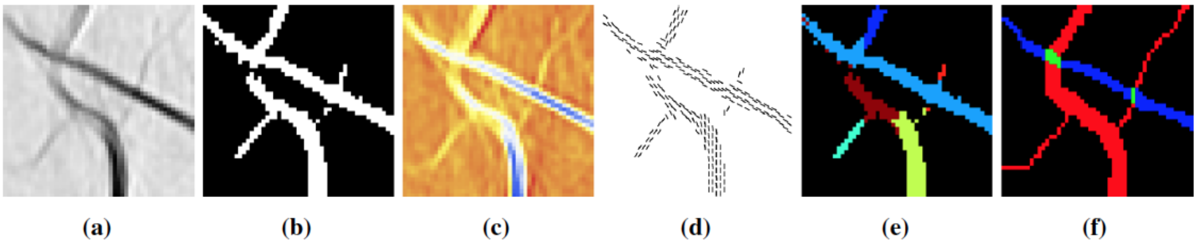

(a) (b) (c) (d) (e) (f)

Figure 7.16: Failure of clustering in presence of highly curved vessels. Columns from left to right: (a) enhanced image; (b) its segmentation; (c) orientation and (d) intensity information; (e) clustering result and (f) the artery/vein labels.

Even though the intensities of arteries and veins in the gray scale enhanced image are very close to each other in some images, adding the intensity term in calculating the final affinity matrix is crucial. By decreasing the value of σ_2, the distance between intensities gets a higher value and it helps in differentiating better between the groups. Figure 7.17 represents a sample 67×67 image patch, which includes two nearby parallel vessels with similar orientations. Figure 7.17 (e) and (f) show the correct and wrong clustering results obtained by changing σ_2 from 0.3 to 1. All other parameters have not changed ($H = 24$, $\sigma = 0.02$ and $n = 100000$). The other important difference between these two results is that the noisy pixels close to the thicker vessel have been totally removed in the correct result.

Although they seem to be oriented with the thick vessel their intensities are totally different. Therefore, by increasing the effect of intensity, they are clustered as several small groups and removed in the final step of the spectral clustering algorithm.

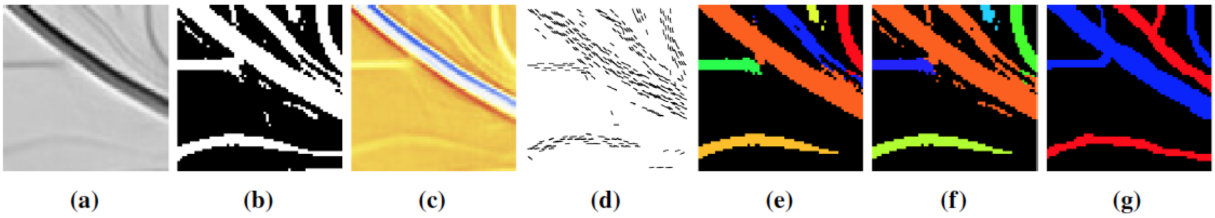


Figure 7.17: The effect of including intensity term in calculating the connectivity kernel. Columns from left to right: (a) enhanced image; (b) its segmentation; (c) orientation and (d) intensity information; (e) correct and (f) wrong clustering results; and (g) the artery/vein labels.

In the next Section we will propose a solution to the limitations described, showing kernels that take into account the curvature of structures in addition to positions and orientations.

7.3 Analysis of vessel connectivities in retinal images - Curvature Integration

In Section 7.2, following Favali et al. [2016a], we have investigated the connectivity analysis in retinal images specifically at junction points, inspired by the mathematical modeling of the geometry of the primary visual cortex (V1). In contrast to the state-of-the-art techniques, the proposed approach was capable of grouping and separating the blood vessels as individual perceptual units, even though there was some information missing due to poor segmentations. It could find the right connections between small vessels and their parents, which are usually missing in the literature and removed during preprocessing, segmentation or a skeleton pruning step. It also individuated nearby parallel vessels, even with presence of a central vessel reflection. The bottleneck of this method was that it was not data-adaptive and it could not follow some of the highly curved vessels, because the introduced kernel was elongated and could not bend as much as the vessel bends at these points.

7.3.1 Cortically-inspired Spectral Clustering

In this Section we show the results presented in Abbasi-Sureshjani et al. [2016a], where we introduce three main novelties:

1) we introduce a new feature detection lifting process to select curvature, extending the curvature extraction technique

to a multiscale approach making it suitable for analysis of multiscale blood vessels;

2) we describe the five-dimensional kernel and the affinity matrix in the lifted space of position, curvature, orientation and intensity;

3) we apply a very efficient second generation clustering algorithm, able to automatically estimate the optimal number of cluster in the grouping process.

7.3.1.1 Features extraction and lifting

In presence of an image, we need to extract the features of orientation and curvature. Orientation at every point can be computed as described in Section 4.1.2.2. It can be selected using Gabor filters, or a family of filters with similar properties, called cake wavelets and introduced by Bekkers et al. [2014]. Calling U_I the output of the filters:

$$-U_I(x, y, \theta) = \int \phi_{x,y,\theta}(x', y')I(x', y'),$$

the orientation at which the maximum response is obtained is assigned to each location. It means the dominant orientation θ_d at location (x, y) is defined as

$$\theta(x, y) = \underset{\theta \in [0,\pi]}{\arg\max}\, Re(-U_I(x, y, \theta)).$$

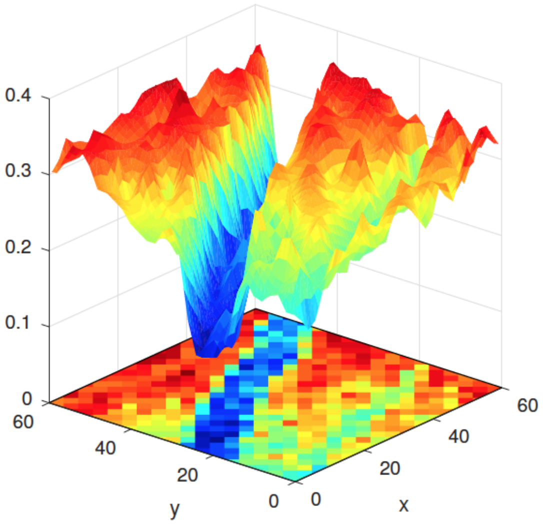


Figure 7.18: The intensity of a sample image patch (I_i', shown in the xy plane) is projected to the z coordinate to depict the Gaussian profile of a sample blood vessel.

In order to evaluate the curvature map for each location $\kappa(x, y)$ we generalize a curvature measurement used by [Bekkers et al., 2015]. This is based is based on the eigensystem analysis of the Gaussian Hessian in the lifted domain. The method performs the best when the scale of the Gaussian filters σ matches the vessel width. By using a single scale, the curvature values are accurate only for the vessels which their width match the used scale. Since the blood vessels in retinal images have different widths, we modify the algorithm and introduce a multi-scale approach which helps in covering various vessel widths available in the image and getting more accurate results. The idea is similar to the multi-scale feature extraction step of the BIMSO segmentation method [Abbasi-Sureshjani et al., 2015], since there the Gaussian filters have been used in the lifted domain as well. Therefore, we altered the method to a multi-scale approach. To do so, confidence and curvature maps have been obtained for several scales ($s_{\sigma_i}(x, y, \theta)$ and $\kappa_{\sigma_i}(x, y, \theta)$, $i = 1, \ldots, n$). Then the curvature map is constructed by assigning to each pixel the curvature value that corresponds to the scale at which the largest confidence value has been obtained, i.e. $\forall (x, y, \theta) \in \mathbb{R}^2 \times S^1$:

$$\kappa(x, y, \theta) = \{\kappa_{\sigma_{max}}(x, y, \theta) | \sigma_{max} = \underset{\sigma_i \in \{\sigma_1 \dots \sigma_n\}}{\arg\max}\ s_{\sigma_i}(x, y, \theta)\}. \tag{7.3.1}$$

Finally we add the feature of intensity, which is independent of the previous ones.

This curvature map is used in Step 3 of Algorithm 2.

7.3.1.2 The new affinity matrix

We can define a new affinity matrix starting from the cortical connectivity kernel. In Sections 4.4 and 6.1, following Favali et al. [2016b,a], Sarti and Citti [2015] we used 3D and 4D connectivity kernels to define corresponding affinity matrices. Similarly, the 5D kernel defined in section 6.2.3 will be used here for defining a 5D affinity matrix.

According to the teoretical description this matrix will be found in two step: first we work in the 4D position, orientation and curvature space. Then we add intensity. In other words we define an additional affinity matrix in order to include the intensity of vessel pixels as a Gaussian weighting. The values of this affinity matrix for each pair of pixels (ϵ_i and ϵ_j) in patch k is calculated as follows:

$$A'_{ij} = e^{\frac{-(f_i - f_j)^2}{\sigma_{int}^2}} \tag{7.3.2}$$

where the intensity values (f_i and f_j) are obtained from the preprocessed image patch (I_k') at corresponding locations and the σ_{int} parameter controls the effectiveness of the intensity similarity. This affinity matrix is normalized (between 0 and 1) and then it is included in the final affinity matrix at Step 6 of Algorithm 2.

As explained above, this cortically inspired kernels and affinity matrices could be potentially augmented with other affinity kernels or matrices including information about other features (width, color), as long as these depend on independent variables.

A summary of the method is presented in Algorithm 2.

Algorithm 2 Proposed perceptual grouping technique for a given image $I \in \mathbb{R}^2$

```
1: Lift the image $I(x,y) \in \mathbb{R}^2$ to $U_I(x,y,\theta) \in \mathbb{R}^2 \times S^1$ by orientation score transform.
2: Calculate the curvature map $\kappa_{map}(x,y,\theta)$ for each point in $U_I(x,y,\theta)$ using the method proposed by Bekkers et al. [2015].
3: Lift the image to $\mathbb{R}^2 \times S^1 \times \mathbb{R}$ so that $U_I'(x,y,\theta,\kappa) = U_I(x,y,\theta)$ if $k_{map}(x,y,\theta) = \kappa$, $\forall (x,y,\theta) \in \mathbb{R}^2 \times S^1$.
4: Calculate all the fundamental solutions of equation 6.2.13, $\Gamma_\kappa'((x,y,\theta),(x',y',\theta'))$ and $\Gamma_{\kappa'}'((x',y',\theta'),(x,y,\theta))$ $\forall$ $\{(x,y,\theta,\kappa),(x',y',\theta',\kappa')\} \in U_I'$ stochastically.
5: Calculate the connectivity kernel $w((x,y,\theta,\kappa),(x',y',\theta',\kappa'))$ for all pairs of points in $U_I'$ and the 5D connectivity kernel $\omega_5((x,y,\theta,f,\kappa),(x',y',\theta',f',\kappa'))$ using Eq. 6.2.15.
6: Create the affinity matrix based on Eq. 7.3.2.
```

The numerical approximations of the 4-D kernel is presented in Section 6.2.3.

7.3.1.3 The optimal clustering algorithm

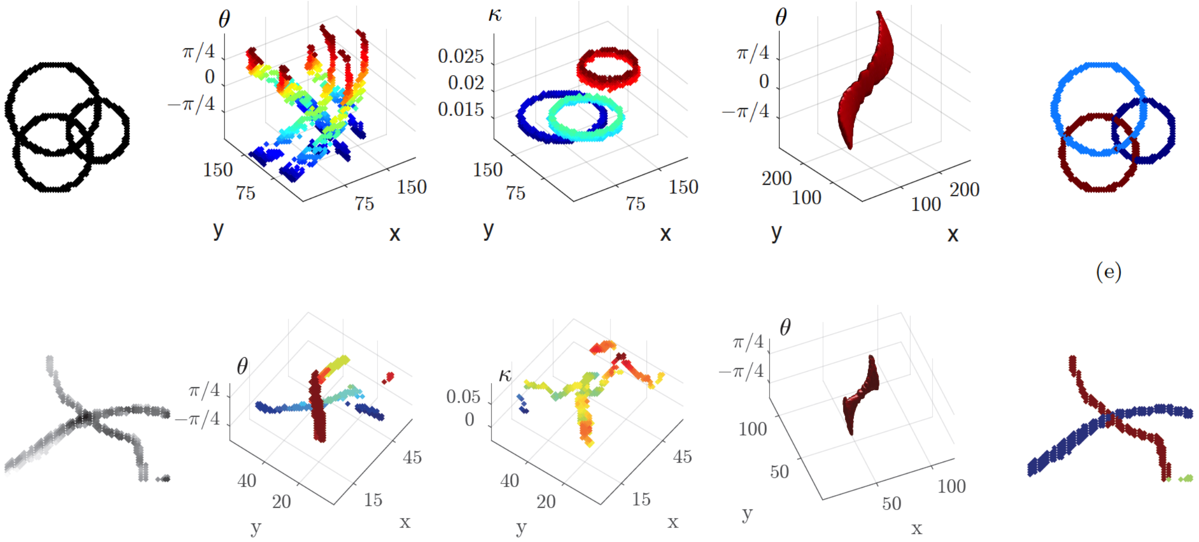
After defining a suitable affinity matrix we apply now a clustering method. Many algorithms have been proposed in this area [Shi and Malik, 2000, Perona and Freeman, 1998, Weiss, 1999, Meila and Shi, 2001, Ng et al., 2002]). Unlike more basic clustering methods, as k-means, where the number of clusters must be assumed a priori, spectral clustering algorithms select from an affinity matrix all eigenvectors whose eigenvalues are sufficiently large. In order to avoid an a priori fixed treshold (as we did in our previous paper [Favali et al., 2016a]), we follow the method proposed by Zelnik-Manor and Perona [2005], which algorithm automatically estimates the number of cluster which optimize grouping. In this algorithm, which is explained step by step in Algorithm 3, we use the affinity matrix defined in Algorithm 2. The structure of the eigenvectors is used to determine the number of groups. The cost function is evaluated from the alignment of the eigenvectors (Step 5). The best number of clusters is considered as the one which minimizes the cost function (J_{min}). Correspondingly, the best clustering quality Q_{clust}, that has a reverse relation to the alignment cost, is obtained in Step 6. In the final step, the noisy elements, that construct small sized groups, are removed.

Algorithm 3 Self-tuning spectral clustering algorithm: Given a set of points $S = s_1, ..., s_n \in R^l$ to cluster

```
1: Define the affinity matrix $A_{i,j}$.
2: Define the diagonal matrix $D$ with $D_{i,i} = \sum_{j=1}^{n} A_{i,j}$ and construct the symmetric normalized graph Laplacian $L = D^{-1/2} A D^{-1/2}$.
3: Find the $C$ largest eigenvectors of $L$ $x_1, ..., x_C$ and construct the matrix $X = [x_1, ..., x_C]$ where $C$ is the highest possible group number.
4: Use a gradient descent scheme to recover the rotation $R$ which best aligns $X$'s columns with the canonical coordinate system.
5: Let Z be the matrix obtained after rotating the eigenvector matrix $X$ and $M_i = max_j Z_{i,j}$. The cost function is defined as: $J = \sum_{i=1}^{n} \sum_{j=1}^{C} \frac{Z_{i,j}^2}{M_i^2}$.
6: Classify the cost of alignment for each group number and set the final group number $C_{best}$ to be the largest group number that provides the minimal cost $J_{min}$. Correspondingly, the best clustering quality is calculated as: $Q_{clust} = 1 - \frac{(J_{min}/n-1)}{C_{best}}$.
7: Consider the alignment result $Z$ of the top $C_{best}$ eigenvectors and assign the points $s_i$ to cluster $c$ if and only if $max_j(Z_{i,j}^2) = Z_{i,c}^2$.
8: Find and remove the clusters that contain less than a minimum number of cluster elements.
```

The scales we used for our SLO images are $\{1.5, 2.5, 3.5\}$ in pixels. Figure 7.20 represents the color-coded orientation (e), confidence (f) and curvature (g) maps of a sample SLO image of the IOSTAR dataset. Note that the depicted confidence and curvature maps are related to one single scale $\sigma = 1.5$, and the absolute curvature value is shown.

Figure 7.19 depicts a sample application of the proposed method for clustering the perceptual units in both an artificial image (the first row) and a small patch of retinal images (the second row). The synthetic image includes three crossing circles with different radii and corresponding curvatures, and the retinal patch includes two crossing vessels. For each case, the orientation and curvature of the lifted images are shown in the column (b) and (c) respectively. The intensity value for the synthetic image is constant all over the circles and for the retinal patch, it is color-coded in (a). A sample level set of corresponding 5D kernels (while keeping two dimensions (κ and f) fixed) has been shown in each row (d). Finally, (e) shows the three detected groups in the synthetic image and three vessels in the retinal image in different colors. The implementation details, validation and application of this proposed method using a set of artificial and retinal images are presented in the next section.



(e)

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

Figure 7.19: A synthetic image consisted of three crossing circles (first row) and a sample retinal image patch including two crossing vessels (the second row). The columns represent the original image, the orientation score, the 3D curvature maps, a level set of the 5D kernel while keeping two dimensions fixed and the final detected clusters in different colors, respectively.

7.3.2 Experiments

In this section, we present a potential application of the proposed connectivity analysis for solving the aforementioned problems in curvilinear structure tracking methods for retinal vasculature analysis. After explaining the material used for validating the method, the details of numerical simulation are described. Then the quantitative and qualitative results of the proposed technique are presented and discussed in detail. Two datasets have been used for validating the method. The specifications and the preparation steps of each dataset are explained in detail as follows.

7.3.2.1 Phantom images

The set of phantom images (201×201) has been generated to include various rotated, curved and interrupted vessel-like structures. The orientation and curvature values in these phantom images are known. Five different groups are created to mimic possible structures that could be present in retinal images, similar to the categories proposed in Favali et al. [2016a]. These categories are (A) crossings; (B) bifurcations; (C) parallel vessels; (D) bifurcations and crossings; and (E) vessels with multiple nearby bifurcations. Each of these categories may also include challenging structures. For instance, they may be interrupted or highly curved or include small junction (crossing/ bifurcation) angles. In order to differentiate between the simple and the challenging cases for each category, we name group X as $X1$ if it is challenging. Figure 7.21 depicts ten different phantom images (first column), two per category, together with their color-coded orientation and curvature maps (second and third columns).

The basic element used for creating these phantom images is a sine wave-like structure, which is generated with several frequencies and amplitudes and it is rotated and located at different positions depending on the target shape. By adjusting the frequency and amplitude of the waves, different curvature values can be created. In addition to the vessel-like structures, other challenging structures such as dashing and the Euler spiral have been also used to examine the strength of the method in grouping these curved structures (e.g. Figure 7.21, A1).

7.3.2.2 Retinal images

This set contains several image patches selected around junctions in the public IOSTAR[1] dataset. The IOSTAR dataset contains images captured using scanning electron ophthalmoscope (SLO) technology. These high contrast images have a resolution of 1024 × 1024 with 45° FOV. The blood vessels, junctions and artery/vein labels have been annotated for 24 images and corrected by two different experts in order to decrease the inter-user variability.

The strength of the method is its ability to individuate junction or crossing where most of the previous methods in the literature have difficulties. Therefore, five images and their annotations are downsampled to half size 512 × 512 and all patches of 51 × 51 pixels with these type of problems have been automatically selected from them. A total of 272 patches have been extracted and classified. As for phantom images, these patches are also manually categorized in five different groups (similar to the phantom images) depending on their structure and com-

[1] Available at: http://www.retinacheck.org/datasets

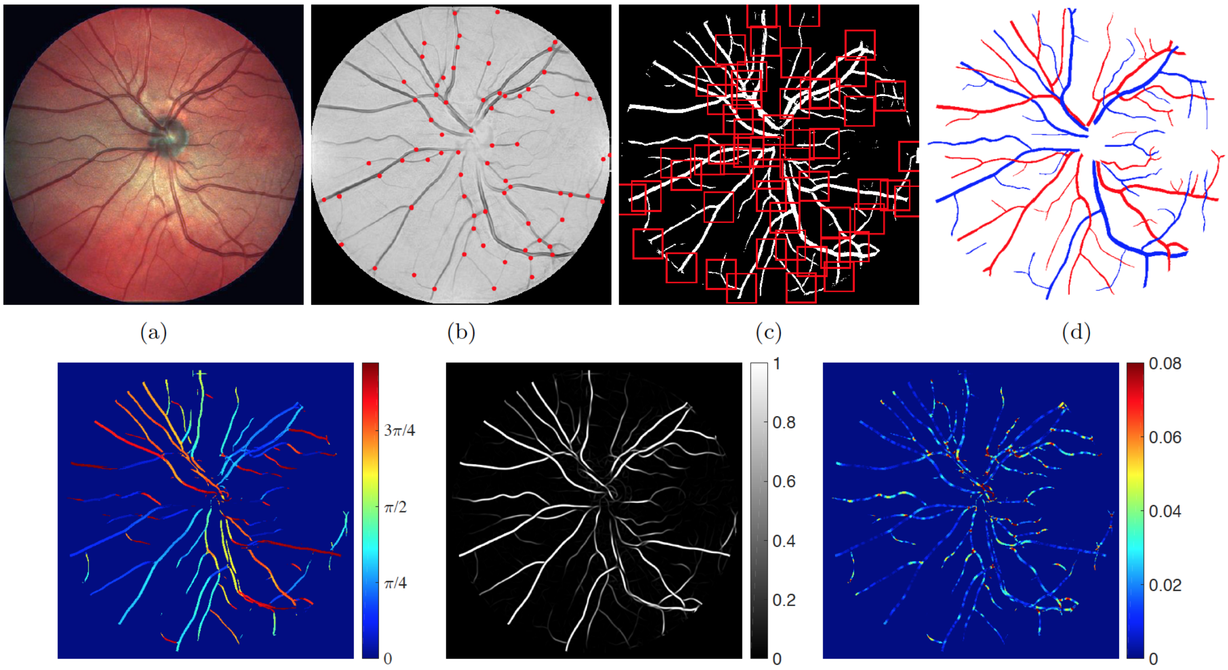
plexity, with the scope of validation of the method.

The proposed method used for creating these patches is explained step by step in Section 7.2.3. The first step is the pre-processing applied on original image ($I_\circ \in \mathbb{R}^2$), so that the luminosity and contrast are normalized, the noise is removed and the blood vessels are enhanced. The same pre-processing technique proposed by [Abbasi-Sureshjani et al., 2015] for the SLO images has been applied on the green ($I_{\circ,g}$) and red ($I_{\circ,r}$) channels separately and then they are combined as $I(x,y) = \sqrt{I^2_{\circ,g}(x,y) + I^2_{\circ,r}(x,y)}$, for all the (x,y) positions in the image $I_\circ$. This increases the difference between the intensity of arteries and veins. The next step is the blood vessel segmentation which provides an initial estimation of the location of the blood vessels. The outcome of segmentation can be a deterministic binary map, in which each pixel is labeled as a vessel (with label 1) or background (label 0). The binary map is often obtained by thresholding the probability map globally. None of the segmentations are perfect and they may contain vessel disconnections or wrongly detected vessel pixels. In this work, the binary segmentations obtained by the BIMSO method proposed by Abbasi-Sureshjani et al. [2015] is used, specifically for the SLO images.

Later a set of junction locations ($\epsilon_i = \{x_i, y_i\},\ i = 1, \dots, M$) is obtained using the BICROS method proposed by Abbasi-Sureshjani et al. [2016b], where M is the number of detected junctions. The obtained segmentation (I_{seg}) is also used as the input to the hybrid step of the BICROS method. Both these methods (BIMSO and BICROS) have been validated before using the IOSTAR dataset. The detected locations are then considered as the centre of patches with a fixed size of 51×51 ($s_\circ = 25$). For an image $H \in \mathbb{R}^2$ the patch (H'_i) centred around junction ϵ_i is selected as:

$$H'_i = \{H(x,y) | x_i - s_\circ \leq x \leq x_i + s_\circ, y_i - s_\circ \leq y \leq y_i + s_\circ\}. \tag{7.3.3}$$

The same operation is applied on the enhanced (I), segmented (I_{seg}) and ground truth (G) images and the corresponding patches are called I'_i, $I'_{seg,i}$ and G'_i (for all $i, 1 \leq i \leq M$) respectively. Figure 7.20 represents a sample SLO image ($I_\circ$), its artery/vein ground truth G (the arteries in red and the veins in blue), enhanced (I) and vessel segmented (I_{seg}) images, and the detected junctions (ϵ). The 51×51 selected patches overlaid on the hard segmentation are shown as well.



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

(e) (f) (g)

Figure 7.20: A sample SLO image, (a) original image (I_o), (b) detected junctions overlaid on the enhanced image (I_{enh}), (c) selected patches overlaid on the vessel segmentation (I_{seg}), (d) the artery/vein ground truth (G), the color-coded (e) orientation, (f) confidence and (g) and absolute curvature maps of this SLO image.

By obtaining the $I'_{seg,i}$ it is possible to perform the connectivity analysis (Algorithm 2) only for pixels labeled as vessels. This helps in reducing the size of the affinity matrix and correspondingly the computational complexity of the spectral clustering step. The vessel locations are found as:

$$v_i = \{(x,y) | I'_{seg,i}(x,y) = 1\}, \ i = 1, \ldots, M. \tag{7.3.4}$$

Therefore, for each image patch (I'_i) only at vessel pixel locations (v_i) the groups are obtained and the resulting clusters are compared with the vessel labels in G'_i. If the detected units in the image patch match the individual vessel labels in the ground truth, then it is perceived as a correct result. It is worth mentioning, in case two vessels have the same label (artery or vein) but they do not belong to the same vessel tree, they are considered as separate units during the comparison.

After this preprocessing, the algorithm presented in the previous section can be applied.

7.3.3 Validation

7.3.3.1 Phantom images

To validate the method using the phantom images, the method presented in Algorithm 2 is directly applied on the lifted images, starting from Step 4. The orientation and curvature values of these images are available from the beginning. The last two columns of Figure 7.21 represent two different clustering results per image (two images per category). For each image, two kernels have been used for obtaining the affinity matrix: the new kernel (adaptive, based on the curvature at each point); and the kernel used in Favali et al. [2016a]. This helps in highlighting the importance of including additional contextual information for connectivity analysis. In these phantom images the intensity was constant, so that the contribution of this feature is always 1. Each color in the final results represents one detected unit. The parameter σ_κ used in these simulations is 0.01 for the 4D kernel [Favali et al., 2016a] and 0.001 for the 5D one. As seen in this figure, the new method is capable of grouping the elongated, rotated and curved structures despite disconnections, high curvature points or small crossing angles. It not only differentiates well between curved crossing structures, but also groups the bifurcations within the main parent structure so that they construct one unique unit.

7.3.3.2 Retinal patches

To validate the method on retinal image patches, 272 image patches with a fixed size of 51×51 pixels have been examined semi-automatically. The number of patches processed per group is presented in the second column of Table 7.3.

The group labels are automatically compared with the clustering results. In case two vessels with similar labels in the ground truth image (artery or vein) belong to separate vessel trees (parents), they get a different label. In addition, we perform a final check to control the final results. Two criteria are used to evaluate the performance. One is the Correct Detection Rate (CDR%) as defined in Section 7.2.4 and in Favali et al. [2016a]. This criterion represents the percentages of correctly grouped patches among all examined cases. The second criterion is Q_{clust} defined in Step 6, which measures the alignment quality of Algorithm 3. As mentioned in Algorithm 3, the best number of clusters is the one which minimizes the defined cost function or maximizes the quality ($0 \leq Q_{clust} \leq 1$). This criterion represents how well aligned the elements of each group are. These two performance values have been measured for all the patches and presented in the last two columns of Table 7.3 for each category separately.

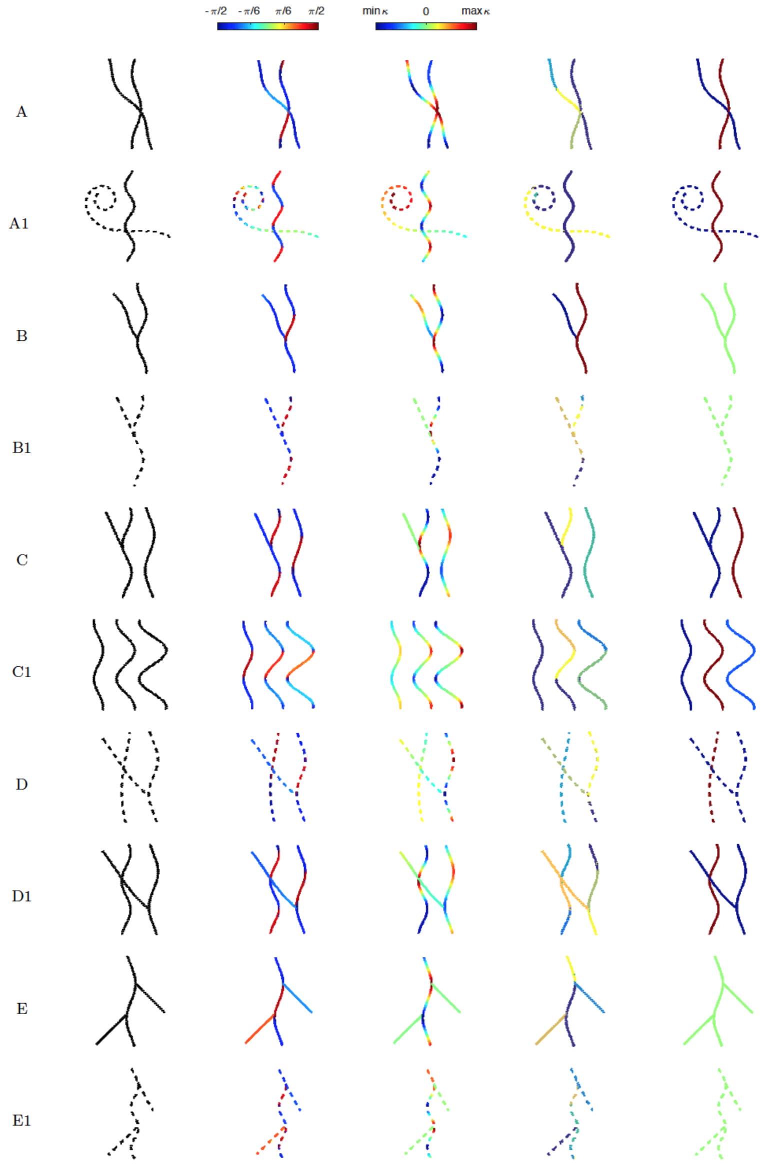


Figure 7.21: Samples of phantom images in different categories. From left to right, the images in each category represent: stimulus, the orientation map, the curvature map and the clustering result with the previous [Favali et al., 2016a] and the new kernel. The color of the curvature maps are scaled between the maximum and minimum values of the curvature in each image.

During the experiments the parameters used in numerical simulations and the calculation time of different parts of the experiments (excluding the patch preparation steps) have been recorded. Several parameters are involved in creating the kernel. Some of them are determined automatically based on the available information in the data and others need to be set manually. The first set includes the size of the kernel in x and y dimension (n_x and n_y respectively). The other one is the number of discrete curvature values (n_κ), which for each the 3D kernel is created. Considering the step size as 1 pixel, n_x and n_y are determined by the difference between maximum and minimum coordinates of the vessel locations in x and y directions. Similarly, considering the step size of 0.05 for discrete curvature values, n_κ is obtained by division of the difference between maximum and minimum of the available curvature values in the patch over the step size. Moreover, the number of steps H used in generating the random paths is set automatically as one third of the patch size. The second set of parameters is set manually. The number of discrete orientations (n_o), number of iterations in the Monte-Carlo simulation (n) and σ_κ (used in equation 6.2.15) were set to 18, 100000 and 1 respectively and kept constant for all the cases. The other parameters are presented in Table 7.3 for each category separately and for all the cases (in *mean* ± *standard deviation* format).

2*Group	2*Size	Parameters		Performance	
		σ_κ	σ_{int}	$CDR\%$	Q_{clust}
A	16	0.0079 ± 0.0247	0.2532 ± 0.0013	0.8125	0.9866
B	42	0.0011 ± 0.0027	0.2996 ± 0.0026	0.7143	0.9979
C	48	0.0031 ± 0.0044	0.2512 ± 0.0093	0.8542	0.9978
D	31	0.0038 ± 0.0047	0.2578 ± 0.0016	0.7742	0.9974
E	19	0.0025 ± 0.0041	0.2699 ± 0.0046	0.7368	0.9967
A1	9	0.0018 ± 0.0024	0.2574 ± 0.0037	0.6667	0.9756
B1	23	0.0004 ± 0.0010	0.2951 ± 0.0054	0.6522	0.9987
C1	41	0.0030 ± 0.0044	0.2993 ± 0.0078	0.8537	0.9975
D1	27	0.0040 ± 0.0048	0.2552 ± 0.0018	0.8148	0.9974
E1	18	0.0021 ± 0.0038	0.2926 ± 0.0018	0.7222	0.9967
All	274	0.0031 ± 0.0087	0.2750 ± 0.0072	0.7602	0.9942

Table 7.3: The number of analyzed patches, the parameters used during numerical simulation and the measured performance values per category and in total.

For recording the calculation times the whole process has been divided into four steps: the discretization step before creating the kernel; creating the kernel for several curvature values; creating the affinity matrix and the spectral clustering step. The times are called t_{disc}, t_{kernel}, $t_{affinity}$ and t_{clust}, respectively, and they are affected by several parameters including the number of vessel pixels in each patch ($|v_i|\ , i = 1, \ldots, M$), the number of discrete orientations (n_o), curvatures (n_κ) and the dimension of the kernel in x and y dimensions (n_x and n_y). To consider these effects, the weighted average of calculation times for each image patch is obtained, so that the weight for each timing is defined as the product of the affecting parameters. It worths mentioning that since the final number of clusters (C_{best}) is determined by comparison among the clustering costs of several cluster sizes (C), t_{clust} is additionally affected by the number of examined cluster sizes ($n_c = 20$ for all the cases), so the final clustering time is divided by n_c. Thus weighted timings are calculated as:

$$
\begin{aligned}
\bar{t}_{disc} &= \sum_{i=1}^{N} t_{disc,i} n_{x,i} n_{y,i} n_{o,i} \Big/ \sum_{i=1}^{N} n_{x,i} n_{y,i} n_{o,i} \\
\bar{t}_{kernel} &= \sum_{i=1}^{N} t_{kernel,i} n_{x,i} n_{y,i} n_{o,i} n_{\kappa,i} \Big/ \sum_{i=1}^{N} n_{x,i} n_{y,i} n_{o,i} n_{\kappa,i} \\
\bar{t}_{affinity} &= \sum_{i=1}^{N} t_{affinity,i} |v_i|^2 \Big/ \sum_{i=1}^{N} |v_i|^2 \\
\bar{t}_{clust} &= 1/n_c \sum_{i=1}^{N} (t_{clust,i}) |v_i|^2 \Big/ \sum_{i=1}^{N} |v_i|^2
\end{aligned}
\tag{7.3.5}
$$

where i and N indicate the patch number and total number of patches respectively. It is worth mentioning that the size of the affinity matrix for patch i is $|v_i^2|$. These weighted times and their normal average over all patches per step are presented in Table 7.4.

	$\bar{t}_{disc}$	$\bar{t}_{kernel}$	$\bar{t}_{affinity}$	$\bar{t}_{clust}$
mean(s)	0.06	43.26	17.86	0.86
weighted mean (s)	0.06	60.64	17.73	1.06

Table 7.4: The weighted and normal mean of the processing time of each step in analyzing retinal patches.

A set of sample results for various kinds of patches is depicted in Figure 7.22. In this figure, the first column shows the cropped patch from the artery/vein ground truth image. The second column depicts the color-coded normalized intensity values taken from the pre-processed image. As seen in these figures, the variation of intensity is too much even for small children vessels belonging to one parent vessel. The third and fourth columns represent the color-coded orientations and curvature values. Finally, the last column represents the clusters found in each patch, each shown in an individual color.

The presented results, parameters and timings are discussed in the next section.

7.3.4 Discussion

Based on the results presented, the main advantage of the new method is that by including the curvature information as an additional contextual information, the kernel adapts itself naturally according to the available data. If the curvature is high, the kernel rotates as well, otherwise it finds a closer path to the points which are collinear with respect to the reference point. In both datasets, the bifurcations are grouped with the parent vessel, but at crossovers with small crossing angles, despite

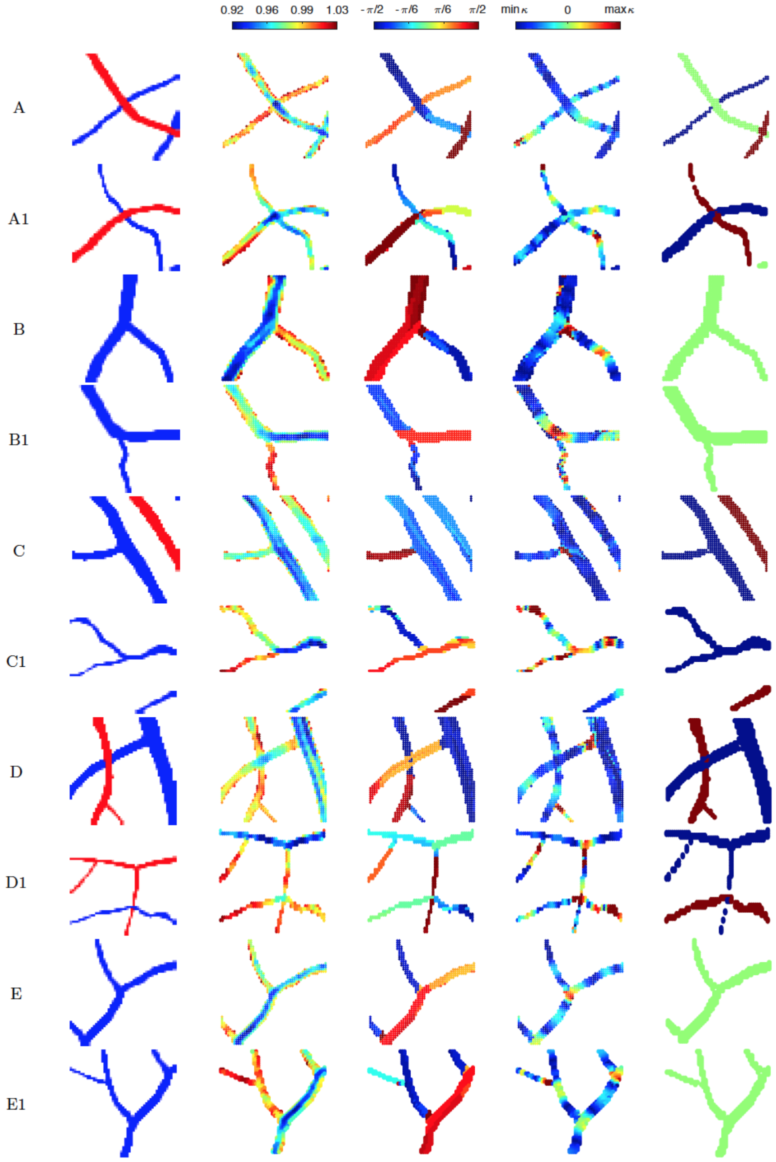


Figure 7.22: Samples of retinal patches in different categories. From left to right, the images in each category represent: artery/vein vessel ground truth, intensity, orientation, curvature and clustering results. The color of the curvature maps are scaled between the maximum and minimum values of the curvature in each image patch.

their similar appearance to junctions, the vessels are totally separated. The main reason is that the curvature at junction points is high (because of sudden change of orientation), while for crossings the orientation for individual vessels changes only slightly (in most of the cases). This is advantageous not only in differentiating between junctions and crossings, but also in separation of arteries from veins or crossing tree structures from each other. Figure 7.23 shows the clustering results for two retinal patches obtained using the new kernel and the previously introduced kernel by Favali et al. [2016a]. This helps in depicting the differences between the two methods visually.

As presented in Figure 7.22, the intensity is a less informative feature compared to the geometrical features because of its

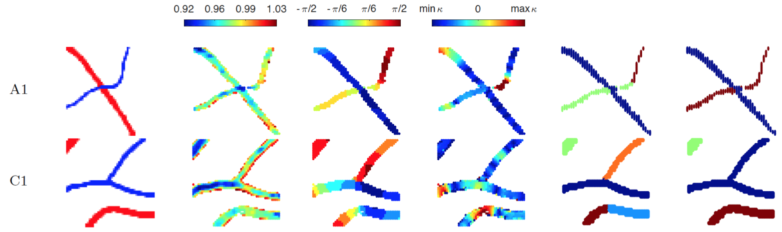
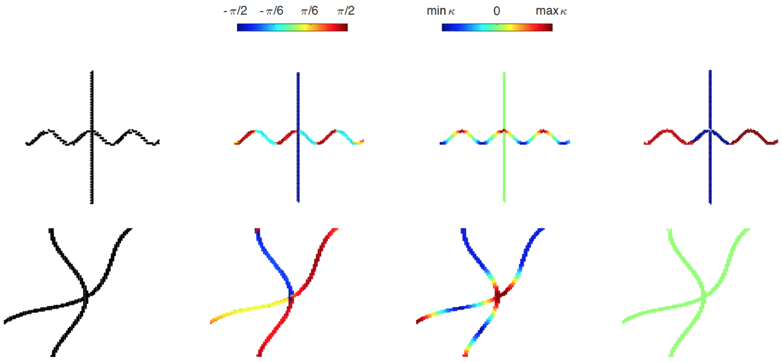


Figure 7.23: Clustering results obtained using the connectivity kernel introduced here and the one proposed in Section 6.1 [Favali et al., 2016a]. From left to right: artery/vein vessel ground truth, intensity, orientation, curvature and clustering results with the previous and the new kernels.

large variation within a small neighborhood; however, in some cases, it is a good local criterion differentiating arteries from veins. Thus it is also included in the final affinity matrix with a smaller effect (using a relatively large σ_{int}).

Some examples of the limitations of the method in clustering the phantom and retinal patches are represented in Figure 7.24 and 7.25 respectively. For phantom cases, the presence of very high curvature combined with the co-circularity and co-linearity of vessels does not allow to obtain a good clustering result. Considering other features, as the intensity, could be helpful in solving this problem. However, as shown in the top row of Figure 7.25, the feature of intensity is not useful for correct clustering. In this image, one of the bifurcations has been assigned as a vessel crossing the other one because it is almost orthogonal to its parent vessel; while the other crossing vessel has been wrongly clustered as a bifurcation. In the bottom row, one of the small bifurcations is totally missing in the segmentation and the other small one is not clustered with its parent vessel because of lack of information.



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

Figure 7.24: From left to right: the stimuli (a), the orientation maps (b), the curvature maps (c) and the clustering results with the new kernel (d).

The statistical analysis on the parameters used during numerical simulations is presented in Table 7.3. Based on these results the curvature diffusion constant parameter (σ_κ) changes in a small range for simple and challenging cases per group, ex-

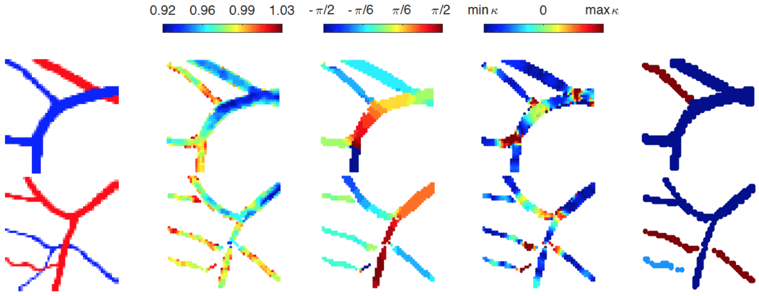


Figure 7.25: Wrong clustering results on two retinal patches. From left to right: the artery/vein ground truth, vessel intensity, orientation, curvature and clustering results.

cept for category A and A1, considering the fact that the number of available challenging patches in category A is small compared to the others. The σ_{int} parameter has a small variation as well. Therefore, it is reasonable to use the mean value in general for examining new patches in each group. It is worth mentioning that all the patches examined in this work have been selected from retinal images with the same resolution and respective pixel size. If the pixel size increases or decreases, the σ_κ controlling the size of the 3D kernel needs to be increased or decreased accordingly. It is also important to mention that compared to previous work, the eigensystem analysis is fully automatic in this work, due to the use of self-tuning spectral clustering. Therefore, no additional parameters need to be tuned for this step.

Qualitative and quantitative results indicate the better performance of the method on all kinds of retinal patches and challenging structures. Compared to Favali et al. [2016a], the *CDR%* performance values have changed for some groups. There are two main reasons. On one hand, in the previous work, the *CDR%* was calculated for 20 patches per group; while in this work each group is categorized into two groups depending on the available structure and also the number of patches per group is different. On the other hand, in this work we evaluated the performance with the assumption that the bifurcations need to be grouped with the main parent vessel, while in previous work, due to use of one elongated kernel, the assumption was to have at least two separate units depending on the bifurcation angle. Another minor difference is that the dataset has changed and the patches have been selected from a different set of retinal images. Therefore, it is not fair to make a one-by-one comparison to the previous results.

Last but not least point is about the computation times. The codes are implemented in Matlab and the times are measured on an Apple Macbook Air, Intel Core i7, 1.7 GHz processor and 8GB of memory. The most time consuming step as presented in Table 7.4 is the calculation of several 3D kernels. The number of kernels depends on the number of discrete curvature values (n_κ) existing in the data, and the computation time as mentioned before depend on its size. The next most time consuming part is the calculation of the affinity matrix which is performed per pair of points. The weighted average times are good indicators of the complexity of each step. Although they are still relatively small, they can be improved both from hardware and implementation points of view.

Concluding, this method allows to analyze the connectivities in images containing elongated, rotated and curved structures. One limitation of the method arises when wrong information is provided as the input of the method, then the failure is natural. If the measured curvature or orientations are not accurate enough, then the method fails. Therefore, it is essential to validate the curvature and orientation measurement methods in advance.

The proposed method has a great potential in discrimination and separation of arteries from veins in retinal images and, in a general view, separation of all the tree structures crossing each other in the vasculature network. Most of the segmentation or artery/vein separation methods are local, pixel based techniques which do not take into account the global connectivity of the blood vessels in the network. By including this global connectivity criterion, most of the errors and wrong detections will be eliminated and the problem of missing information will be handled appropriately.

8 Theory of the cortical model in encoding-decoding techniques

In this section we apply the geometrical models of the cortex to identification and reconstruction of images from human brain activity [Kay et al., 2008].

Following Kay et al. [2008], the first problem can be stated as follows:

"Let us considered a large, arbitrary set of images. The observer picks an image from the set and views it while brain activity is measured. Is it possible to use the measured brain activity to identify which specific image was seen?"

Even more challenging is the second problem:

"Is it possible to reconstruct the image seen from the measured brain activity?"

Every image can be represented in term of receptive profiles and we develop a decoding method based on quantitative receptive fields models that characterize the relationship between visual stimuli and fMRI activity in early visual area. We will obtain optimal identification performances, while low spatial resolution of available fMRI data is a serious obstacle to reconstruction.

8.1 Introduction to Functional MRI

The problem of linking a cognitive function to different anatomical structures of the brain represents one of the oldest debates in neuroscience. Several brain imaging techniques are available today, in this section we introduce one of the principal, the functional MRI (fMRI), that represents the most powerful tool available for measuring human brain activity.

The functional MRI provides an indirect and nonlinear measure of neuronal activity, because it does not measure neuronal activity directly, but rather measures changes in blood oxygenation caused by metabolic processes in neurons.

Functional imaging techniques detects physiological activities such as blood flow. Due to its good spatial resolution ($1mm^3$), the fMRI is used to localize the brain activity in response to a given task or experimental condition. Moreover, it allows to study the whole brain coverage and not only predefined regions or layers. The spatial resolution is determined by the physical constraints of the fMRI scanner (as the limits on the strength of the magnetic fields that can be produced and limits on the power of the radio frequency energy that can be deposited safely in the tissue).

The first fMRI method developed, which is also the one used to record the data that we will consider, measures the oxygen change in blood flow and is known as BOLD (*Blood-oxygen-level dependent*) contrast. It is well known that blood oxygenation in the brain is closely linked to neural activity. The BOLD contrast can be explained considering the function of an oxygen carrier in the blood cells, the hemoglobin. The principle behind its functioning is that oxygenated and deoxygenated hemoglobin yield different reactions to an externally applied magnetic field, giving distinct magnetic resonance responses [Ogawa et al., 1990, Thulborn et al., 1982]. This allows to trace the presence or the absence of neural activity, as shown in Figure 8.1.

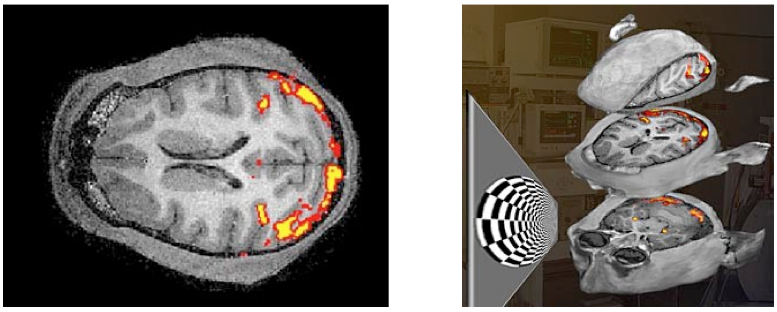


Figure 8.1: Functional magnetic resonance of the monkey brain under visual stimulation. Blood Oxygen Level Dependence (BOLD) technique, field strength 4.7 Tesla. Left: Clearly a marked activity is measured in V1. Right: different cut-away views from the brain of the anesthetized monkey. Source: [Romeny, 2008].

The spatial resolution of fMRI is given by the size of a *voxel*, a 3-D cuboid given by a measure of the scanner. Its size is between 4mm to 1mm; smaller voxels have fewer neurons on average, include less blood flow and have less signal to noise ratio than the larger ones. fMRI data are collected in the form of slices, which are then projected onto a standardized anatomical volume. Datasets are commonly measured using repeated 2D imaging methods, resulting in a temporal offset between slices and the time resolution is given by the repetition time of successive image acquisitions [Pedregosa-Izquierdo, 2015].

In our example we will consider data from Kay et al. [2008], where functional BOLD data were recorded from occipital cortex at a spatial resolution of 2mm x 2mm x 2.5mm and a temporal resolution of 1 Hz.

8.2 Encoding and decoding model in the problem of identification of images from fMRI activity

8.2.1 Overview of encoding and decoding results in fMRI analysis

Several scientific achievements have been made in the last few decades in brain encoding and decoding using fMRI. The goal of many fMRI studies is to understand what sensory, cognitive or motor information is represented in some specific region of the brain.

Functional magnetic resonance imaging studies have decoded orientation [Haynes and Rees, 2005, Kamitani and Tong, 2005], position [Thirion et al., 2006] and object category [Cox and Savoy, 2003, Haxby et al., 2001] from activity in visual cortex.

While encoding uses external stimuli to predict brain activity, decoding uses brain activity to predict information about external stimuli. In particular, encoding models goal is to understand how well the brain activity can be predicted from the modeled external stimuli. Decoding models aim at studying how much of the external stimuli can be learned considering the brain activity [Haxby et al., 2001, Kay et al., 2008, Miyawaki et al., 2008, Naselaris et al., 2011]. In this sense, encoding and decoding can be seen as complementary methods to understand the fundamental mechanisms of brain functions via neural codes ([Dayan et al., 2001, Gerstner et al., 1997, Haynes and Rees, 2006, Trappenberg, 2009].

In this field, Haynes and Rees [2006] discussed the general research problem of *brain reading*, focusing on visual perception and other types of mental state. Kay and Gallant [2009] summarized several advancements of brain decoders of visual stimuli via fMRI including the ones in Kay et al. [2008], Miyawaki et al. [2008], Thirion et al. [2007]. They provided analysis on the future research direction and potential application of brain decoding. Existing studies have been reviewed by Hasson et al. [2010] that examined the reliability of cortical activity within or between human subjects in response to natural visual stimulation. An extensive overview of recent experimental methodology advancements in voxel-based decoding models of visual stimuli has been provided by Naselaris et al. [2011], where it is described an estimate of encoding model for every voxel in an fMRI scan. The estimated encoding model is used to perform decoding.

This voxel-based methods for brain decoding have been widely used in the literature due to its simplicity and effectiveness. In several voxel-based encoding models [Naselaris et al., 2011, Mitchell et al., 2008, Thirion et al., 2007], authors tried to predict the functional activity in single voxels, evoked by different stimuli. Those models contain a quantitative description of how stimulus information is represented in the functional activity of individual voxels. In Kay et al. [2008] several thousands of voxels located in the V1, V2 and V3 areas of the visual cortex were used for learning the predictive receptive-field models. We will be inspired by these results in the following Sections.

In order to study the fMRI signals and informations, researchers have also tried to extract fMRI BOLD signals from the region of interest (ROI) [Cox and Savoy, 2003, Walther et al., 2009, Hu et al., 2012].

Because of the remarkable structural and functional variation across individual brains, neuroimage registration algorithms are still insufficient to accurately establish correspondences in different brains [Liu, 2011]. For these reasons, both these methods have limitations [Liu, 2011]. To overcome the limitations, solution have been proposed by Zhu et al. [2012b,a], to represent structural and functional brain architectures by a set of reproducible and consistent brain landmarks that can be accurately and reliably localized in each individual brain. Moreover, in recent years, works that employed connectivity-based measurements for quantification of the brain's responses have been presented [Hu et al., 2010, Ji et al., 2011, Richiardi et al., 2011, Pantazatos et al., 2012]. In this way, functional connectivity can be used for measurements of functional brain responses, potentially used for brain decoders. This method represents a promising opportunities for the advancements of brain encoding and decoding applications in the near future.

A summarization of these brain encoding and decoding studies is presented more in details in Chen et al. [2014].

8.2.2 Identifying natural images from human brain activity

We will be particularly inspired by the results in Kay et al. [2008]. In his paper the authors develop a decoding method based on receptive-field models that characterize the relation between visual stimuli and fMRI activity in early visual areas.

The original analysis of Kay et al. [2008] is composed by two stages (see Figure 7.2). The first one is model estimation, where fMRI data are recorded while subjects viewed 1750 training images. These data are used to estimate a receptive-field model for each voxel. The second stage is image identification and fMRI data are recorded while subjects viewed a collection of 120 novel images, that form the validation set. By using the encoding model evaluated at the first stage, the goal was to identify which image had been seen by the subjects, comparing the measured fMRI activity with the predicted signal in response to the presentation of these images.

The model estimation scheme is represented in Figure 8.2.

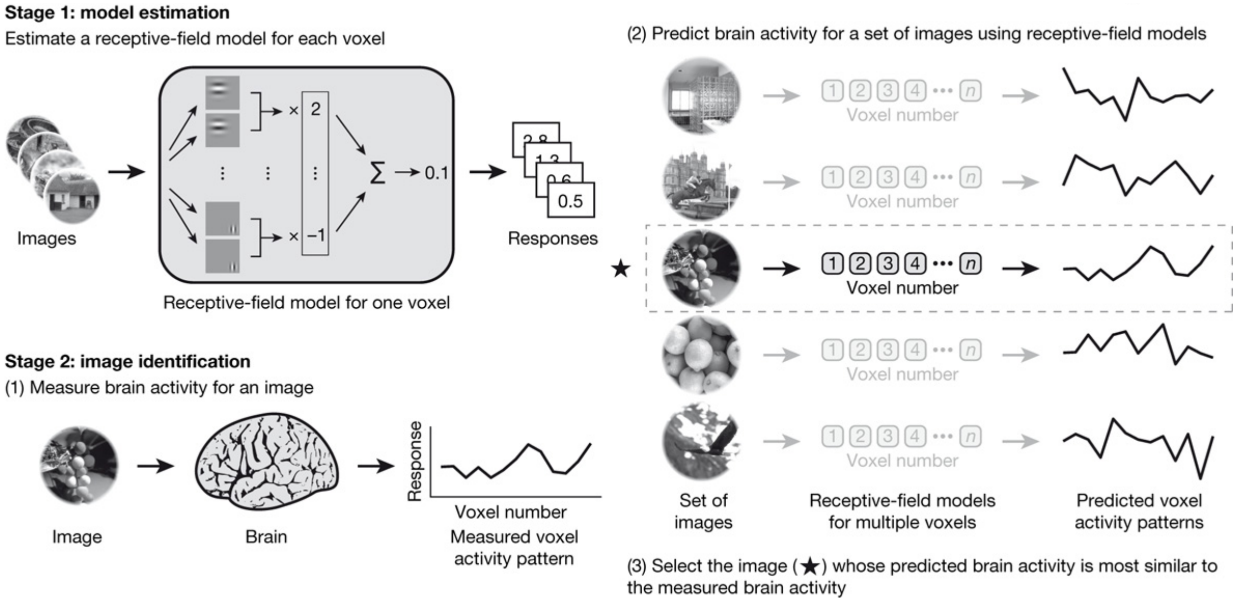


Figure 8.2: The two stage of the experiment described in Kay et al. [2008]: model estimation and image identification. Source: [Kay et al., 2008].

8.2.2.1 Visual stimulus description

Visual stimulus consist of grayscale images I of size 128×128 pixels. The model in Kay et al. [2008] is based on a representation of the images in terms of a Gabor wavelet pyramid (GWP) [Jones and Palmer, 1987, Daugman, 1985, Lee, 1996] and describe tuning along the dimensions of space, orientation and spatial frequency (Figure 8.3).

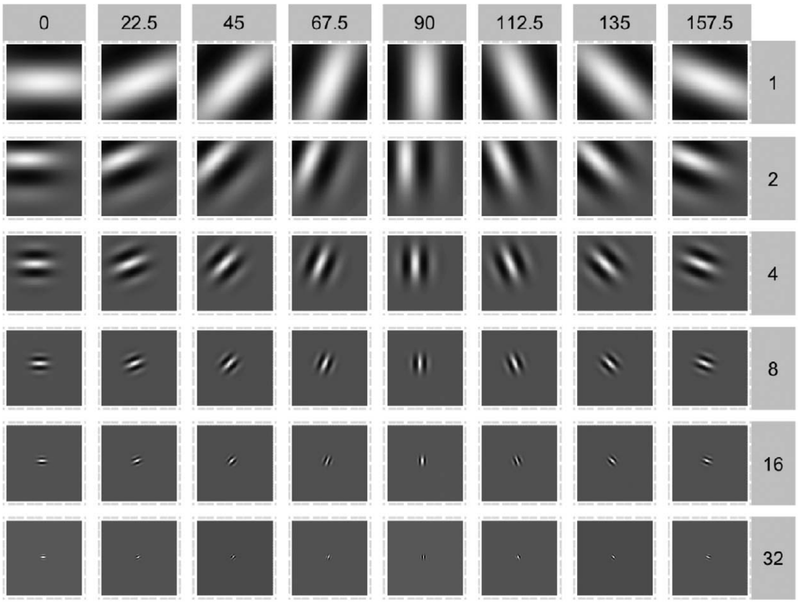


Figure 8.3: Simple cell receptive fields of the GWP model. Each square is of size 128x128 pixels and shows an even-symmetric Gabor wavelet. The receptive fields spanned eight orientations and six spatial frequencies. Source: [Vu et al., 2011]

Concretely, the GWP model is a hand-designed population of quadrature-phase Gabor wavelets that spanned a range of locations, orientations and spatial frequencies as shown in Figure 8.3. Each wavelet is fully connected to the input stimuli. A receptive-field model is estimated for each voxel based on its responses to the training images; fMRI encoding model predicts voxel responses.

Let's consider the 2D Gabor filters:

$$\psi_{g_k}(x,y) = e^{-\frac{(\tilde{x}^2+\tilde{y}^2)}{2\sigma_k^2}} e^{2\pi i\omega\tilde{y}} \tag{8.2.1}$$

where g_k denotes, as in Section 3.2.1, the discrete choice of parameters defining the Gabor filter $g_k = (x_k, y_k, \theta_k, \sigma_k)$:

$$\begin{aligned} \tilde{x} &= (x - x_k)\cos(\theta_k) + (y - y_k)\sin(\theta_k) \\ \tilde{y} &= -(x - x_k)\sin(\theta_k) + (y - y_k)\cos(\theta_k). \end{aligned} \tag{8.2.2}$$

For every image I_i the Gabor energy associated to a Gabor pair ψ_{g_k} is computed as :

$$\tilde{E}_{ki} = |\int I_i(x,y)\psi_{g_k}(x,y)dxdy| \tag{8.2.3}$$

and is called *contrast energy*. A total of n Gabor filters have been considered. Hence the contrast energy is represented as a matrix $\tilde{E}_{ki}$, where the index i identifies the image $i = 1 \cdots N$, with $N = 1750$, while the index k, with $k = 1, \cdots, n$ identifies the coefficient of the k Gabor filter in the representation.

8.2.2.2 The fMRI measurement

We use the estimated responses evaluated in Kay et al. [2008] where peak BOLD responses to each of the training and validation images were estimated from the preprocessed data.

For every image I_i the corresponding fMRI signal is measured. This is represented as 3-D cuboid given by a measure of the scanner. In the present case it is a $64 \times 64 \times 18$ volume. Also, of the 73,728 ($64 \times 64 \times 18$) voxels recorded for each scan, only $\sim 25,000$ voxels in or near the cortex were selected for each subject. For simplicity we will identify each voxel in the volume with an index v, varying from 1 to $M \sim 25,000$. Hence the full set of measured fMRI signal will be a matrix $y_{i,v}$, $i = 1, \cdots, N$, $v = 1, \cdots, M$, where the index i identifies the stimulus image i and the index v identifies the voxel.

8.2.2.3 Encoding procedure

In Kay et al. [2008] the authors assume this relation between energies $\tilde{E}_{ki}$ and measured activities y_{iv}:

$$y_{iv} = \sum_{k=1}^{n} \tilde{E}_{ki}\beta_{kv} + \beta_{0_v} \tag{8.2.4}$$

where the matrix β and the vector β_0 are unknown. In the matrix β_{gv} the index g varies from 1 to n and identifies the Gabor filter, the index v varies from 1 to M and identifies the voxel. The vector β_0 only depends on the voxel v.

This model assumes that the measured activities are a weighted sum of a fixed transformation of the local contrast energy features. They are interested in finding β such that the linear combination with coefficients of the contrast energy of a set of images should approximate in the closest way the fMRI measured responses to those images. For every voxel v, this is formalized through the minimization problem:

$$\min_{\beta} F_{1,v}(\beta) \tag{8.2.5}$$

where

$$F_{1,v}(\beta) = \sum_{i=1}^{N} (y_{iv} - \sum_{k=1}^{n} \tilde{E}_{ki}\beta_{kv} - \beta_{0v})^2 \tag{8.2.6}$$

is the mean square error between the measured and the predicted response. Gradient descent with early stopping is used in Kay et al. [2008] to learn a predictor of voxel activity on the training set. Early stopping is a form of regularization used to avoid overfitting, which occurs when a model describes the noise rather than the underlying relationships that one aims to characterize.

For every voxel the corresponding receptive-field model can be evaluated as:

$$RF_v = \sum_{k=1}^{n} \beta_{kv} |\psi_{g_k}| \tag{8.2.7}$$

where ψ_{g_k} is the Gabor filter defined in (8.2.1). An example is shown in Figure 8.4.

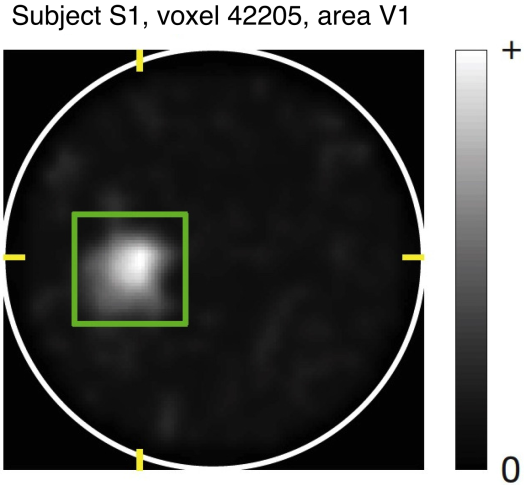


Figure 8.4: Receptive-field model for a representative voxel from Kay et al. [2008]. The intensity of each pixel indicates the sensitivity of the receptive field to that location. The white circle delineates the bounds of the stimulus.

8.2.2.4 Decoding and image identification

In the second stage, the image identification, they use the encoding model estimated in the first stage to predict the voxel activity pattern evoked by each of the images (see Figure 8.2). From the stimuli in the validation set they predict the activation coefficients that they then use to identify the correct image. The image whose predicted voxel activity pattern is most correlated (Pearson's correlation) with the measured one is selected. The Pearson correlation coefficient is a measure of the linear dependence (correlation) between two variables and it is evaluated as their covariance divided by the product of the standard deviation of the two responses. It has a value between +1 and −1 inclusive, where +1 is total positive linear correlation, 0 is no linear correlation, and−1 is total negative linear correlation.

To optimize the performance of the identification algorithm, they first select voxels whose receptive field models have the highest predictive power. This predictive power is evaluated as the Pearson's correlation between measured y_m and predicted y_p responses for the images used in the image identification runs:

$$\rho_{y_p,y_m} = \frac{\sigma_{y_p y_m}}{\sigma_{y_p}\sigma_{y_m}}. \tag{8.2.8}$$

In their experiment approximately 5000 voxels are located in the stimulated portions of visual areas V1, V2 and V3; the predictive power of the receptive-field models is significantly variant for different voxels. Optimal performance is achieved using the first 500 voxels whose receptive-field models had highest predictive power; most of these voxels are located in area V1, where predictive power is higher.

Identification performances for both subjects are illustrated in Figure 8.5. For subject S1 they obtain 92% (110/120) of the images correctly identified, whereas chance performance is just 0.8% (1/120). For subject S2, 72% (86/120) of the images are identified correctly.

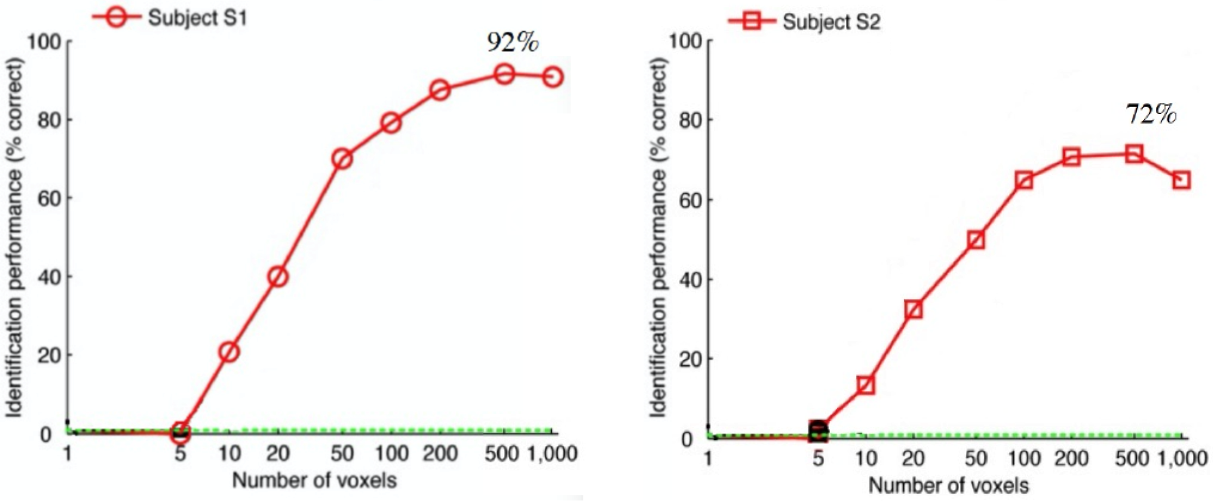


Figure 8.5: Identification performances for subjects S1 and S2. Source: [Kay et al., 2008]

These high performance levels demonstrate the validity of their decoding approach and indicate that their receptive-field models accurately characterize the selectivity of individual voxels to natural images.

8.2.3 A modified version of the encoding model

The aim of this section is to provide a modified and more efficient model of encoding. We start from the Gabor Wavelet Pyramid model of Figure 8.3 composed by 6 spatial frequencies and 8 possible orientations θ, for a total of $(1^2 + 2^2 + 4^2 + 8^2 +$

$16^2 + 32^2) \times 8 = 10,920$ phase-invariant complex cells [Kay et al., 2008]. Variants of this model were used in a series of seminal encoding and decoding studies [Nishimoto et al., 2011, Naselaris et al., 2009, Kay et al., 2013]. We will provide a new model, with the following main features:

- following the idea of Vu et al. [2011], we will add a L^1 penalization, called Lasso model, in order to ensure sparsity of the representation and we will add a nonlinearity of logaritmic type in order to better model the structure of the cortex;
- we will add a regularization in the subriemannian cortical structure.

We will see that our new encoding model has better identification performances compared to the model described in Kay et al. [2008].

8.2.3.1 The Lasso method

The main property of each voxel is to have receptive fields localized around a very precise location in the visual field. The model in Kay et al. [2008] has already this property, however, following Vu et al. [2009, 2011] we propose here a modified version of the model based on a different method of evaluation of coefficients. It is based on the so called Lasso method [Tibshirani, 1996]. This is a popular method that uses an L^1 penalty to achieve a sparse solution. That is, the minimizer of the Lasso problem we will have has many vanishing components $\beta_j = 0$ as it is possible and the solutions will be more localized. This would not be true with ridge regression, that penalizes with the L^2 norm instead of the L^1 norm, due to the strong regularizing effect of the L^2 norm. Moreover, while ridge regression improves prediction error by shrinking large regression coefficients to reduce overfitting but it does not perform covariate selection, Lasso is able to achieve both of these goals by forcing the sum of the absolute value of the regression coefficients to be less than a fixed value.

Let us consider to have predictors $\tilde{E}_{ki}$, $i = 1, ..., N$ $k = 1, \cdots, n$ and outcome values y_{iv} for the ith observation, for $i = 1, 2, .., N$. We fix the value v of the voxel, and use the same notation of the previous section. The Lasso solves the problem:

$$\min_{\beta} F_{1,v} \text{ subject to } \sum_{k=1}^{n} |\beta_{kv}| \leq s. \tag{8.2.9}$$

Proposition 4. *By convexity, for each s, there is always at least one solution in equation (8.2.9).*

Note that the solution is not necessarly unique if, for example, $\tilde{E}$ has not maximum rank. The bound s is usually chosen by a model selection procedure as cross validation. By Lagrange multiplier Theorem, the solution of equation (8.2.9) also minimizes the following functional, generally called Lagrangian of the system:

$$F_{2,v}(\beta, \lambda) = \frac{1}{2}\sum_{i=1}^{N}(y_i - \sum_{k=1}^{n}\tilde{E}_{ik}\beta_{kv} - \beta_{0,v})^2 + \lambda\sum_{k=1}^{n}|\beta_{kv}| \quad (8.2.10)$$

where $\lambda \geq 0$.

Precisely the following proposition holds:

Proposition 5. *If $\hat{\beta}$ minimizes equation* (8.2.9) *there is a value λ such that $\hat{\beta}$ minimizes equation* (8.2.10). *If $\hat{\beta}(\lambda)$ minimizes equation* (8.2.10) *then it also solves equation (8.2.9) with $s = \sum_{k=1}^{n} |\hat{\beta}_k(\lambda)|$.*

In other words, studying equation (8.2.10) we introduce a new variable λ but we study an unconstraint optimization problem, instead of a constraint one.

Different algorithms have been proposed for solving this problem [Efron et al., 2004, Friedman et al., 2007]. Considering uncorrelated multiple predictors, the Lasso solutions are soft-thresholded versions of the individual least squares estimates [Friedman et al., 2007]. For λ fixed a coordinate-wise optimization methods is applied and the equation is iteratively solved for every fixed component β_j.

At each iteration step we denote $\hat{\beta}_k(\lambda)$ the values of the parameters found at the previous iteration step. Then we fix each component $\hat{\beta}_k$ with $k \neq j$ and consider F as function of β_j alone. Then equation (8.2.10) becomes:

$$f_{2,v}(\beta_j, \lambda) = \frac{1}{2}\sum_{i=1}^{N}(y_i - \sum_{k\neq j}\tilde{E}_{ki}\hat{\beta}_k - \tilde{E}_{ji}\beta_j)^2 + \lambda\sum_{k\neq j}|\hat{\beta}_k| + \lambda|\beta_j|. \quad (8.2.11)$$

For every component j the minimum is computed and at the end of the iteration step the values $\hat{\beta}_k(\lambda)$ of minimizers are updated. For the fixed value of λ the convergence of the method to the minimum $\hat{\beta}(\lambda)$ of $F_{2,v}(\cdot, \lambda)$ is guaranteed.

The next step is to minimize with respect to λ. To do so, we apply a cross-validation technique, which is a technique that uses different subsets of the data to fit the model in order to limit problems like overfitting [Hastie et al., 2011]. A good estimator estimated from a training set should perform well on a new, independent set of data. Hence the available data are organized into both training and test sets. Here use $K-$fold cross-validation, i.e. we partition the training data into $K = 5$ separate sets S_h of equal size N/K, $h = 1, \cdots 5$. For each of these subset S_h of data we minimize $F_{2,v}$ on the training set excluding the elements of S_h:

$$F_{2,v}(\beta, \lambda) = \frac{1}{2} \sum_{i \notin S_h} (y_i - \sum_{k=1}^{n} \tilde{E}_{ik} \beta_{kv} - \beta_{0v})^2 + \lambda \sum_{k=1}^{n} |\beta_{kv}|. \quad (8.2.12)$$

If $\hat{\beta}^h(\lambda)$ is the minimizer, we evaluate the prediction error of the model on the elements of S_h:

$$CV(h, \lambda) = \frac{K}{N} (\sum_{i \in S_h} (y_i - \sum_{k=1}^{n} \tilde{E}_{ik} \hat{\beta}^h_{kv} - \hat{\beta}^h_{0_v})^2 + \lambda \sum_{k=1}^{n} |\hat{\beta}^h_{kv}|). \quad (8.2.13)$$

The overall cross-validation error of the model is then:

$$CV(\lambda) = \frac{1}{K} \sum_{h=1}^{K} CV(h, \lambda). \quad (8.2.14)$$

The optimal paramenter λ is the one which minimizes the overall CV error.

8.2.3.2 A nonlinear model

In Vu et al. [2011] it is underlined how the model described in Kay et al. [2008] didn't take into account systematic nonlinearity across voxels. Following their idea, who analysed encoding and decoding V1 fMRI responses to natural images using sparse nonparametric models, and the work of Güçlü and van Gerven [2014] we consider a variation of the model described in Kay et al. [2008] adding a non linearity in the computation of the energy term.

In particular, considering the contrast energy equation (8.2.3) we evaluate a logarithmic version of the model:

$$log(1 + \tilde{E}_{ki}) \quad (8.2.15)$$

that was used by Naselaris et al. [2009] to analyze the same data set; this additional nonlinear transformation could absorb

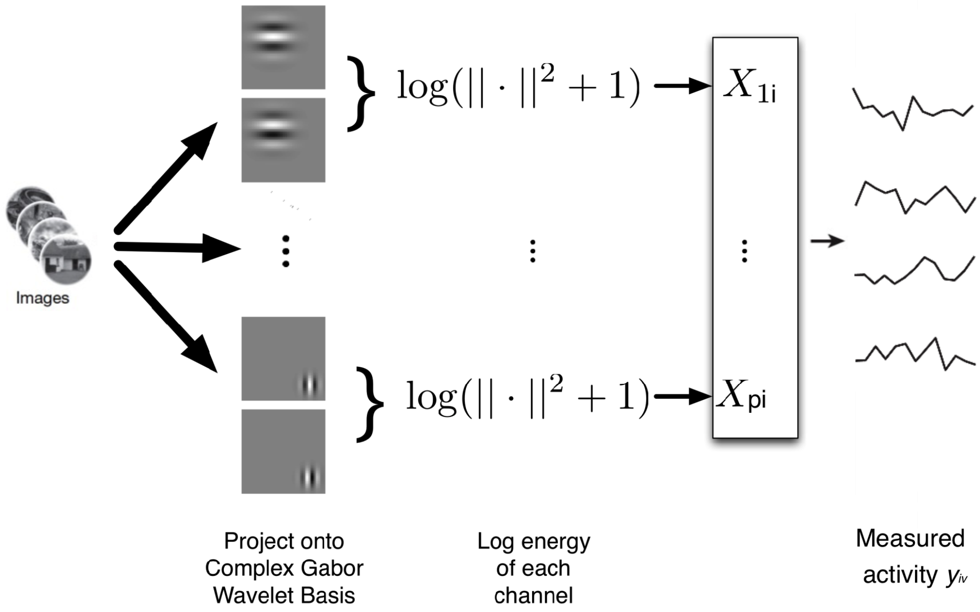


Figure 8.6: Modified model estimation stage. Adapted from: [Kay et al., 2008].

some of the residual nonlinearity in the model of Kay et al. [2008].

The new model estimation scheme is represented in Figure 8.6, where fMRI images are codified by a matrix y_{iv}.

We assume the same relation between energies defined in equation (8.2.4) and measured activities:

$$y = log(1 + \tilde{E})\beta + \beta_0 \tag{8.2.16}$$

where the coefficients $\beta = \beta_{kv}$ and the vector β_{0_v} are unknown. We use Lasso method [Tibshirani, 1996], described in Section 8.2.3.1, in order to achieve a sparse solution.

Let us consider a single voxel v. In our case, the functional we consider is:

$$F_{3,v}(\beta) = (\frac{1}{2N}\sum_{i=1}^{N}(y_{iv} - \beta_{0_v} - \sum_{k=1}^{n} log(1 + \tilde{E}_{ki})\beta_{kv})^2 + \lambda\sum_{k=1}^{n}|\beta_{kv}|) \tag{8.2.17}$$

and we minimize it:

$$min_{\beta,\beta_0} F_{3,v}(\beta) \tag{8.2.18}$$

where β and β_0 are p-vector and scalar components and $\lambda \geq 0$ is a regularization parameter.

As clarified in the previous section, when λ increases the number of nonzero components of β decreases. The first term of equation (8.2.17) is called *fiducial term*, because its minimization brings to the evaluation of coefficients and allows to obtain data closest to the measured ones.

8.2.3.3 Encoding with cortical connectivity

The encoding model presented here is a minimization in the L^2 space. However, we have seen in Section 4.1.2.3 that the functional geometry of the space of simple cells receptive profiles is endowed by the cortical connectivity with a differential structure of Subriemannian type. Hence we will introduce this physiological term in the model.

For any fixed voxel v, the unknown β of the minimization problem is a function of the variable k, which is induced by a discretization of the group parameters $g_k = (x_k, y_k, \theta_k, \sigma_k)$. For very fixed value of σ_k we have that $\beta_{vk} = \beta_{v\sigma_k}(x_k, y_k, \theta_k)$. If we consider to begin with a continuous model in the variables (x, y, θ), the unknowns $\beta_{v\sigma}$ are defined on $R^2 \times S^1$. The metric of the space has been defined in Section 4.1.2.3, and the gradient of the space is defined as:

$$\nabla_{SR}\beta_{v\sigma}(x, y, \theta) = (X_1\beta_{v\sigma}(x, y, \theta), X_2\beta_{v\sigma}(x, y, \theta)). \tag{8.2.19}$$

The Dirichlet functional in this space will be represented as:

$$||\nabla_{SR}\beta_{v,\sigma}||_2^2 = \int (|X_1\beta_{v\sigma}(x, y, \theta)|^2 + |X_2\beta_{v\sigma}(x, y, \theta)|^2)dxdyd\theta. \tag{8.2.20}$$

While restricting to the discrete space (x_k, y_k, θ_k), we substitute the derivatives with finite differences. This discretization brings to the definition of the discrete subriemannian gradient $D_{SR}\beta_{vk}$, which norm will be $\sum_k |D_{SR}\beta_{vk}|^2$.

As a consequence, the functional taking into account these terms will be:

$$F_{4,v}(\beta) = \lambda_3 \sum_k |D_{SR}\beta_{vk}|^2 + (\frac{1}{2N}\sum_{i=1}^{N}(y_{iv} - \beta_{0_v} - \sum_{k=1}^{n} log(1 + \tilde{E}_{ki})\beta_{kv})^2 + \lambda \sum_{k=1}^{n} |\beta_{kv}|) \tag{8.2.21}$$

and the final model is the minimization of the functional $F_{4,v}$:

$$minF_{4,v}(\beta). \tag{8.2.22}$$

As described in Section 8.2.3.1, the minimum is evaluated using the Lasso method with cross validation, considering both the regularization term in equation (8.2.10) and the discrete subriemannian gradient $D_{SR}\beta_{v,k}$ (equation 8.2.21). In this case, for simplicity, we use $\lambda_3 = 1$ but this parameter can be evaluated more precisely, as λ, using cross validation. We will see that the first term of equation (8.2.21) helps in the problem of reconstruction of images, while the others play a central role in the problem of identification from fMRI.

8.2.3.4 Our receptive field model

Minimizing the functional $F_{4,v}$, we are interested in finding β such that the linear combination with coefficients of the contrast energy of a set of images should approximate in the closest way the fMRI measured responses to those images, taking into account in the minimization the regularization term. In this sense, the coefficients β should give a measure of how much the receptive profile ψ_{g_k} is relevant in representing the neuronal response to an image.

In this sense the receptive profile will be reconstructed as:

$$RF_v = \sum_{k=1}^{n} \beta_{kv} |\psi_{g_k}| \tag{8.2.23}$$

An example of receptive-field model for a representative voxel is shown in Figure 8.7 and it is possible to notice the similarity between our (right) and the model described in [Kay et al., 2008] (left). Moreover, in our case, the Lasso term allows to obtain sparse receptive-field models.

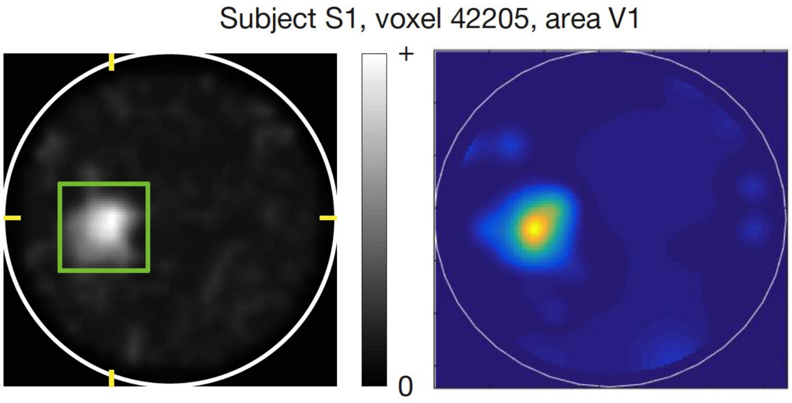


Figure 8.7: Receptive-field model for a representative voxel from Kay et al. [2008] (left) and our model (right). The intensity of each pixel indicates the sensitivity of the receptive field to that location. The white circle delineates the bounds of the stimulus.

8.2.3.5 Image identification performances

In the image identification stage, we use the new encoding model estimated to predict the voxel activity pattern evoked by the images in the validation set. The second and the third term of equation (8.2.21) play a central role to solve this problem.

The predicted image is the one yielding the highest correlation with the measured activity, using maximum likelihood and covariance matrix regularization as described in Naselaris et al. [2009]. In order to optimize the performance of the identification algorithm, we first select voxels whose receptive field models have the minimun mean square error calculated during cross validation. As in Kay et al. [2008], optimal performances are achieved using about 500 voxels; most of these voxels are located in area V1 and V2, where predictive power is higher (see Figure 8.8).

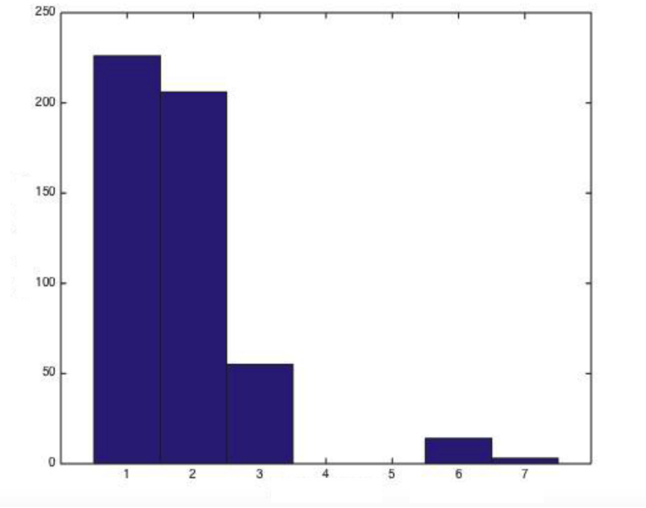


Figure 8.8: Histogram of distribution of the first 500 significative voxels. On the x axis the Region of interest (0: other, 1: V1, 2: V2, 3: V3, 4: V3A, 5: V3B, 6: V4, 7: Lateral Occipital area), on the y axis the number of voxels in that area.

To do that, we define $p(y|i)$ as the likelihood of the observed response given a sampled image:

$$p(y|i) \propto exp[-\frac{1}{2}(y - \hat{y}(i))\Lambda^{-1}(y - \hat{y}(i))^T] \tag{8.2.24}$$

where Λ is a covariance matrix, obtained as the covariance of the residuals, evaluated as the difference between the measured and the predicted response in the training set:

$$\Lambda = \langle (y - \hat{y}(i))^T(y - \hat{y}(i))\rangle. \tag{8.2.25}$$

In most cases the covariance matrix is singular or close to singular and it is not possible to calculate the inverse of Λ in a stable manner. To overcome this problem a Tikhonov regularization is used to estimate the inverse [Nishimoto et al., 2011].

The image whose predicted voxel activity pattern is most correlated (highest p(y|i)) with the measured one is selected.

Identification performances for both subjects are illustrated in Figure 8.9, comparing our results with the ones of Kay et al. [2008]. For subject S1 they obtained 92% (110/120) of the images correctly identified, whereas chance performance is just 0.8% (1/120). For subject S2, 72% (86/120) of the images were identified correctly. In our case, we obtain 99.2% of identification performances for subject S1 and 92.5% for subject S1.

The validity of our decoding method is underlined by these high performances, showing how our receptive-field models

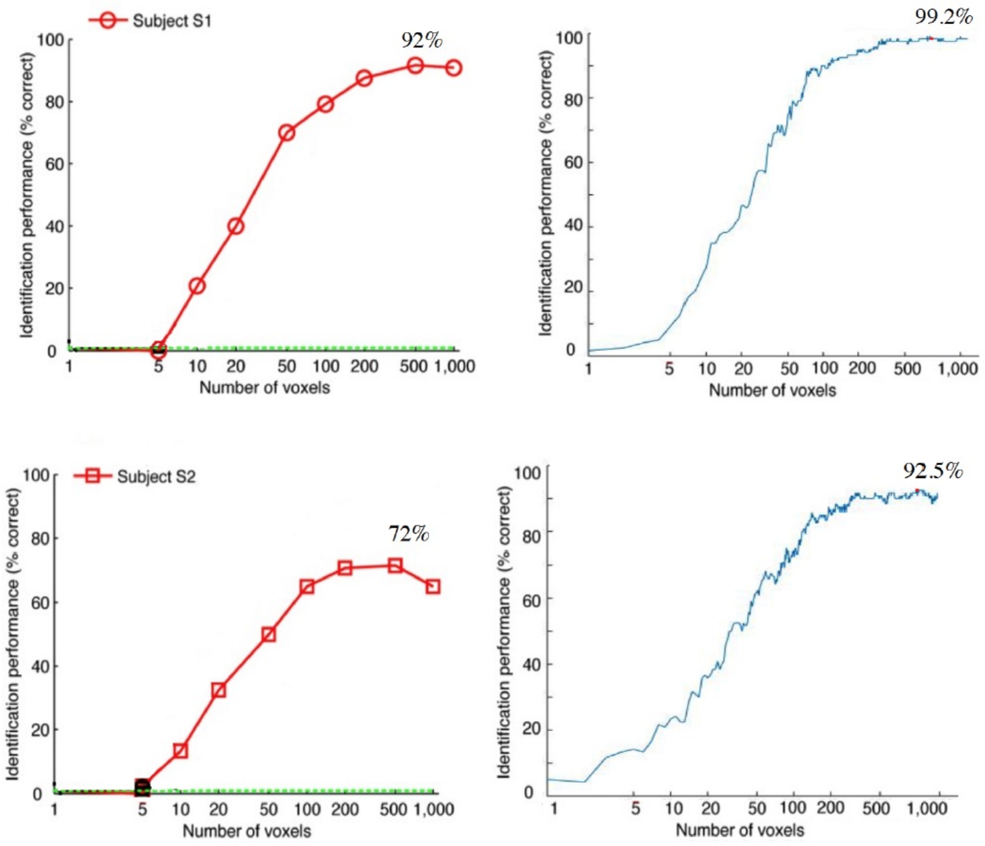


Figure 8.9: Identification performances for subjects S1 and S2, comparing Kay et al. [2008] (left) and our (right) results.

characterize the selectivity of individual voxels to natural images.

8.2.4 Analysis of fMRI data

In order to afford the problem of reconstruction of images, we first analyze fMRI data, projecting them on a standardized anatomical 3D space. For this analysis, we use FreeSurfer, a software for the elaboration and visualization of neuroimaging data [Fischl, 2012] (see Figure 8.10).

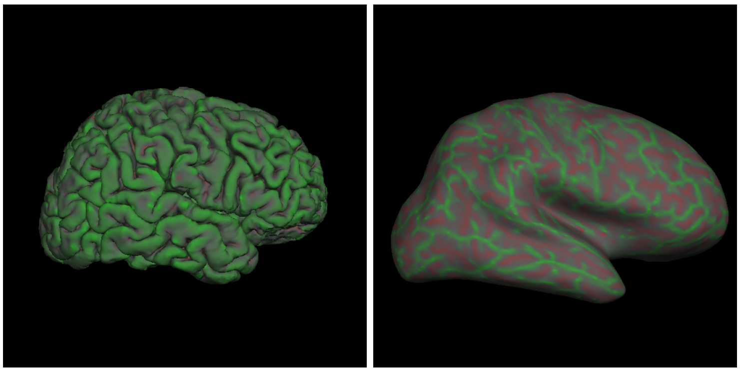


Figure 8.10: A visualization of the right hemishpere of a subject with superimposed the feature of curvature obtained through FreeSurfer. On the left the main and on the right the inflated surface.

An example of fMRI activity projected on the right hemisphere of a subject is represented in Figure 8.11.

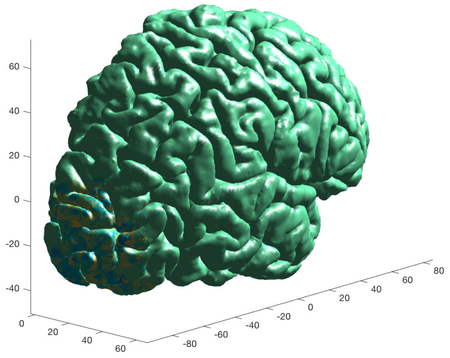


Figure 8.11: Example of fMRI activity represented on the right hemisphere of a subject.

As visualized in Figure 8.10 (left), regions which are close together in the volume may be relatively far apart along the cortical surface due to the folded structure of the brain. In order to overcome this limitation, a set of procedures for modifying the shape of the surface to obtain a more suitable representation of the cortical surface has been designed by Dale et al. [1999].

In Dale et al. [1999] it has been developed a general procedure for minimizing metric distortion in a variety of contexts, such as surface inflation, flattening, as well as mapping to other parameterizable surfaces such as a sphere. Constructing this type of mapping is a difficult task due to the complex and highly folded nature of the original surface. To obtain a 2D representation of the occipital area of the cortex and to visualize the cortical activity on it, it has been considered the flattening technique described in Fischl et al. [1999]. In particular, it is possible to flatten both the full surface and only a portion of it.

In our case we consider the occiput surface shown in Figure 8.12, this cortical patch is particularly useful for displaying the results of visual experiments. In particular we represent on this surface the informations of voxels of the primary visual cortex.

8.2.4.1 Parameters estimation of the retino-cortical model

In order to evaluate suitable values for the constants a and k of equation (3.5) (page 35) we consider the method described in Montobbio [2016]. We first create a set of representative stimuli, evaluating their contrast energy and deforming their energy images applying the coordinate change given by equation (3.5).

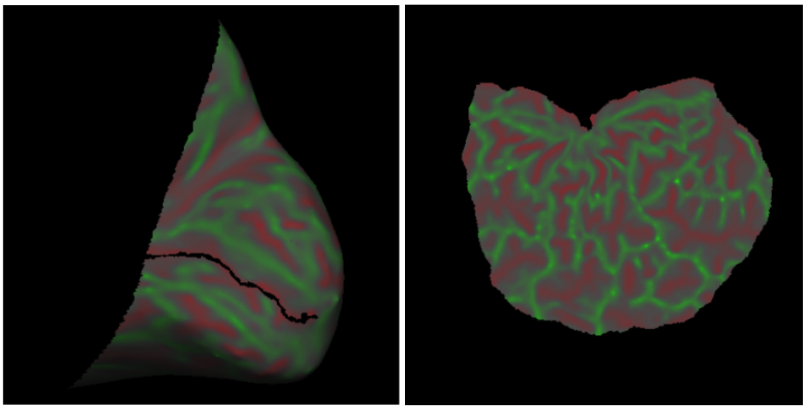


Figure 8.12: A cutting scheme for occipital flattening (left) and a visualization of the flattened left occipital lobe (right). The algorithm implemented in FreeSurfer follows the procedure described in Fischl et al. [1999].

The set of artificial stimuli is represented in Figure 8.13; grayscale images with a resolution of $128x128$ pixels, as the stimuli in Kay et al. [2008].

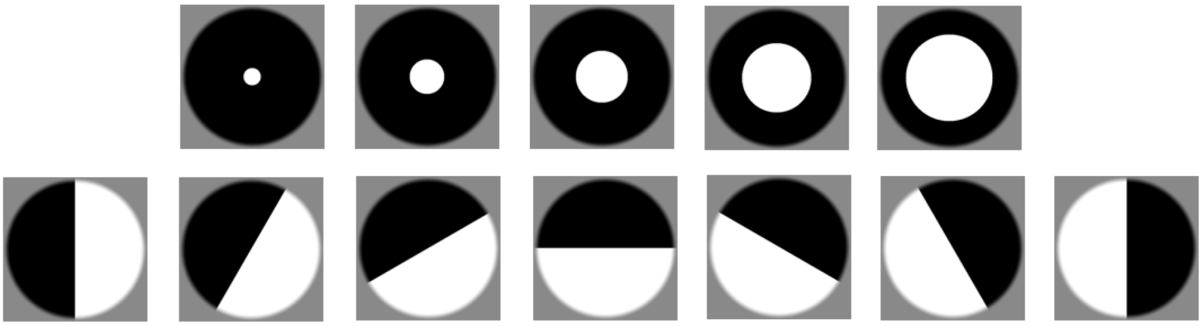


Figure 8.13: The set of artificial stimuli used in the simulation.

We used these stimuli because the result of their transformation in the cortex considering the logarithmic model is clear. Two examples of simulated fMRI responses compared with the complex logarithm mapping are visualized in Figure 8.14. Since we worked with one brain hemisphere at a time, we splitted the images in two, separating the left and right hemifields of view. The stimuli are represented on the left, where the yellow boxes correspond to the right half of the stimuli as we were considering the projection of the response on the left hemisphere of the brain, visualized in blue to yellow values in the middle of Figure 8.14. On the right the complex logarithm mapping. The red lines in the stimuli represented a contrast in the image that can be identified in the projection on the cortical surface; the same line are also visualized in the logarithm mapping.

Starting from the stimuli, we evaluate their energy considering a family of Gabor filters. Applying the coordinate change given by $l_{a,k}$, the deformed energy images are visualized. We

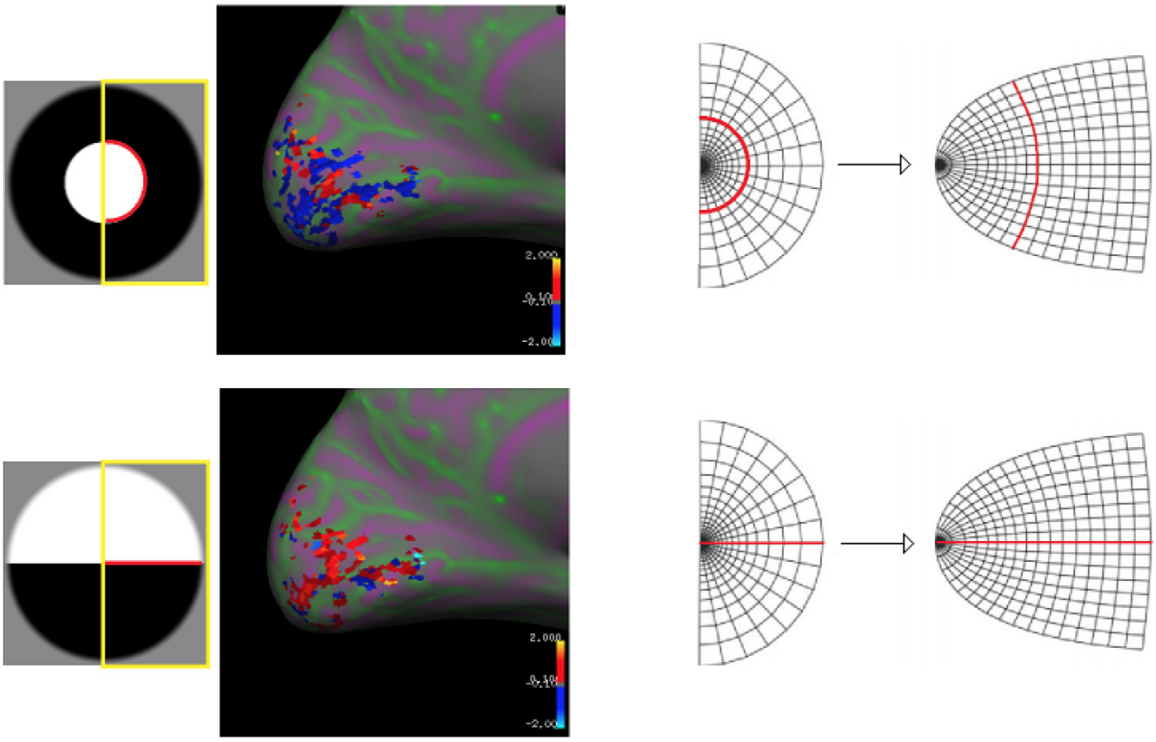


Figure 8.14: From left to right: the stimuli, the projections of the fMRI simulated activity in response to these stimuli on the left hemisphere and the complex logarithm mapping.

use the encoding model presented in Section 8.2.3 to simulate the fMRI responses to the stimuli and after selecting the voxels corresponding to V1, we represent their response on the flattened surface. In this case adding the term of regularization in the subriemannian cortical structure allows to obtain better reconstruction results.

Finally comparing the simulated fMRI activation maps with the deformed energy images, it is possible to evaluate the best fitting parameters for the constants a and k of the two-parameter logarithmic function $l_{a,k}$ (equation (3.5)). An example of this method applied to one image is shown in Figure 8.15.

To do this, it is necessary the flattening of the cortical surface and its representation in 2D coordinates. This step is performed using FreeSurfer and allows to obtain a visualization of the simulated fMRI activity on the flattened cortex.

8.2.5 Inverse mapping: reconstruction of the stimulus

We have seen in Section 8.2.2.4 and 8.2.3.5 the great performances achieved in the problem of identification of images from human brain activity. The approach presented here is an extension to the problem of reconstruction of images from fMRI

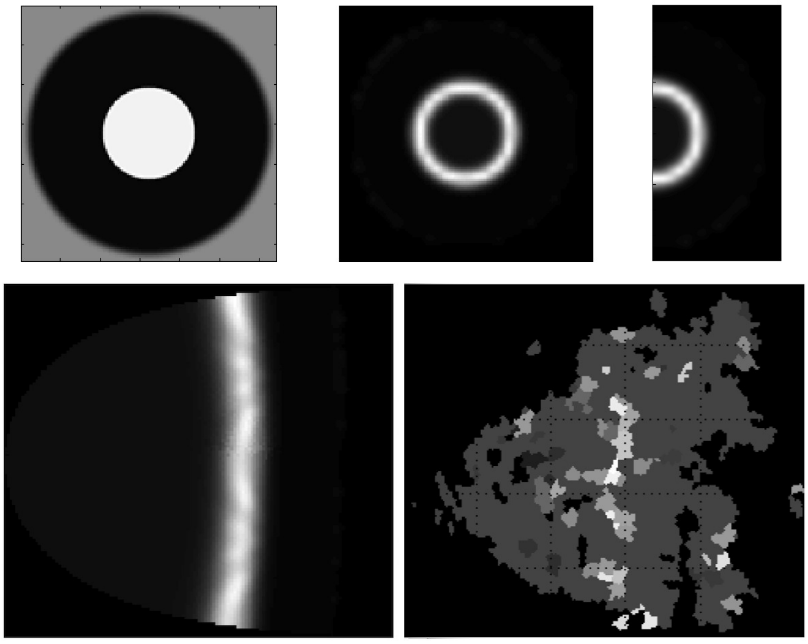


Figure 8.15: Top: on the left one of the stimuli, in the middle the energy and on the right the energy relative to its right half. Bottom: on the left the energy of the half stimulus in cortex coordinates, on the right the simulated fMRI activity on the flattened cortex.

activity, trying to reproduce the image seen by the observer knowing only its fMRI response. This still unsolved problem is called inverse mapping. The idea consists of representing the fMRI response from the flattened cortex to the retinal coordinates. In this method, the fMRI activity obtained on the flattened V1 is interpreted as the image of the logarithmic function. In order to obtain the inverse mapping and so the fMRI response in retinal coordinates, we make a coordinate change using the exponential function. In Figure 8.16 are visualized two examples: on the left the stimulus and its energy, in the middle the simulated fMRI responses to the stimuli projected on the flattened primary visual cortex, on the right the fMRI response in retinal coordinates, after applying the inverse mapping $l^{-1}_{a,k}$. In red are represented the marked contrasts of the stimuli; the white lines on the right underline the activation areas. It is possible to notice the correspondence between the original energy image and the fMRI responses; in both the examples we can notice the presence of a circle almost tangent to the borders of the stimuli and this is due to the contrast with the gray mask applied to every stimulus.

The results considering simple and artificial images seem very promising, the similarity with the original energy image is visible. For the moment, we have considered for simplicity this artificial images but the purpose is to improve the retino-cortical model in order to reproduce with this technique also

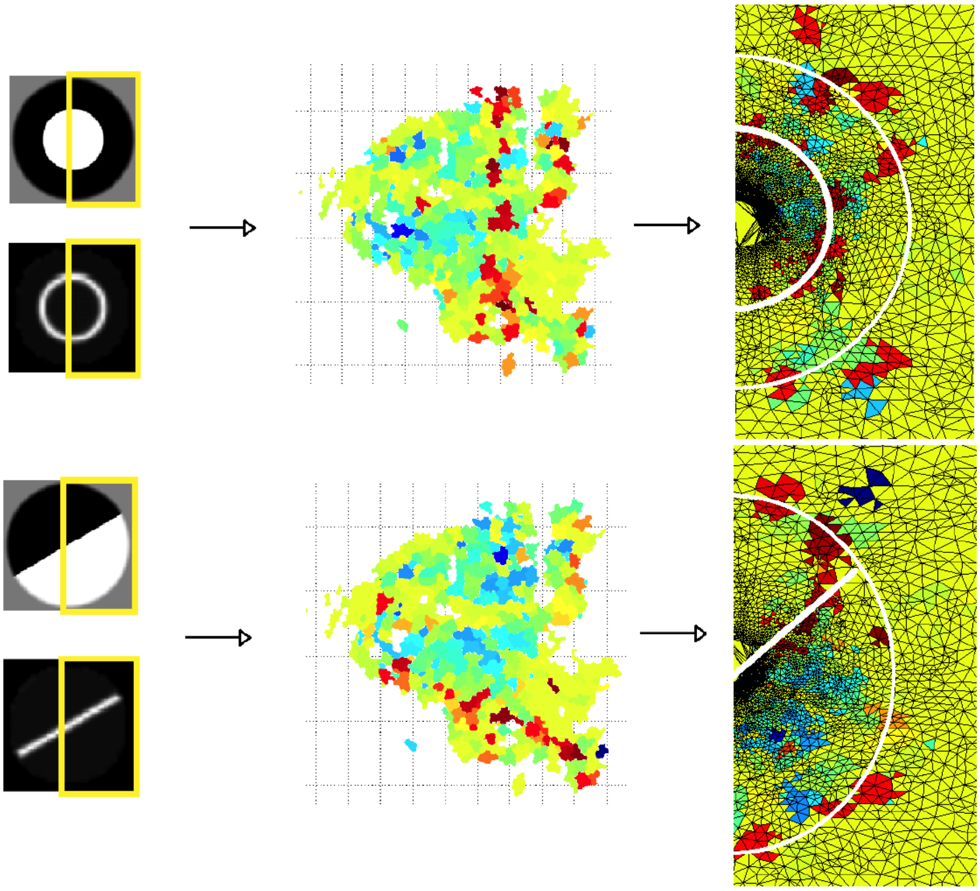


Figure 8.16: Two examples of stimuli and their energy, the fMRI responses in the flattened V1 and the fMRI responses in retinal coordinates.

natural ones.

Before considering the problem of reconstruction of natural images, we consider the 2D retinal coordinates and we project on it the Receptive Profiles of voxels of V1 and V2, each one weighted for the corresponding Pearson correlation coefficients evaluated in Section 8.2.2.4.

In this mapping pixels related to voxels with higher predictive power are represented in yellow, where most of the information is contained, while the ones related to voxels with lower predictive power are represented in blue. This will allow us to predict which part of the image can be better reconstructed.

8.3 Reconstruction of images

In this Section we afford the problem of reconstruction of images from Gabor wavelts, from the contrast energy and finally from the human brain activity. We compare the results, showing the increasing difficulty in the resolution of this problem. We will present a model for the retinal receptive fields showing the reconstruction of images obtained considering a simi-

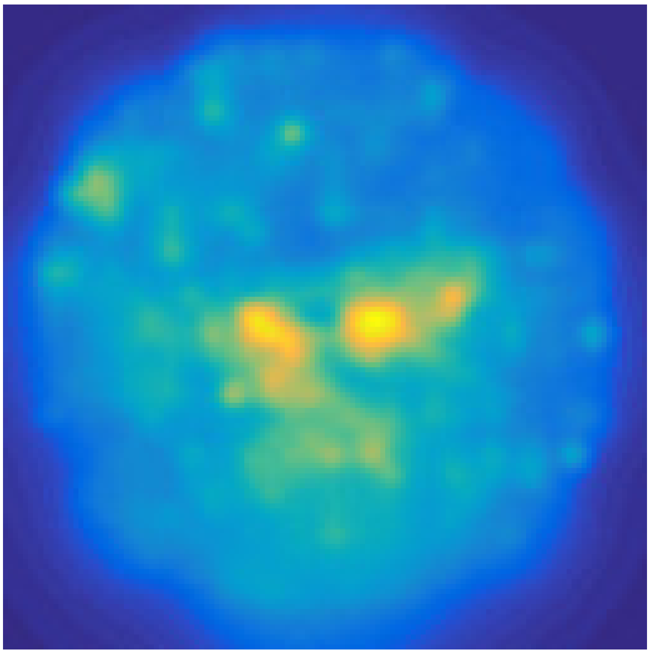


Figure 8.17: Projection of the Pearson correlation coefficients in retinal coordinates.

lar distribution of wavelets and underlying how the neurophysiological constraints about the distribution of wavelets in the retina can represent a limit in the reconstruction of images. We will also see how a different positions of wavelets in the image determin a good reconstruction nearby that area.

8.3.1 Reconstruction of images from Gabor wavelets

Gabor filters provide the most famous example of continuous wavelet decomposition. The basis idea is to reconstruct a function I in L^2 from its Gabor transform, which is nothing but the action of Gabor filters defined on Section 3.2.1.

Definition 5. *If $I \in L^2$ we call Gabor Transform of the function f the lifted function defined as:*

$$h_I(g) = \int \psi_g(x,y) I(x,y) dxdy,$$

where $g = (x, y, \theta, \sigma)$.

[Lee, 1996] derives the conditions under which a set of continuous 2D Gabor wavelets will provide a complete representation of any image and how it is possible to obtain their reconstruction. Precisely a function $I \in L^2$ can be reconstructed from its Gabor transform.

Proposition 6. *Any function $I \in L^2$ can be represented as:*

$$I(x,y) = \int h_I(g) \bar{\psi}_g(x,y) dg$$

where $g = (x, y, \theta, \sigma)$.

The same result is valid for a discretized family of Gabor filters, if the choosen filters are correctly overlapping. In this section we describe this result applied to our dataset.

In particular we use the following algorithm:

- let $I(i)$ be an image and ψ_{g_k} a Gabor pair;
- evaluate the Gabor wavelet transform as the integral $h_I(g_k)$ between the image $I(i)$ and the Gabor filters ψ_{g_k},

$$h_I(g_k) = \sum_i I(i)\psi_{g_k}(i).$$

- apply the Inverse Gabor wavelet transform to obtain the reconstructed image, using the result of proposition

$$I(i) = \sum_k h_I(g_k)\bar{\psi}_{g_k}(i).$$

Naturally, the reconstruction of images depends on the size, the position and the orientation of the Gabor wavelets used. In our case we used the Gabor wavelet pyramid represented in Figure 8.3.

Figure 8.18 illustrates 4 examples of original images (left) and their reconstructions from Gabor wavelets (right). These results show how it is possible to reconstruct images using Direct and Inverse Gabor wavelet transform.

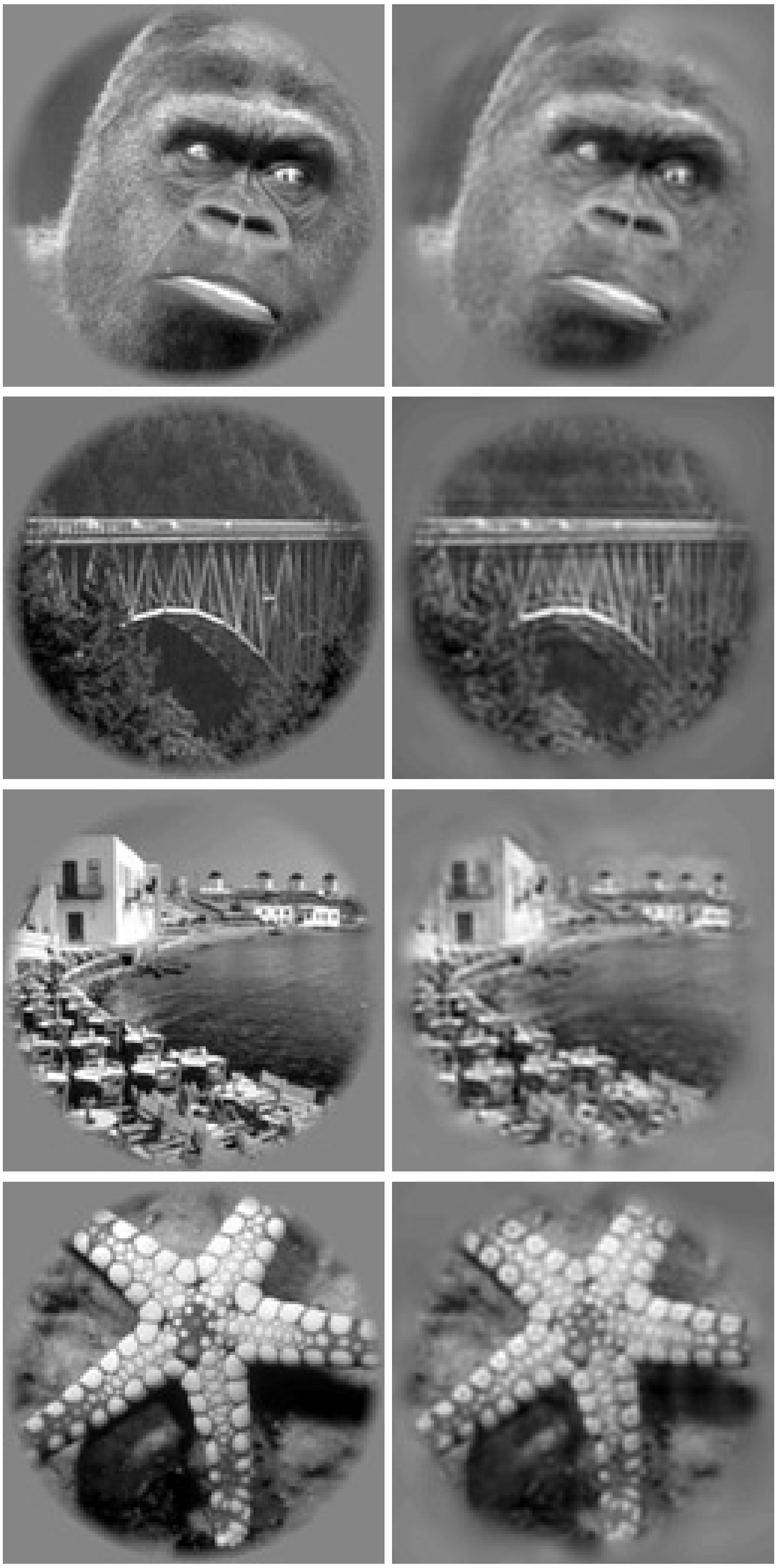


Figure 8.18: Examples of images (left) and their reconstructions from Gabor wavelets (right).

8.3.2 Reconstruction of images from the contrast energy

In Shams and Von Der Malsburg [2002] it is shown the role of complex cells in object recognition, in particular it is underlined how population responses contain sufficient information to capture the perceptual essence of images. Formally this reduces to reconstruct images from the contrast energy defined in equation (8.2.3).

Formally they proved the following Theorem, called "Gabor Magnitude Theorem".

Theorem 3. *If I_1, I_2 are images with the same Gabor Energy $E_{I1} = E_{I2}$ then it follows that $I1 = \pm I2$.*

Figure 8.19 illustrates a diagram of the reconstruction algorithm.

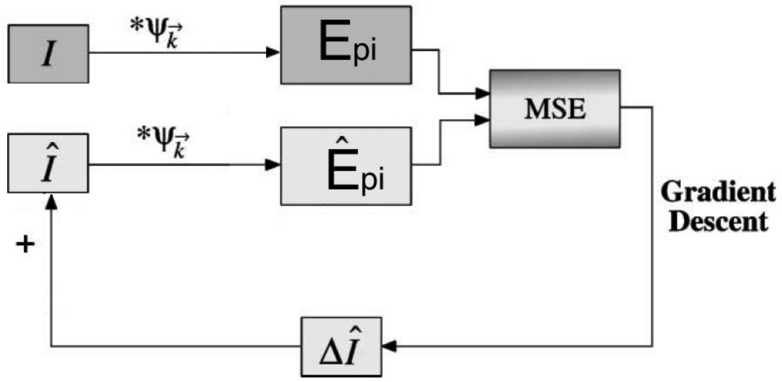


Figure 8.19: The iterative algorithm used to reconstruct target image I from the contrast energy. Adapted from: [Shams and Von Der Malsburg, 2002].

As a consequence, following Shams and Von Der Malsburg [2002], we use this algorithm to recover an image from its Gabor Energy:

- let I_i be an image and E_{ki} the Gabor energy associated to a Gabor pair ψ_{g_k} ;
- starting with an arbitrary seed image, iteratively compute a sequence of images $\hat{I}_i$ whose energies $\hat{E}_{ki}$ approximate those of I_i with higher accuracy;
- compute the mean squared error between the squared magnitudes of the target and trial image transforms respectively, using the gradient descent method: $min_{\hat{I}_i} \sum_k (E^2_{ki} - \hat{E}^2_{ki})^2$;
- defined ϵ a parameter to scale the update speed, update $\Delta\hat{I}_i$ to the trial image, where:
  $\Delta\hat{I}_i = \epsilon(E^2_{ki} - \hat{E}^2_{ki})(\sum_k \psi_{g_k} \hat{I}_i \psi_{g_k})$

- stop the iteration when the median error $e = \frac{(E_{ki} - \hat{E}_{ki})}{E_{ki}}$ falls below a threshold.

In particular, considering the starting image $\hat{I}$ evaluated as:

$$\hat{I} = \sum_{k} \frac{E_{ki}}{2}(Re(\psi_{g_k}) + Im(\psi_{g_k})), \tag{8.3.1}$$

we obtain the reconstructions shown in Figure 8.20.

The target images are shown in the first column. For each target image two reconstructions are displayed (in the second and in the third column). Starting from the contrast energy, that contains both the real and the imaginary part of the Gabor filter, it is not possible to discriminate between them and this is underlined in the reconstruction results visualized in Figure 8.20: we can obtain with the same probability the reconstruction visualized in the second column or its inverse, represented in the third column.

These results show how the representation provided by a population of complex cells implicitly encodes the phase information that is needed for object recognition.

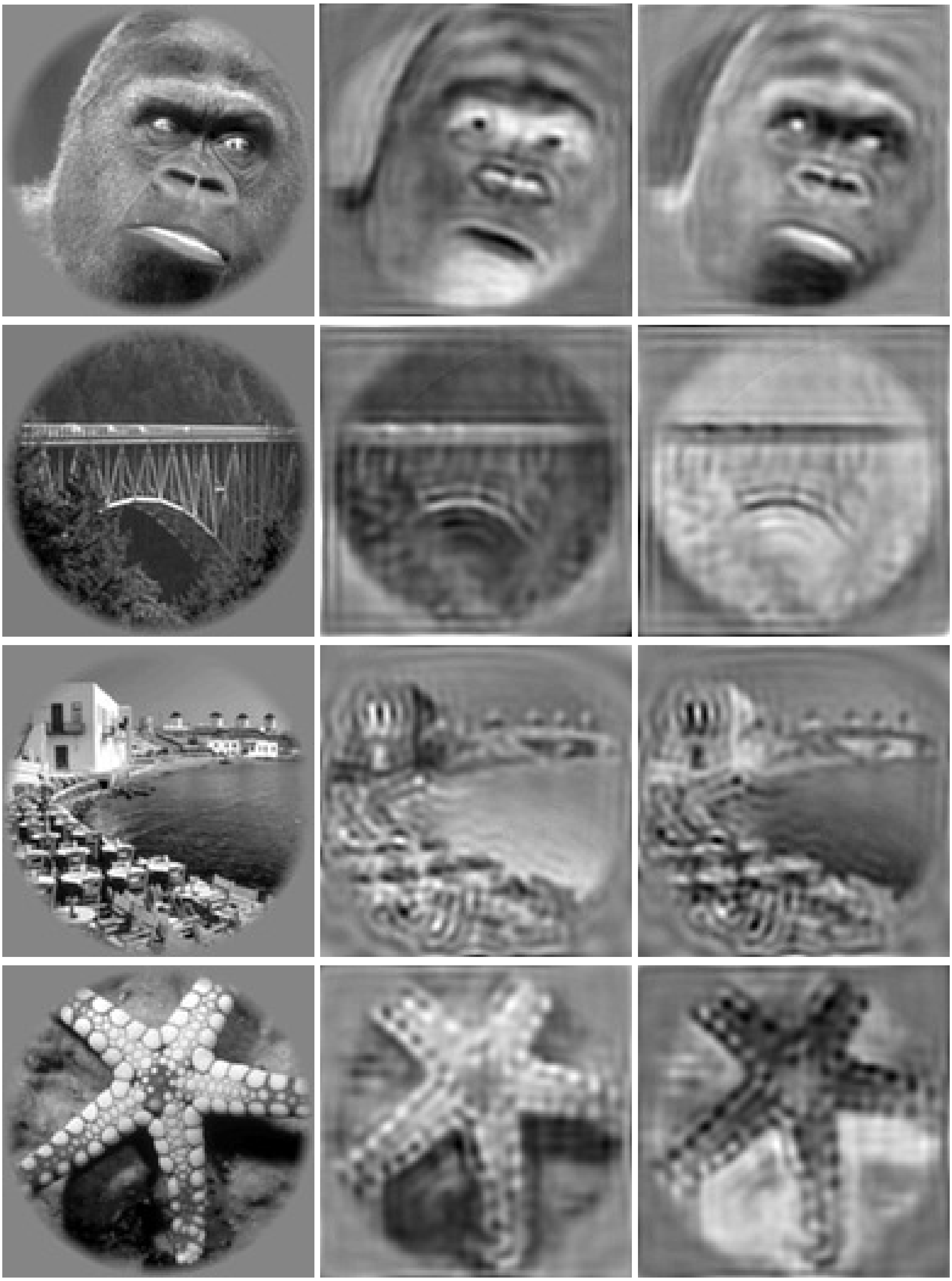


Figure 8.20: Reconstruction from the energy: the target images are shown in the first column, for each target image one reconstruction and its inverse are displayed in the second and third column.

8.3.3 Reconstruction of images from fMRI activity

8.3.3.1 Overview of previous results

Although the goal of both identification and reconstruction of images from human brain activity is to determine the specific image that produced a given activity pattern, in the identification a set of images is provided, in the recostruction no set is provided.

Previous fMRI decoding papers presented algorithms for reconstructing the spatial layout of simple geometrical patterns composed of high-contrast flicker patches [Thirion et al., 2006, Miyawaki et al., 2008]. In Naselaris et al. [2009] it is shown how combining a structural encoding model that characterizes responses in early visual areas, a semantic encoding model that characterizes responses in anterior visual areas and prior information about the structure and semantic content of natural images, it is possible to obtain a decoder that produces reconstructions. Reconstructions with structural encoding model and two different types of prior information are visualized in Figure 8.21. In the first columns, two target images are shown; column two shows reconstructions obtained using a flat prior that does not bias reconstructions. Flat prior assigns the same probability to all possible images. Regions of the target images that have low texture contrast are depicted as smooth gray patches, while regions that have substantial texture contrast are depicted as textured patches. The flat prior reconstructions reveal the distribution of texture contrast in the target images but cannot readily be interpreted. Reconstructions obtained using a sparse Gabor prior are shown in the third column. This ensures that reconstructions possess the lower-order statistical properties of natural images. These reconstructions appear to be smoothed versions of those obtained with the flat prior and they also cannot be readily interpreted (numbers in bottom right corner of the second row of image reconstructions indicate structural accuracy, see Naselaris et al. [2009] for details).

Moreover, in Nishimoto et al. [2011] they constructed a Bayesian decoder that provides remarkable reconstruction of the viewed movies from BOLD signals.

In Section 8.2.5 it is presented an approach to solve the problem of inverse mapping, applied to artificial and simple stimuli. Starting from the encoding model presented in Section 8.2.3, we study in Section 8.3.3.2 the problem of reconstruction of images from human brain activity.

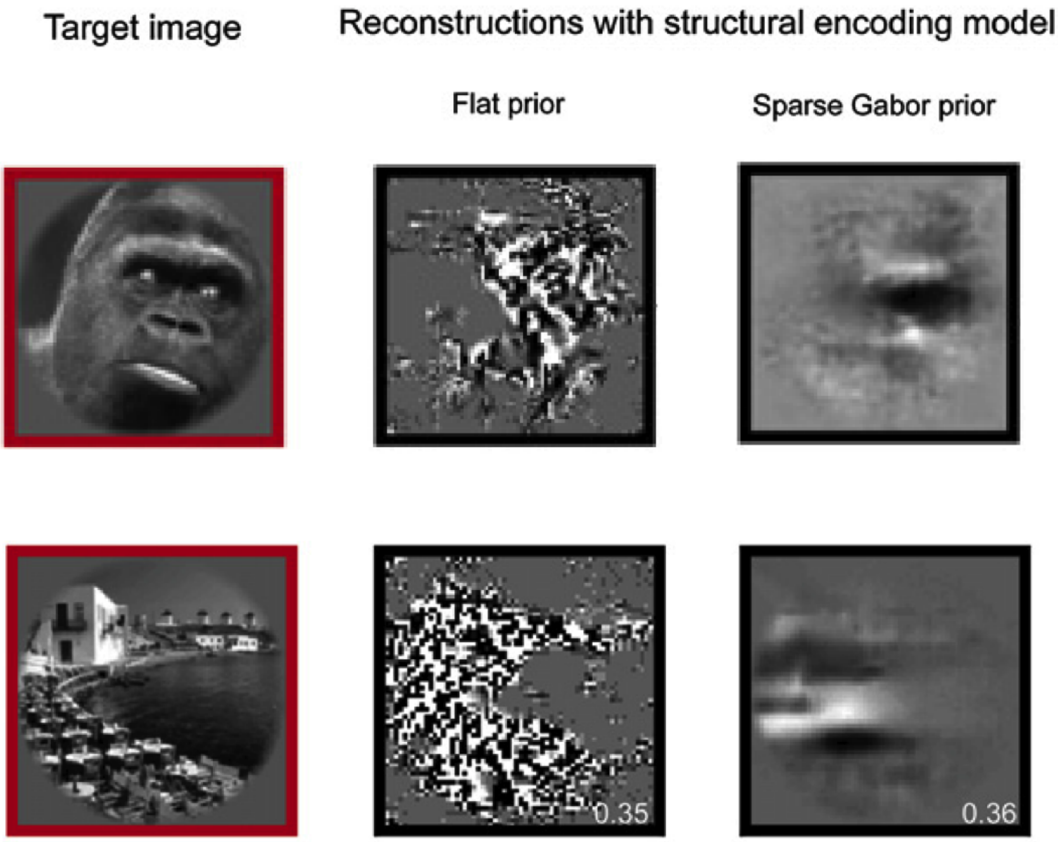


Figure 8.21: Two target images are shown in the first column. The second and the third columns show reconstructions obtained using the structural encoding model and two different types of prior information. Source: [Naselaris et al., 2009].

8.3.3.2 An approach to a new functional

Considering a starting image I_0 evaluated as the mean of 10 images of the dataset, we evaluate the distance between the fMRI measured activity and the predicted one as:

$$f(I) = \frac{1}{2}\sum_{v=1}^{N}(y_{iv} - \sum_{k=1}^{n} log(1 + \tilde{E}_{ki}(I))\beta_{kv})^2 \tag{8.3.2}$$

where the coefficients β_{kv} have been evaluated with the Lasso method described in Section 8.2.3 and E_{ik} is the logarithm version of the contrast energy defined in equation (8.2.15).

Considering the derivative of equation (8.3.2), we minimize with the help of gradient descent and we obtain:

$$\frac{\partial f}{\partial E_{ki}} = \sum_{v=1}^{N} \frac{\beta_{vk}(y_{iv} - log(1 + \tilde{E}_{ki})\beta_{kv})}{1 + log(1 + \tilde{E}_{ki})} \tag{8.3.3}$$

and the reconstructed image can be written as:

$$\nabla f(I) = \sum_{k} \frac{\partial f}{\partial E_{ik}}(Re(\psi_{g_k}) + Im(\psi_{g_k})). \tag{8.3.4}$$

Defined the reconstructed image as:

$$\begin{cases} \frac{\partial I}{\partial t} = \nabla f(I); \\ I(0) = I_0; \end{cases} \tag{8.3.5}$$

we add a regularization term:

$$\begin{cases} \frac{\partial I}{\partial t} = \nabla f(I) + \lambda_s \Delta(I); \\ I(0) = I_0; \end{cases} \tag{8.3.6}$$

where $\Delta(I)$ is the Laplacian of the image, that allows to regularize in the 2D domain and where $\lambda_s = 0.002$ is a constant that contains the speed of convergence of the algorithm. This process is then iterated until it converges and we obtain the results of simulation visualized in Figure 8.22.

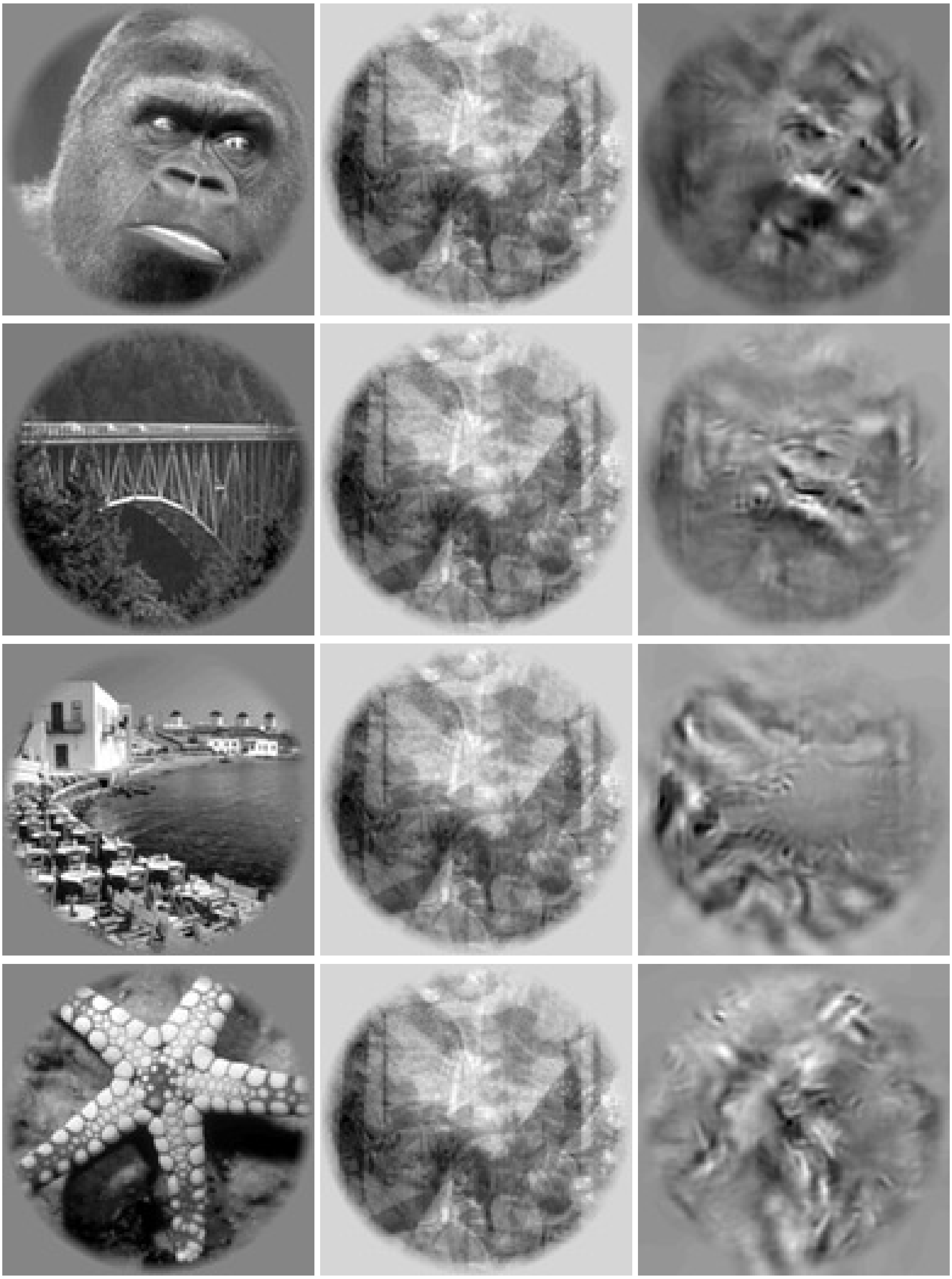


Figure 8.22: Reconstruction from fMRI. From first to third column: the target images, the starting image and for each target image I_0 one reconstruction is displayed.

A comparison betweenthe results obtained by Naselaris et al. [2009] and our method is represented in Figure 8.23, where the first and the third target images of Figure 8.22 have been considered. It is possible to notice that both methods contain informations about the stimuli. In particular they are both able to determine the regions of the image in which the informations are. Moreover with our model we notice the presence of more distinct informations regarding the stimuli: in the first reconstruction it is possible to notice the details of the face (the monkey's eyes, nose and mouth), in the second the presence of arcs of the bridge, in the third the contour delimiting the sea and in the last the 5-point of the star.

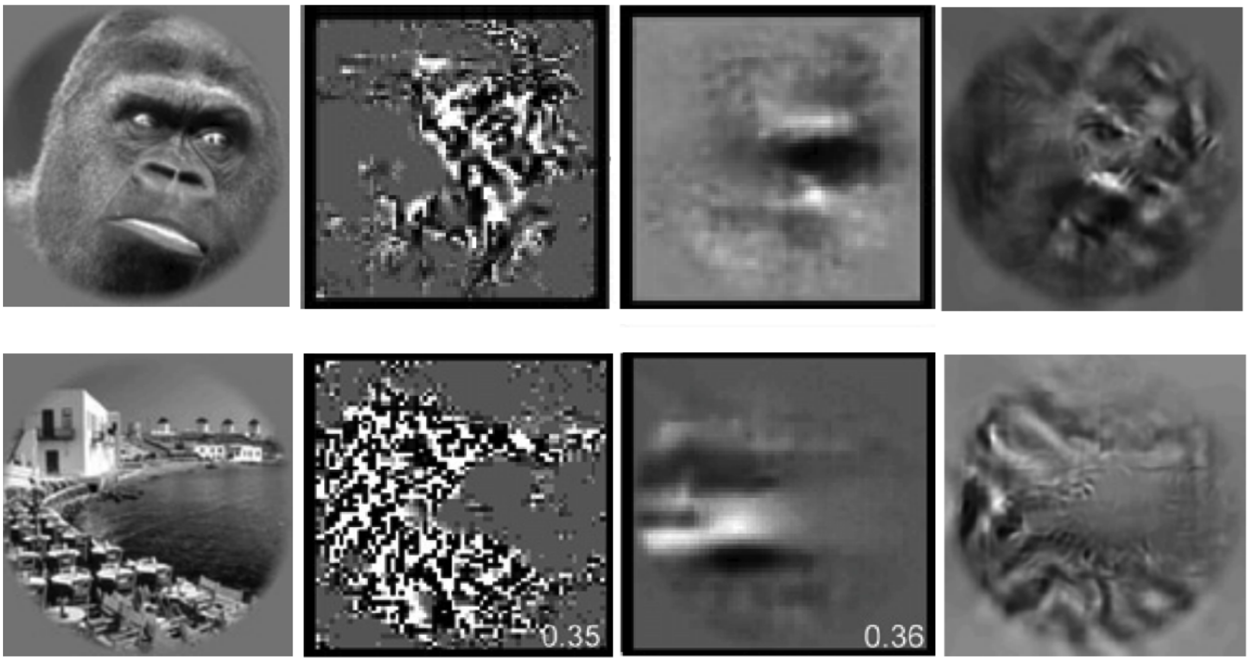


Figure 8.23: Comparison between the results of Naselaris et al. [2009] and our method for the first and the third target images visualized in Figure 8.22.

The map in Figure 8.17, that contains pixels related to voxels with higher predictive power, is in good agreement with the reconstructions obtained in Figure 8.22: the parts of the images correctly reconstructed correspond to the pixels with higher predictive power.

Even if it is not a trial and already solved problem, we think that this could be a first step to improve the results in the field of reconstruction of images from fMRI activity and we guess that considering fMRI data with higher resolution could help in these reconstructions.

9 Conclusions

The aim of this thesis is to develop mathematical models of visual perception based on cortical architecture and to apply them both on phenomenological experiments and on natural images.

In this work we have presented a neurally based model for figure-ground segmentation and grouping of different perceptual units using spectral methods, where segmentation has been performed by computing eigenvectors of affinity matrices. Our objective was to model these experimental data with a suitable mathematical framework and different connectivity kernels that are compatible with the functional architecture of the primary visual cortex have been presented. We have modelled them as fundamental solution of Fokker-Planck, Sub-Riemannian Laplacian and isotropic Laplacian equations and compared their properties. With this model we have identified perceptual units of different illusory Kanizsa figures [Favali et al., 2016b]d, showing that this can be considered a good quantitative model for the constitution of perceptual units equipped by their saliency. The proposed mathematical model was then able to integrate local and global gestalt laws as a process implemented in the functional architecture of the visual cortex.

The same method has been applied to retinal images [Favali et al., 2016a]. An extension to more general kernels able to detect geometrical features different from orientation as curvature [Abbasi-Sureshjani et al., 2016a] has been described. Based on the results shown in the numerical simulations, we were able to detect the salient groups in retinal images and this can be considered as an excellent quantitative model for the constitution of perceptual units.

Finally we have considered the problem of identification and reconstruction of natural images from human brain activity and we have integrated the structure of the cortex, described through the cortical models previously defined, with fMRI data. In particular we have proposed a first approach to improve these results adding into the analysis neurophysiological features, even if it is not an already solved problem.

This thesis leaves more than one open questions:

- concerning the connectivity kernels, an extension to other parts of the visual cortex (like V2) can be analysed, simulating the behaviour of cells in subsequent processing stages of the neural visual path;
- new experiments, as bistable illusory figures, could be evaluated to test some aspect of the model and to afford more complicated cases;
- another interesting application of these models is the possibility to train the connectivity kernels on the basis of past experience, including this feature in the connections to account for a priori knowledge;
- considering retinal images, the method has been examined only on small patches to reduce the computational complexity. However, this can easily be improved from implementation and hardware points of view and full retinal images can be analyzed;
- considering fMRI data with higher resolution and bigger dataset could help in the problem of identification of images from fMRI activity;
- improving the retino cortical mapping and considering the flattened V2, could bring to better reconstruction of images from artificial stimuli and from fMRI activity.

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

10 Bibliography

Samaneh Abbasi-Sureshjani, Iris Smit-Ockeloen, Jiong Zhang, and Bart Ter Haar Romeny. Biologically-inspired supervised vasculature segmentation in SLO retinal fundus images. In *Image Analysis and Recognition*, volume 9164, pages 325–334. Springer, 2015.

Samaneh Abbasi-Sureshjani, Marta Favali, Giovanna Citti, Alessandro Sarti, and Bart M Romeny. Cortically-inspired spectral clustering for connectivity analysis in retinal images: Curvature integration. *arXiv preprint arXiv:1608.08049*, 2016a.

Samaneh Abbasi-Sureshjani, Iris Smit-Ockeloen, Erik Bekkers, Behdad Dashtbozorg, and Bart ter Haar Romeny. Automatic detection of vascular bifurcations and crossings in retinal images using orientation scores. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pages 189–192. IEEE, 2016b.

Bashir Al-Diri, Andrew Hunter, and David Steel. An active contour model for segmenting and measuring retinal vessels. *IEEE Transactions on Medical Imaging*, 28(9):1488–1497, 2009.

Alessandra Angelucci, Jonathan B Levitt, Emma JS Walton, Jean-Michel Hupe, Jean Bullier, and Jennifer S Lund. Circuits for local and global signal integration in primary visual cortex. *The Journal of Neuroscience*, 22(19):8633–8646, 2002.

Roberto Annunziata, Ahmad Kheirkhah, Shruti Aggarwal, Pedram Hamrah, and Emanuele Trucco. A fully automated tortuosity quantification system with application to corneal nerve fibres in confocal microscopy images. *Medical image analysis*, 32:216–232, 2016.

Jonas August and Steven W Zucker. The curve indicator random field: Curve organization via edge correlation. In *Perceptual organization for artificial vision systems*, pages 265–288. Springer, 2000.

Jonas August and Steven W. Zucker. Sketches with curvature: The curve indicator random field and markov processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(4): 387–400, 2003.

Davide Barbieri, Giovanna Citti, Gonzalo Sanguinetti, and Alessandro Sarti. An uncertainty principle underlying the functional architecture of v1. *Journal of Physiology-Paris*, 106(5):183–193, 2012.

Erik Bekkers, Remco Duits, Tos Berendschot, and Bart ter Haar Romeny. A multi-orientation analysis approach to retinal vessel tracking. *Journal of Mathematical Imaging and Vision*, pages 1–28, 2014. ISSN 0924-9907.

Erik J Bekkers, Jiong Zhang, Remco Duits, and Bart M ter Haar Romeny. Curvature based biomarkers for diabetic retinopathy via exponential curve fits in se (2). 2015.

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

André Bellaïche. The tangent space in sub-riemannian geometry. In *Sub-Riemannian geometry*, pages 1–78. Springer, 1996.

Ohad Ben-Shahar and Steven Zucker. Geometrical computations explain projection patterns of long-range horizontal connections in visual cortex. *Neural computation*, 16(3):445–476, 2004.

Tobias Bonhoeffer, Amiram Grinvald, et al. Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature*, 353(6343):429–431, 1991.

Ugo Boscain, Jean Duplaix, Jean-Paul Gauthier, and Francesco Rossi. Anthropomorphic image reconstruction via hypoelliptic diffusion. *SIAM Journal on Control and Optimization*, 50(3):1309–1336, 2012.

William H Bosking, Ying Zhang, Brett Schofield, and David Fitzpatrick. Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *The Journal of neuroscience*, 17(6):2112–2127, 1997.

Paul C Bressloff and Jack D Cowan. The functional geometry of local and horizontal connections in a model of v1. *Journal of Physiology-Paris*, 97(2):221–236, 2003.

Paul C Bressloff, Jack D Cowan, Martin Golubitsky, Peter J Thomas, and Matthew C Wiener. What geometric visual hallucinations tell us about the visual cortex. *Neural Computation*, 14(3): 473–491, 2002.

Katja Bühler, Petr Felkel, and Alexandra La Cruz. *Geometric methods for vessel visualization and quantification—a survey*. Springer, 2004.

Ali Can, Hong Shen, James N Turner, Howard L Tanenbaum, and Badrinath Roysam. Rapid automated tracing and feature extraction from retinal fundus images using direct exploratory algorithms. *IEEE Transactions on Information Technology in Biomedicine*, 3(2):125–138, 1999.

Mo Chen, Junwei Han, Xintao Hu, Xi Jiang, Lei Guo, and Tianming Liu. Survey of encoding and decoding of visual stimulus via fmri: an image analysis perspective. *Brain imaging and behavior*, 8(1):7–23, 2014.

Pascal Chossat and Olivier Faugeras. Hyperbolic planforms in relation to visual edges and textures perception. *PLoS Comput Biol*, 5(12):e1000625, 2009.

O Chutatape, Liu Zheng, and SM Krishnan. Retinal blood vessel detection and tracking by matched Gaussian and Kalman filters. In *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE*, volume 6, pages 3144–3149. IEEE, 1998.

Giovanna Citti and Alessandro Sarti. A cortical based model of perceptual completion in the roto-translation space. *Journal of Mathematical Imaging and Vision*, 24(3):307–326, 2006.

Giovanna Citti, Loukas Grafakos, Carlos Pérez, Alessandro Sarti, and Xiao Zhong. *Harmonic and Geometric Analysis*. Springer, 2015.

Giacomo Cocci. Spatio-temporal models of the functional architecture of the visual cortex. 2014.

Giacomo Cocci, Davide Barbieri, Giovanna Citti, and Alessandro Sarti. Cortical spatiotemporal dimensionality reduction for visual grouping. *Neural computation*, 2015.

Ronald R Coifman and Stéphane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006.

David D Cox and Robert L Savoy. Functional magnetic resonance imaging (fmri)“brain reading”: detecting and classifying distributed patterns of fmri activity in human visual cortex. *Neuroimage*, 19(2):261–270, 2003.

Anders M Dale, Bruce Fischl, and Martin I Sereno. Cortical surface-based analysis: I. segmentation and surface reconstruction. *Neuroimage*, 9(2):179–194, 1999.

PM Daniel and D Whitteridge. The representation of the visual field on the cerebral cortex in monkeys. *The Journal of physiology*, 159(2):203–221, 1961.

John G Daugman. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision research*, 20(10):847–856, 1980.

John G Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *JOSA A*, 2(7):1160–1169, 1985.

PETER Dayan, LF Abbott, and L Abbott. Theoretical neuroscience: computational and mathematical modeling of neural systems. *Philosophical Psychology*, pages 563–577, 2001.

Jaydeep De, Tengfei Ma, Huiqi Li, Manoranjan Dash, and Cheng Li. Automated tracing of retinal blood vessels using graphical models. In *Image Analysis*, pages 277–289. Springer, 2013.

Jaydeep De, Huiqi Li, and Li Cheng. Tracing retinal vessel trees by transductive inference. *BMC Bioinformatics*, 15(1):20, 2014.

Konstantinos K Delibasis, Aristides I Kechriniotis, C Tsonos, and Nicholas Assimakis. Automatic model-based tracing algorithm for vessel segmentation and diameter estimation. *Computer Methods and Programs in Biomedicine*, 100(2):108–122, 2010.

Allan Dobbins, Steven W Zucker, and Max S Cynader. Endstopped neurons in the visual cortex as a substrate for calculating curvature. *Nature*, 329(6138):438–441, 1987.

Allan Dobbins, Steven W Zucker, and Max S Cynader. Endstopping and curvature. *Vision research*, 29(10):1371–1387, 1989.

R Duits, MHJ Janssen, J Hannink, and GR Sanguinetti. Locally adaptive frames in the roto-translation group and their applications in medical imaging. *Journal of Mathematical Imaging and Vision*, pages 1–36, 2016.

Remco Duits and EM 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). *Quarterly of Applied Mathematics, AMS*, 68(2):255–292, 2010a.

Remco Duits and Erik Franken. Line enhancement and completion via linear left invariant scale spaces on se (2). In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 795–807. Springer, 2009.

Remco Duits and Erik 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. *Quarterly of Applied Mathematics, AMS*, 68(2):293–331, 2010b.

Remco Duits and Markus Van Almsick. The explicit solutions of linear left-invariant second order stochastic evolution equations on the 2d euclidean motion group. *Quarterly of Applied Mathematics*, pages 27–67, 2008.

Bradley Efron, Trevor Hastie, Iain Johnstone, Robert Tibshirani, et al. Least angle regression. *The Annals of statistics*, 32(2):407–499, 2004.

G Bard Ermentrout and JD Cowan. Large scale spatially organized activity in neural nets. *SIAM Journal on Applied Mathematics*, 38(1):1–21, 1980.

Olivier Faugeras, Romain Veltz, and Francois Grimbert. Persistent neural states: stationary localized activity patterns in nonlinear continuous n-population, q-dimensional neural networks. *Neural Computation*, 21(1):147–187, 2009.

Marta Favali, Samaneh Abbasi-Sureshjani, Bart Haar Romeny, and Alessandro Sarti. Analysis of vessel connectivities in retinal images by cortically inspired spectral clustering. *Journal of Mathematical Imaging and Vision*, 56(1):158–172, 2016a.

Marta Favali, Giovanna Citti, and Alessandro Sarti. Local and global gestalt laws: A neurally based spectral approach. *accepted to Neural Computation, arXiv preprint arXiv:1512.06566*, 2016b.

Petr Felkel, Rainer Wegenkittl, and Armin Kanitsar. Vessel tracking in peripheral CTA datasets-an overview. In *Computer Graphics, Spring Conference on, 2001.*, pages 232–239. IEEE, 2001.

David J Field, Anthony Hayes, and Robert F Hess. Contour integration by the human visual system: Evidence for a local "association field". *Vision research*, 33(2):173–193, 1993.

Bruce Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012.

Bruce Fischl, Martin I Sereno, and Anders M Dale. Cortical surface-based analysis: Ii: inflation, flattening, and a surface-based coordinate system. *Neuroimage*, 9(2):195–207, 1999.

Marco Foracchia, Enrico Grisan, and Alfredo Ruggeri. Luminosity and contrast normalization in retinal images. *Medical Image Analysis*, 9(3):179 – 190, 2005. ISSN 1361-8415.

Erik Franken and Remco Duits. Crossing-preserving coherence-enhancing diffusion on invertible orientation scores. *International Journal of Computer Vision*, 85(3):253–278, 2009.

Erik Franken, Remco Duits, and Bart ter Haar Romeny. Nonlinear diffusion on the 2d euclidean motion group. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 461–472. Springer, 2007.

M.M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A.R. Rudnicka, C.G. Owen, and S.A. Barman. Blood vessel segmentation methodologies in retinal images – a survey. *Computer Methods and Programs in Biomedicine*, 108(1):407 – 433, 2012. ISSN 0169-2607.

Jerome Friedman, Trevor Hastie, Holger Höfling, Robert Tibshirani, et al. Pathwise coordinate optimization. *The Annals of Applied Statistics*, 1(2):302–332, 2007.

Wulfram Gerstner, Andreas K Kreiter, Henry Markram, and Andreas VM Herz. Neural codes: firing rates and beyond. *Proceedings of the National Academy of Sciences*, 94(24):12740–12741, 1997.

Charles D Gilbert, Aniruddha Das, Minami Ito, Mitesh Kapadia, and Gerald Westheimer. Spatial integration and cortical dynamics. *Proceedings of the National Academy of Sciences*, 93(2):615–622, 1996.

Germán González, Engin Türetken, François Fleuret, and Pascal Fua. Delineating trees in noisy 2D images and 3D image-stacks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2799–2806, 2010.

Stephen Grossberg and Ennio Mingolla. Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychological review*, 92(2):173, 1985.

Umut Güçlü and Marcel AJ van Gerven. Unsupervised feature learning improves prediction of human brain activity in response to natural images. *PLoS Comput Biol*, 10(8):e1003724, 2014.

William E Hart, Michael Goldbaum, Brad Côté, Paul Kube, and Mark R Nelson. Measurement and classification of retinal vascular tortuosity. *International journal of medical informatics*, 53 (2):239–252, 1999.

Uri Hasson, Rafael Malach, and David J Heeger. Reliability of cortical activity during natural stimulation. *Trends in cognitive sciences*, 14(1):40–48, 2010.

Trevor J.. Hastie, Robert John Tibshirani, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction.* Springer, 2011.

James V Haxby, M Ida Gobbini, Maura L Furey, Alumit Ishai, Jennifer L Schouten, and Pietro Pietrini. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539):2425–2430, 2001.

John-Dylan Haynes and Geraint Rees. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature neuroscience*, 8(5):686–691, 2005.

John-Dylan Haynes and Geraint Rees. Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, 7(7):523–534, 2006.

Desmond J Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, 43(3):525–546, 2001.

William C Hoffman. The visual cortex is a contact bundle. *Applied Mathematics and Computation*, 32(2):137–167, 1989.

Qiao Hu, Michael D Abràmoff, and Mona K Garvin. Automated separation of binary overlapping trees in low-contrast color retinal images. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, pages 436–443. Springer, 2013.

Xintao Hu, Fan Deng, Kaiming Li, Tuo Zhang, Hanbo Chen, Xi Jiang, Jinglei Lv, Dajiang Zhu, Carlos Faraco, Degang Zhang, et al. Bridging low-level features and high-level semantics via fmri brain imaging for video classification. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 451–460. ACM, 2010.

Xintao Hu, Kaiming Li, Junwei Han, Xiansheng Hua, Lei Guo, and Tianming Liu. Bridging the semantic gap via functional brain imaging. *IEEE Transactions on Multimedia*, 14(2):314–325, 2012.

David H Hubel. *Eye, brain, and vision.* Scientific American Library/Scientific American Books, 1995.

David H Hubel and Torsten N Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106–154, 1962.

David H Hubel and Torsten N Wiesel. Ferrier lecture: Functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 198 (1130):1–59, 1977.

David S Jerison and Antonio Sánchez-Calle. Estimates for the heat kernel for a sum of squares of vector fields. *Indiana University mathematics journal*, 35(4):835–854, 1986.

Xiang Ji, Junwei Han, Xintao Hu, Kaiming Li, Fan Deng, Jun Fang, Lei Guo, and Tianming Liu. Retrieving video shots in semantic brain imaging space using manifold-ranking. In *2011 18th IEEE International Conference on Image Processing*, pages 3633–3636. IEEE, 2011.

Judson P Jones and Larry A Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of neurophysiology*, 58(6):1233–1258, 1987.

Vinayak S Joshi, Mona K Garvin, Joseph M Reinhardt, and Michael D Abramoff. Automated method for the identification and analysis of vascular tree structures in retinal vessel network. In *SPIE Medical Imaging*, pages 79630I–79630I. International Society for Optics and Photonics, 2011.

Angelos A Kalitzeos, Gregory YH Lip, and Rebekka Heitmar. Retinal vessel tortuosity measures and their applications. *Experimental eye research*, 106:40–46, 2013.

Yukiyasu Kamitani and Frank Tong. Decoding the visual and subjective contents of the human brain. *Nature neuroscience*, 8(5):679–685, 2005.

Eric R Kandel, James H Schwartz, Thomas M Jessell, Steven A Siegelbaum, and AJ Hudspeth. *Principles of neural science*, volume 4. McGraw-hill New York, 2000.

Gaetano Kanizsa. *Organization in vision: Essays on Gestalt perception.* Praeger Publishers, 1979.

Gaetano Kanizsa. *Grammatica del vedere: saggi su percezione e gestalt.* Il mulino, 1980.

Kendrick N Kay and Jack L Gallant. I can see what you see. *Nature neuroscience*, 12(3):245–245, 2009.

Kendrick N Kay, Thomas Naselaris, Ryan J Prenger, and Jack L Gallant. Identifying natural images from human brain activity. *Nature*, 452(7185):352–355, 2008.

Kendrick N Kay, Jonathan Winawer, Ariel Rokem, Aviv Mezer, and Brian A Wandell. A two-stage cascade model of bold responses in human visual cortex. *PLoS Comput Biol*, 9(5):e1003079, 2013.

Philip J Kellman and Thomas F Shipley. A theory of visual interpolation in object perception. *Cognitive psychology*, 23(2):141–221, 1991.

Christof Koch and Shimon Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. In *Matters of intelligence*, pages 115–141. Springer, 1987.

Jan J Koenderink. The structure of images. *Biological cybernetics*, 50(5):363–370, 1984.

Jan J Koenderink and Andrea J van Doorn. Representation of local geometry in the visual system. *Biological Cybernetics*, 55(6):367–375, 1987.

K Koflka. Principles of gestalt psychology. *New York: Har*, 1935.

W Kohler. Gestalt psychology (1929). *New York, NY: Liveright*, 1947.

Tai Sing Lee. Image representation using 2d gabor wavelets. *IEEE Transactions on pattern analysis and machine intelligence*, 18(10):959–971, 1996.

Tai Sing Lee. Dynamics of subjective contour formation in the early visual cortex. *Proceedings of the National Academy of Sciences*, 98(4):1907–1911, 2001.

Tianming Liu. A few thoughts on brain rois. *Brain imaging and behavior*, 5(3):189–202, 2011.

Jean Lorenceau and David Alais. Form constraints in motion binding. *Nature neuroscience*, 4(7): 745–751, 2001.

Marina Meila and Jianbo Shi. A random walks view of spectral segmentation. 2001.

Maurice Merleau-Ponty and Colin Smith. *Phenomenology of perception.* Motilal Banarsidass Publishe, 1996.

Tom M Mitchell, Svetlana V Shinkareva, Andrew Carlson, Kai-Min Chang, Vicente L Malave, Robert A Mason, and Marcel Adam Just. Predicting human brain activity associated with the meanings of nouns. *science*, 320(5880):1191–1195, 2008.

Yoichi Miyawaki, Hajime Uchida, Okito Yamashita, Masa-aki Sato, Yusuke Morito, Hiroki C Tanabe, Norihiro Sadato, and Yukiyasu Kamitani. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron*, 60(5):915–929, 2008.

Noemi Montobbio. Variational techniques in encoding fmri data for cortical architecture modeling. Master's thesis, 2016.

David Mumford. *Elastica and computer vision.* Springer, 1994.

Harihar Narasimha-Iyer, Vijay Mahadevan, James M Beach, and Badrinath Roysam. Improved detection of the central reflex in retinal vessels using a generalized dual-Gaussian model and robust hypothesis testing. *IEEE Transactions on Information Technology in Biomedicine*, 12(3): 406–410, 2008.

Thomas Naselaris, Ryan J Prenger, Kendrick N Kay, Michael Oliver, and Jack L Gallant. Bayesian reconstruction of natural images from human brain activity. *Neuron*, 63(6):902–915, 2009.

Thomas Naselaris, Kendrick N Kay, Shinji Nishimoto, and Jack L Gallant. Encoding and decoding in fmri. *Neuroimage*, 56(2):400–410, 2011.

Andrew Y Ng, Michael I Jordan, Yair Weiss, et al. On spectral clustering: Analysis and an algorithm. *Advances In Neural Information Processing Systems*, 2:849–856, 2002.

Shinji Nishimoto, An T Vu, Thomas Naselaris, Yuval Benjamini, Bin Yu, and Jack L Gallant. Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology*, 21(19):1641–1646, 2011.

Seiji Ogawa, Tso-Ming Lee, Asha S Nayak, and Paul Glynn. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnetic resonance in medicine*, 14(1):68–78, 1990.

Kenichi Ohki, Sooyoung Chung, Prakash Kara, Mark Hübener, Tobias Bonhoeffer, and R Clay Reid. Highly ordered arrangement of single neurons in orientation pinwheels. *Nature*, 442 (7105):925–928, 2006.

Bernt Oksendal. *Stochastic differential equations: an introduction with applications.* Springer Science & Business Media, 2013.

Cheryl A Olman, Pierre-Francois Van de Moortele, Jennifer F Schumacher, Joseph R Guy, Kâmil Uğurbil, and Essa Yacoub. Retinotopic mapping with spin echo bold at 7t. *Magnetic resonance imaging*, 28(9):1258–1269, 2010.

Martin Aastrup Olsen, Daniel Hartung, Christoph Busch, and Rasmus Larsen. Convolution approach for feature detection in topological skeletons obtained from vascular patterns. In *Computational Intelligence in Biometrics and Identity Management (CIBIM), 2011 IEEE Workshop on*, pages 163–167. IEEE, 2011.

Nobuyuki Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11 (285-296):23–27, 1975.

Spiro P Pantazatos, Ardesheer Talati, Paul Pavlidis, and Joy Hirsch. Decoding unattended fearful faces with whole-brain correlations: an approach to identify condition-dependent large-scale functional connectivity. *PLoS Comput Biol*, 8(3):e1002441, 2012.

Pierre Parent and Steven W Zucker. Trace inference, curvature consistency, and curve detection. *IEEE Transactions on pattern analysis and machine intelligence*, 11(8):823–839, 1989.

Fabian Pedregosa-Izquierdo. *Feature extraction and supervised learning on fMRI: from practice to theory.* PhD thesis, Université Pierre et Marie Curie-Paris VI, 2015.

Pietro Perona and William Freeman. A factorization approach to grouping. In *European Conference on Computer Vision*, pages 655–670. Springer, 1998.

Jean Petitot. The neurogeometry of pinwheels as a sub-riemannian contact structure. *Journal of Physiology-Paris*, 97(2):265–309, 2003.

Jean Petitot. *Neurogéométrie de la vision: modeles mathematiques et physiques des architectures fonctionnelles.* Editions Ecole Polytechnique, 2008.

Jean Petitot and Yannick Tondut. Vers une neurogéométrie. fibrations corticales, structures de contact et contours subjectifs modaux. *Mathématiques informatique et sciences humaines*, (145): 5–102, 1999.

Jonathan Pillow and Nava Rubin. Perceptual completion across the vertical meridian and the role of early visual cortex. *Neuron*, 33(5):805–813, 2002.

Stephen Lucian Polyak. The retina: the anatomy and the histology of the retina in man, ape, and monkey, including the consideration of visual functions, the history of physiological optics, and the histological laboratory technique. 1941.

Kelvin Poon, Ghassan Hamarneh, and Rafeef Abugharbieh. Live-vessel: Extending livewire for simultaneous extraction of optimal medial and boundary paths in vascular images. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007*, pages 444–451. Springer, 2007.

Francis KH Quek and Cemil Kirbas. Vessel extraction in medical images by wave-propagation and traceback. *IEEE Transactions on Medical Imaging*, 20(2):117–131, 2001.

Jonas Richiardi, Hamdi Eryilmaz, Sophie Schwartz, Patrik Vuilleumier, and Dimitri Van De Ville. Decoding brain states from fmri connectivity graphs. *Neuroimage*, 56(2):616–626, 2011.

Dario L Ringach. Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of neurophysiology*, 88(1):455–463, 2002.

Christian Robert and George Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.

Bart M Haar Romeny. *Front-end vision and multi-scale image analysis: multi-scale computer vision theory and applications, written in mathematica*, volume 27. Springer Science & Business Media, 2008.

Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

Gonzalo Sanguinetti. Invariant models of vision between phenomenology, image statistics and neurosciences. 2011.

Gonzalo Sanguinetti, Giovanna Citti, and Alessandro Sarti. Image completion using a diffusion driven mean curvature flow in a sub-riemannian space. In *Int. Conf. Comput. Vision Theory and Appl.(VISAPP'08)*, pages 22–25, 2008.

Gonzalo Sanguinetti, Giovanna Citti, and Alessandro Sarti. A model of natural image edge co-occurrence in the rototranslation group. *Journal of Vision*, 10(14):37, 2010.

ALESSANDRO Sarti and GIOVANNA Citti. On the origin and nature of neurogeometry. *La Nuova Critica*, 2011.

Alessandro Sarti and Giovanna Citti. The constitution of visual perceptual units in the functional architecture of v1. *Journal of computational neuroscience*, 38(2):285–300, 2015.

Alessandro Sarti, Giovanna Citti, and Jean Petitot. The symplectic structure of the primary visual cortex. *Biological Cybernetics*, 98(1):33–48, 2008.

Eric L Schwartz. Spatial mapping in the primate sensory projection: analytic structure and relevance to perception. *Biological cybernetics*, 25(4):181–194, 1977.

Eric L Schwartz. Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding. *Vision research*, 20(8):645–669, 1980.

Ladan Shams and Christoph Von Der Malsburg. The role of complex cells in object recognition. *Vision Research*, 42(22):2547–2554, 2002.

Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000.

Thomas F Shipley and Philip J Kellman. Perception of partly occluded objects and illusory figures: Evidence for an identity hypothesis. *Journal of Experimental Psychology: Human Perception and Performance*, 18(1):106, 1992.

Thomas F Shipley and Philip J Kellman. Spatiotemporal boundary formation: boundary, form, and motion perception from transformations of surface elements. *Journal of Experimental Psychology: General*, 123(1):3, 1994.

Joes Staal, Michael D Abràmoff, Meindert Niemeijer, Max A Viergever, and Bram van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4):501–509, 2004.

SA Talbot and WH Marshall. Physiological studies on neural mechanisms of visual localization and discrimination. *American Journal of Ophthalmology*, 24(11):1255–1264, 1941.

Bertrand Thirion, Edouard Duchesnay, Edward Hubbard, Jessica Dubois, Jean-Baptiste Poline, Denis Lebihan, and Stanislas Dehaene. Inverse retinotopy: inferring the visual content of images from brain activation patterns. *Neuroimage*, 33(4):1104–1116, 2006.

Bertrand Thirion, Philippe Pinel, Sébastien Mériaux, Alexis Roche, Stanislas Dehaene, and Jean-Baptiste Poline. Analysis of a large fmri cohort: Statistical and methodological issues for group analyses. *Neuroimage*, 35(1):105–120, 2007.

Karvel K Thornber and Lance R Williams. Analytic solution of stochastic completion fields. *Biological Cybernetics*, 75(2):141–151, 1996.

Keith R Thulborn, John C Waterton, Paul M Matthews, and George K Radda. Oxygenation dependence of the transverse relaxation time of water protons in whole blood at high field. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 714(2):265–270, 1982.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

Robert B Tootell, Eugene Switkes, Martin S Silverman, and Susan L Hamilton. Functional anatomy of macaque striate cortex. ii. retinotopic organization. *The Journal of Neuroscience*, 8(5):1531–1568, 1988.

Thomas Trappenberg. *Fundamentals of computational neuroscience*. OUP Oxford, 2009.

Daniel Y Ts'o, Charles D Gilbert, and Torsten N Wiesel. Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis. *The Journal of Neuroscience*, 6(4):1160–1170, 1986.

Markus Van Almsick, Remco Duits, Erik Franken, and Bart ter Haar Romeny. From stochastic completion fields to tensor voting. In *Deep Structure, Singularities, and Computer Vision*, pages 124–134. Springer, 2005.

R Von Der Heydt, F Heitger, and E Peterhans. Perception of occluding contours: Neural mechanisms and a computational model. *Biomedical research*, 14:1–6, 1993.

Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

Vincent Q Vu, Bin Yu, Thomas Naselaris, Kendrick Kay, Jack Gallant, and Pradeep K Ravikumar. Nonparametric sparse hierarchical models describe v1 fmri responses to natural images. In *Advances in Neural Information Processing Systems*, pages 1337–1344, 2009.

Vincent Q Vu, Pradeep Ravikumar, Thomas Naselaris, Kendrick N Kay, Jack L Gallant, and Bin Yu. Encoding and decoding v1 fmri responses to natural images with sparse nonparametric models. *The annals of applied statistics*, 5(2B):1159, 2011.

Johan Wagemans, James H Elder, Michael Kubovy, Stephen E Palmer, Mary A Peterson, Manish Singh, and Rüdiger von der Heydt. A century of gestalt psychology in visual perception: I. perceptual grouping and figure–ground organization. *Psychological bulletin*, 138(6):1172, 2012.

Dirk B Walther, Eamon Caddigan, Li Fei-Fei, and Diane M Beck. Natural scene categories revealed in distributed patterns of activity in the human brain. *The Journal of Neuroscience*, 29 (34):10573–10581, 2009.

Yael Weiss. Segmentation using eigenvectors: a unifying view. In *Computer vision, 1999. The proceedings of the Seventh IEEE International Conference on*, volume 2, pages 975–982. IEEE, 1999.

Max Wertheimer. Laws of organization in perceptual forms. 1938.

Lance R Williams and David W Jacobs. Local parallel computation of stochastic completion fields. *Neural computation*, 9(4):859–881, 1997a.

Lance R Williams and David W Jacobs. Stochastic completion fields: A neural model of illusory contour shape and salience. *Neural computation*, 9(4):837–858, 1997b.

Clare M Wilson, Kenneth D Cocker, Merrick J Moseley, Carl Paterson, Simon T Clay, William E Schulenburg, Monte D Mills, Anna L Ells, Kim H Parker, Graham E Quinn, et al. Computerized analysis of retinal vessel width and tortuosity in premature infants. *Investigative ophthalmology & visual science*, 49(8):3577–3585, 2008.

Robert H Wurtz and Eric R Kandel. Central visual pathways. *Principles of neural science*, 4: 523–545, 2000.

Xiayu Xu, Meindert Niemeijer, Qi Song, Milan Sonka, Mona K Garvin, Joseph M Reinhardt, and Michael D Abràmoff. Vessel boundary delineation on fundus images using graph-based approach. *IEEE Transactions on Medical Imaging*, 30(6):1184–1191, 2011.

Semir Zeki. *A Vision of the Brain.* Oxford Univ Press, 1993.

Lihi Zelnik-Manor and Pietro Perona. Self-tuning spectral clustering. 2005.

Jiong Zhang, Remco Duits, Gonzalo Sanguinetti, and Bart M ter Haar Romeny. Numerical approaches for linear left-invariant diffusions on se (2), their comparison to exact solutions, and their applications in retinal imaging. *Numerical Mathematics: Theory, Methods and Applications*, 9(01):1–50, 2016.

Dajiang Zhu, Kaiming Li, Carlos Cesar Faraco, Fan Deng, Degang Zhang, Lei Guo, L Stephen Miller, and Tianming Liu. Optimization of functional brain rois via maximization of consistency of structural connectivity profiles. *NeuroImage*, 59(2):1382–1393, 2012a.

Dajiang Zhu, Kaiming Li, Lei Guo, Xi Jiang, Tuo Zhang, Degang Zhang, Hanbo Chen, Fan Deng, Carlos Faraco, Changfeng Jin, et al. Dicccol: dense individualized and common connectivity-based cortical landmarks. *Cerebral cortex*, page bhs072, 2012b.

S Zucker. Differential geometry from the frenet point of view: boundary detection, stereo, texture and color. In *Handbook of mathematical models in computer vision*, pages 357–373. Springer, 2006.