



HAL
open science

Retina-Inspired Image Coding Schemes

Khaled Masmoudi

► **To cite this version:**

Khaled Masmoudi. Retina-Inspired Image Coding Schemes. Information Theory [cs.IT]. Université Nice Sophia Antipolis, 2012. English. NNT: . tel-00850287

HAL Id: tel-00850287

<https://theses.hal.science/tel-00850287>

Submitted on 12 Aug 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITY OF NICE - SOPHIA ANTIPOLIS
DOCTORAL SCHOOL STIC
SCIENCES ET TECHNOLOGIES DE L'INFORMATION
ET DE LA COMMUNICATION

PHD THESIS

to obtain the title of

Doctor of Science

of the University of Nice - Sophia Antipolis

Speciality : Signal and Image Processing

Defended by

Khaled MASMOUDI

Retina-Inspired Image Coding Schemes

prepared at the I3S laboratory

Sophia Antipolis, MEDIACODING Team

INTRODUCTION

The goal of my thesis work is to conceive novel image coders inspired from the retina. Indeed, while the issue of coding is important for energy and bandwidth saving, we are convinced that little is still to be gained if no shift is made in the philosophy underlying the conception of new image coders/decoders. So that, my work aims at laying the groundwork for the design of original image and video compression schemes that are based on models of biological visual systems. Here, we will focus on the retina as it is the organ responsible of the acquisition of the visual stimulus and the first coding device within the human visual system. The subject treated is very challenging and motivating for two main reasons. The first one is that the video compression techniques are now essential for most standard equipments such as HDTV and DVD, and that recent technical progress allows us to imagine more sophisticated coding schemes. The second one is that the recent discoveries in neurosciences about the human visual system could be a source of inspiration to propose new ideas, especially if we are able to better understand the neural code of the retina. Thus, we expect that an interdisciplinary approach can help achieve our goals. This approach would combine the signal processing techniques and the knowledge acquired by neurophysiologists. Hopefully, our effort will lead us towards novel coding algorithms that go beyond the standards.

The challenges underlying our work are numerous. First of all, we expect our effort to be at the origin of efficient bio-inspired image coders/decoders and a fortiori we expect it to be the basis of novel bio-inspired 2D+t and 3D coding algorithms. Another important issue is related to the numerous hypotheses concerning the retina coding schemes. Indeed, cracking the neural code of the retina is one crucial, though long term, goal. This issue is at the heart of several applications ranging from object recognition to retina prosthetics and brain machine interfaces.

The fundamental hypothesis behind my work is that the retina generates a code for the visual stimulus that has a suitable rate/quality trade-off. In fact, the retina transforms the visual stimulus into a brain-friendly signal: the retinal neural code. So that, the main novelty introduced here is to show how could the code generated by the retina be exploited in the context of still image compression.

In the first part of this manuscript, we will focus on the rank order coding. The rank order coding is one of several hypotheses about the coding features within the retinal neural code. The rank ordering is based on the hypothesis that the retina represents the visual stimulus by the order in which its cells are activated. We will introduce the rank order coding founding bases, formalize mathematically the concept of rank ordering, and detail a retina model specified in [Van Rullen 2001b] intended to support it. We will revisit this model by proposing an original and exact synthesis procedure for it. Indeed, the classical decoding procedure employed yields reconstruction errors that limit the model Rate/Quality performances when used as an image codec. Here we solve this problem in an original fashion by using the *frames theory*, where a frame of a vector space designates an extension for the notion of basis. Our contribution encompasses an (i) original mathematical demonstration that the filter bank used is a frame, (ii) an algorithm that guarantees an errorless reconstruction, and (iii) a novel out-of-core algorithm that computes the dual frame. Then, we will present a novel retina-inspired coding scheme for static images. We aim at using this coder to transmit the compressed visual information over low-bandwidth channels. Our novel coder/decoder combines the spiking retina model under study and well known data compression techniques. Then, we will compare our results to the JPEG standards and show that our coder/decoder has comparable performances with a simpler implementation under strong bandwidth restrictions. In particular, the coder/decoder that we devised has an interesting feature that we denote by the “time scalability”. Here, the time scalability feature designates the ability of tuning the rate and quality through the choice of the reconstruction time. We also study the case when the visual data is contaminated with noise and show that our coder acts as a coder coupled to a denoiser.

The second part of my thesis will lead this effort towards more biological plausibility. Though the first model considered takes into account some of the retinal visual processing mechanisms, several other processing stages are ignored. So that, we will rely on a retina simulation software called *Virtual Retina* [Wohrer 2009a] to enhance our first coder/decoder. We will design a novel bio-inspired image coder that keeps a strong plausibility with regard to the mammals retina behavior. In order to do this, we will first consider the deep retina layers and inspire ourselves from their behavior to design an original and dynamic A/D converter. Interestingly, the retina and especially its deepest layers operate a quantization process. The bio-inspired A/D converter that we will introduce offers several interesting features as the scalability, and the introduction of time dependency in the coding system. These features inset implicit bit-allocation. Then, we will model the retinal noise by a dither

signal, and thenceforth we identify the retina behavior to a non-subtractive dithered quantizer. This hypothesis gives a possible interpretation for the non-determinism observed in the spike-based neural code of the retina. When introduced in our A/D converter, the dithering process enables several interesting features. Indeed as we will show later in this chapter, the dithering whitens the reconstruction error and decorrelates it from the input stimuli. Finally, we will propose the design of an original scalable image coder/decoder that is highly inspired from the mammalians retina. Our coder accounts for the time-dependent and also non-deterministic behavior of the actual retina. We will introduce within our coder a multiscale dither signal in order to gain interesting perceptual features. The dithering process introduced here is an extension to what we presented in our previously specified A/D converter in the sense that it accounts for the multiscale nature of the transform. The dithering process that we propose adds several interesting features to our image coder. Integrating the multiscale dither noise in our coder/decoder allows a faster recognition of the fine details of the image during the decoding process.

INTRODUCTION

Mon travail de thèse vise à concevoir de nouveaux codeurs/décodeurs d'images inspirés de la rétine. En effet, la problématique de la compression est importante pour des questions d'économie d'énergie et de bande passante, notamment pour les équipements embarqués. Nous sommes convaincus qu'un changement est nécessaire dans la philosophie qui sous-tend la conception des codeurs/décodeurs d'images afin d'aboutir à des systèmes innovants et performants. Notre travail vise donc à poser les jalons de futurs systèmes originaux de compression d'images et vidéos qui soient basés sur des modèles de systèmes visuels biologiques. Ici, nous allons nous concentrer sur la rétine puisqu'elle est l'organe chargé de l'acquisition du stimulus visuel et le premier dispositif de codage dans le système visuel humain. Le sujet traité est extrêmement stimulant et motivant pour deux raisons principales. La première est que les systèmes de compression d'images et de vidéos sont aujourd'hui présents dans divers équipements standards tels que la TVHD, et que les progrès techniques récents nous permettent d'imaginer des schémas de codage plus élaborés qu'auparavant. La seconde est que les résultats établis par les neuroscientifiques ainsi que les récentes découvertes concernant le système visuel pourraient être une source d'inspiration pour proposer de nouvelles idées pour le codage d'images. Ainsi, nous pensons qu'une approche interdisciplinaire peut aider à atteindre nos objectifs. Cette approche devra combiner les techniques de traitement du signal et les connaissances acquises par les neurophysiologistes. Cet effort nous conduira vers l'élaboration de nouveaux algorithmes de codage qui vont au-delà des standards actuels.

Les défis sous-tendant notre travail sont nombreux. Tout d'abord, nous nous attendons à ce que notre effort soit à l'origine de codeurs/décodeurs d'images bio-inspirés efficaces et a fortiori nous nous attendons à ce que ce travail soit à la base d'algorithmes de codage bio-inspirés 2D + t et 3D originaux. Une autre question importante concerne les nombreuses hypothèses quant au schémas de codage adoptés par la rétine. En effet, le décryptage du code neural de la rétine est un objectif crucial, quoi que de long terme. Cette question est au cœur de plusieurs applications allant de la reconnaissance d'objets aux prothèses de rétine et autres interfaces cerveau-machine.

L'hypothèse fondamentale de mon travail est que la rétine représente le stimulus visuel sous la forme d'un code ayant un rapport débit/distorsion satisfaisant pour la compression d'images. En effet, la rétine transforme le

stimulus visuel en un signal manipulable par les zones corticales du cerveau consacrées à la vision. Ce signal est le code neuronal de la rétine. La principale nouveauté introduite ici est de montrer comment le code généré par la rétine peut être exploité dans le contexte de la compression d'images.

Dans la première partie de ce manuscrit, nous considérerons le codage d'images par rang. Le codage par rang est une hypothèse parmi plusieurs autres concernant les caractéristiques supposées être significatives au sein du code neuronal de la rétine. Le codage par rang se base sur l'hypothèse selon laquelle la rétine représente le stimulus visuel par l'ordre dans lequel ses cellules sont activées. Nous allons introduire les fondements du codage par rang, nous les formaliserons mathématiquement, et détaillerons un modèle de rétine spécifié dans [Van Rullen 2001b] destiné à soutenir cette hypothèse. Nous revisiterons ce modèle en proposant une procédure de synthèse originale et garantissant une reconstruction exacte. En effet, la procédure classique de décodage employée donne des erreurs de reconstruction qui limitent les performances du modèle en terme de rapport débit/distorsion lorsqu'il est utilisé comme un codeur/décodeur d'images. Ici, nous résolvons ce problème d'une manière originale en utilisant la théorie des *frames*, où une frame d'un espace vectoriel désigne une extension de la notion de base. Ma contribution englobe (i) une démonstration mathématique que le banc de filtres utilisé est une frame, (ii) un algorithme qui garantit une reconstruction sans erreur, et (iii) un nouvel algorithme out-of-core qui calcule la frame duale. Ensuite, nous présenterons un nouveau schéma de codage/décodage inspiré de la rétine. Ce nouveau codeur/décodeur est adapté à la transmission de données à travers des canaux de faible bande passante. Notre codeur/décodeur combine le modèle rétine étudié et des techniques éprouvées de compression de données. Ensuite, nous comparerons les résultats à ceux des normes JPEG et montrerons que notre codeur/décodeur a des performances comparables avec une implémentation plus simple sous une forte contrainte de bande passante. En particulier, le codeur/décodeur que nous avons imaginé possède une caractéristique intéressante que nous désignerons par "scalabilité temporelle". Dans notre cas, la scalabilité temporelle désigne la capacité d'ajuster la qualité et le coût de l'image à travers le choix du temps accordé à la reconstruction. Nous étudierons également le cas où les données sont contaminées par du bruit et nous montrerons que notre codeur agit comme un codeur couplé à un débruiteur.

La deuxième partie de ma thèse nous conduira vers l'élaboration d'un codeur/décodeur ayant plus de plausibilité biologique. En effet, le premier modèle de rétine considéré prend en compte plusieurs mécanismes impliqués

dans le traitement rétinien du stimulus. Cela étant, plusieurs autres mécanismes sont ignorés. Nous nous appuyons donc sur le logiciel de simulation de rétine appelée *Virtual Retina* [Wohrer 2009a] pour améliorer notre premier codeur/décodeur. Afin de concevoir notre nouveau codeur d'images bio-inspiré, nous allons d'abord considérer les couches profondes et la rétine et nous inspirer de leur comportement pour spécifier un convertisseur A/D original et dynamique. En effet, la rétine et en particulier ses couches les plus profondes opèrent une quantification. Le convertisseur A/D bio-inspiré que nous introduiront offre plusieurs fonctionnalités intéressantes comme la scalabilité et l'incorporation d'une évolutivité temporelle dans notre système de codage. Ces caractéristiques permettent un mécanisme d'allocation de débit implicite. Ensuite, nous modéliserons le bruit observé dans les signaux de la rétine par un bruit *dither*. Nous identifierons donc le comportement de la rétine à celui d'un quantificateur avec dither non-soustractif. Cette hypothèse donne une interprétation possible pour le non-déterminisme observé dans le code neuronal rétinien. Lorsqu'il est introduit dans notre convertisseur A/D, le processus de dithering lui adjoint plusieurs fonctionnalités intéressantes. En effet, comme nous le montrerons, le dithering rend blanc le spectre de l'erreur de reconstruction et il le décorrèle des stimuli d'entrée. Enfin, nous proposerons un design pour un codeur/décodeur d'images scalable qui est fortement inspiré de la rétine des mammifères. Notre codeur/décodeur a un comportement dépendant du temps et qui prend en compte le non-déterminisme des processus de codage rétinien. Nous incorporons au sein de notre codeur/décodeur un processus de dithering multi-échelle original. Le processus de dithering que nous proposons ajoute des fonctionnalités intéressantes à notre codeur d'images. En particulier, intégrer le bruit dither multi-échelle dans notre codeur/décodeur (i) décorrèle les coefficients d'analyse de l'erreur de reconstruction et (ii) permet une reconnaissance plus rapide des détails fins de l'image pendant le processus de décodage.

Organization

My Thesis consists of an introductory chapter followed by two parts of three and two chapters. The manuscript is organized as follows:

- *) In Chapter 1, we overview the mammalians visual system from a “work-flow” point of view. We describe the main cortical areas crossed by the information within the two visual pathways: the ventral one and the dorsal one. We also introduce fundamental concepts about the neural code generated by the retina and the different hypotheses about the coding features within.
- 1. Part I is dedicated to the conception of a bio-inspired image coder starting from the rank order coding hypothesis. We especially consider the retina model defined in [Van Rullen 2001b] as a basis for our novel coder/decoder and make the adequate modifications and enhancements to adapt it to the coding application.
 - 1) In Chapter 2, we detail our main working hypothesis for the first part of this work: the rank order coding. We describe its guiding principles and report the neurophysiologic results supporting it. We also give a new mathematical formalism that rigorously defines the rank order code. We then recall the specification of a simplified model of the retina in [Van Rullen 2001b] that is intended to support the rank order coding.
 - 2) In Chapter 3, we revisit the retina model specified in [Van Rullen 2001b] by proposing an original and exact synthesis procedure for it. The solution that we propose relies on the frames theory. Our contribution encompasses a theoretical and a technical aspect. We published these results in [Masmoudi 2012b, Masmoudi 2012a]
 - 3) In Chapter 4, we present an original retina-inspired coding scheme for static images. Our coder has three stages that combine the spiking retina model presented in Chapter 2 and data compression techniques from the literature. We detail the specifications of our new coder stages, then we compare its performances to state of the art coders. This chapter is based mainly on the results that we published in [Masmoudi 2010d].

2. Part II is based also on the model specified in [Van Rullen 2001b]. We enhance it through the introduction of time dynamics in the coding and decoding processes. We also introduce several other mechanisms involved in the retina behavior and that are reproduced in the retina simulation software called *Virtual Retina* [Wohrer 2009a]. The bio-inspired image coder/decoder that we present in this part mimics as faithfully as possible the retina behavior.
 - 1) In Chapter 5, we explore the behavior of the inner layers of the mammalian retina. We take the *Virtual Retina* simulation software as a basis for our study. In a first step, we design a dynamic quantization scheme that relies on a rate code. This scheme encompasses three stages mimicking the inner layers of the retina. In a second step, we hypothesize that the noise observed in the retina is a dither signal. We then overview a few interesting features of our retina-inspired analog-to-digital (A/D) converter. This chapter is based mainly on the results that we published in [Masmoudi 2010a, Masmoudi 2010b, Masmoudi 2010c].
 - 2) In Chapter 6, we propose the design of an original scalable image coder/decoder that is inspired from the mammalian retina. Our coder accounts for the time-dependent and also non-deterministic behavior of the actual retina. This chapter is based mainly on the results that we published in [Masmoudi 2012c] and [Masmoudi 2011].

Contents

| | | |
|----------|--|-----------|
| 1 | Basics about the image compression and the HVS | 1 |
| 1.1 | Basics about image compression | 4 |
| 1.1.1 | Generalities | 4 |
| 1.1.2 | Pyramid image transforms | 6 |
| 1.1.3 | Rate-distortion theory | 8 |
| 1.1.4 | Typical image coders design <i>vs</i> the retina | 8 |
| 1.2 | The visual pathways | 10 |
| 1.2.1 | The primal visual pathway | 11 |
| 1.2.2 | The dorsal and ventral streams | 17 |
| 1.2.3 | The feedforward and feedback connections | 20 |
| 1.3 | A few fundamental notions | 23 |
| 1.3.1 | The retinal neural code shape | 23 |
| 1.3.2 | Possible coding metrics in the retina | 25 |
| 1.4 | Discussion | 30 |
| I | From the ROC to a Bio-Inspired Scalable Codec | 33 |
| 2 | The Rank Order Code | 35 |
| 2.1 | The rank order coding | 38 |
| 2.1.1 | The rank order coding principles | 38 |
| 2.1.2 | The experimental evidence behind | 38 |
| 2.2 | A mathematical formalism for the rank order coding | 42 |
| 2.3 | A retina model based on rank order coding | 43 |
| 2.3.1 | The retina cells: model and architecture | 43 |
| 2.3.2 | Image analysis | 47 |
| 2.3.3 | Image synthesis | 50 |
| 2.4 | Discussion | 51 |
| 3 | Frames to Invert ROC | 53 |
| 3.1 | The ROC limitations | 56 |
| 3.1.1 | Matrix notations for ROC | 56 |
| 3.1.2 | Limitations and straightforward solution | 60 |
| 3.2 | Inverting the retina model | 63 |
| 3.2.1 | Introduction of a low-pass scaling function | 65 |
| 3.2.2 | The DoG transform is a frame operator | 66 |
| 3.2.3 | Synthesis using the dual DoG frame | 69 |

| | | |
|-----------|---|------------|
| 3.3 | Results | 74 |
| 3.4 | Discussion | 75 |
| 4 | A novel bio-inspired coding scheme | 81 |
| 4.1 | The coding pathway | 85 |
| 4.1.1 | System overview | 85 |
| 4.1.2 | The image transform in the retina model | 85 |
| 4.1.3 | Spikes coding using the stack-run algorithm | 88 |
| 4.1.4 | Arithmetic coding | 92 |
| 4.2 | Results | 96 |
| 4.2.1 | Quality metrics | 96 |
| 4.2.2 | Comparison to JPEG standards | 97 |
| 4.3 | Discussion | 98 |
| II | From <i>Virtual Retina</i> to a Bio-Inspired Codec | 101 |
| 5 | A Retina-Inspired A/D converter | 105 |
| 5.1 | <i>Virtual Retina</i> : A bio-plausible model | 109 |
| 5.1.1 | The outer layers | 110 |
| 5.1.2 | The inner layers | 112 |
| 5.1.3 | The ganglionic layer | 113 |
| 5.2 | A retina-inspired A/D converter | 114 |
| 5.2.1 | Study context | 114 |
| 5.2.2 | The coding pathway | 117 |
| 5.2.3 | The decoding pathway | 122 |
| 5.3 | Our A/D converter <i>vs</i> standards | 124 |
| 5.3.1 | Our A/D converter overall behavior | 124 |
| 5.3.2 | A beyond the standards A/D converter | 126 |
| 5.4 | The dithering hypothesis | 130 |
| 5.5 | Discussion | 135 |
| 6 | Streaming an image through the retina | 137 |
| 6.1 | The coding pathway | 140 |
| 6.1.1 | The image transform: The outer layers of the retina | 140 |
| 6.1.2 | The A/D converter | 143 |
| 6.2 | The decoding pathway | 146 |
| 6.3 | Results: The noiseless case | 148 |
| 6.4 | Introducing the noise in the coder | 151 |
| 6.4.1 | Study context | 151 |
| 6.4.2 | Introducing the multiscale NSD | 152 |
| 6.5 | Results: The dithered case | 156 |

| | |
|---------------------------------------|-------------|
| Contents | xiii |
| <hr/> | |
| 6.6 Conclusion | 160 |
| 7 Discussion and perspectives | 163 |
| A Out-of-core matrix inversion | 167 |
| Bibliography | 173 |

List of Figures

| | | |
|------|--|----|
| 1.1 | Example of a progressive transmission/reconstruction with three types of pyramid transforms | 7 |
| 1.2 | The analytical rate-distortion function in the case of a memoryless independent Gaussian source. | 9 |
| 1.3 | Possible analogy between a classical image coder and the retina. | 11 |
| 1.4 | The visual Pathway | 12 |
| 1.5 | OCT scan and schematic view of the retina and its cells | 13 |
| 1.6 | Diversity of retinal cell types | 15 |
| 1.7 | Center-surround receptive field structure of an OFF-center midget bipolar cell | 16 |
| 1.8 | The LGN layers | 17 |
| 1.9 | The ventral and dorsal streams | 18 |
| 1.10 | The feedforward and feedback connections in the ventral stream | 21 |
| 1.11 | A spike, a spike train, and a raster plot | 24 |
| 1.12 | Some possible relevant measures in the spiking code | 25 |
| 1.13 | Trial averaging for rate coding | 27 |
| 1.14 | Population averaging for rate coding | 28 |
| | | |
| 2.1 | The first wave of spikes | 39 |
| 2.2 | Invariance of ROC to stimuli contrast and dynamic range . . . | 41 |
| 2.3 | Behavior of the DoG filter | 45 |
| 2.4 | DoG filters spread over a dyadic grid | 46 |
| 2.5 | Spectra of the DoG filters | 47 |
| 2.6 | Examples of ROC look-up-tables | 49 |
| 2.7 | Progressive stimulus synthesis using the ROC | 52 |
| | | |
| 3.1 | The image transform through the retina model | 58 |
| 3.2 | Templates of the analysis and synthesis matrices | 59 |
| 3.3 | Histograms comparison between the original retina model and the corrected one | 63 |
| 3.4 | Error image with the original retina model decoder | 64 |
| 3.5 | Spectra of the DoG filters augmented with a scaling function . | 66 |
| 3.6 | Error image with the dual DoG frame decoder | 71 |
| 3.7 | Comparison between the DoG filters and their duals | 72 |
| 3.8 | <i>Lena</i> reconstruction using the dual DoG filters | 77 |
| 3.9 | Cameraman reconstruction using the dual DoG filters | 78 |

| | | |
|------|--|-----|
| 3.10 | The reconstruction PSNR comparison: primal Vs. dual DoG filters | 79 |
| 4.1 | Block diagram of the compression scheme | 86 |
| 4.2 | The look-up-table estimated for Lena | 87 |
| 4.3 | Histograms of the two stack run code files | 92 |
| 4.4 | Block diagram of the stack run coder coupled to a context based arithmetic coding | 92 |
| 4.5 | An example of arithmetic coding | 94 |
| 4.6 | Bloc diagram of a context based adaptive arithmetic coder | 95 |
| 4.7 | Comparison between JPEG, JPEG2000 and our new codec | 97 |
| 4.8 | Robustness to noise | 98 |
| 4.9 | Rate/Quality behavior in the case of noisy data | 99 |
| 5.1 | Schematic view of <i>Virtual Retina</i> | 111 |
| 5.2 | Input/output map of the INL layer | 119 |
| 5.3 | Input/output map of the IPL layer cascaded to the INL layer | 120 |
| 5.4 | The time behavior of the ganglion cell voltage | 121 |
| 5.5 | Behavior evolution of our retina-inspired A/D converter | 122 |
| 5.6 | Time evolution of the decoding map | 123 |
| 5.7 | Time-evolution of the characteristic function | 125 |
| 5.8 | The analogy between the retina and the standard A/D converters | 126 |
| 5.9 | Behavior evolution of the INL stage | 127 |
| 5.10 | Behavior evolution of the INL/IPL stages | 128 |
| 5.11 | A comparison between the INL/IPL stages and the standard gain amplification/companing algorithms. | 129 |
| 5.12 | A comparison between our retina-inspired A/D converter and the standard Lloyd-Max quantizer | 130 |
| 5.13 | The LIF neuron seen as an uniform scalar quantizer | 132 |
| 5.14 | Cross correlation of the quantization error and the input stimuli. | 134 |
| 5.15 | Comparison of the reconstruction visual quality between non-dithered and dithered quantizing ganglion cell | 134 |
| 6.1 | Overview of our bio-inspired codec | 141 |
| 6.2 | Example on Lena of the DoG coefficients generated by the retina model | 143 |
| 6.3 | Time delays D_{t_k} introduced in the coding process | 144 |
| 6.4 | The inner layers behavior | 145 |
| 6.5 | Progressive image reconstruction of cameraman | 149 |
| 6.6 | Illustration of the concept of time scalability | 150 |
| 6.7 | Estimation of the LIF neuron quantization step | 154 |

| | | |
|------|--|-----|
| 6.8 | Example of dither noise introduced at the input of the ganglionic layer | 155 |
| 6.9 | Error whitening and decorrelation in the DoG transform domain induced by the dither noise addition | 157 |
| 6.10 | Perceptual impact of the multiscale dithering on the reconstruction of cameraman | 159 |
| 6.11 | Perceptual impact of the multiscale dithering on the reconstruction of baboon | 160 |

BASICS ABOUT THE IMAGE COMPRESSION AND THE HUMAN VISUAL SYSTEM

Contents

| | | |
|------------|--|-----------|
| 1.1 | Basics about image compression | 4 |
| 1.1.1 | Generalities | 4 |
| 1.1.2 | Pyramid image transforms | 6 |
| 1.1.3 | Rate-distortion theory | 8 |
| 1.1.4 | Typical image coders design <i>vs</i> the retina | 8 |
| 1.2 | The visual pathways | 10 |
| 1.2.1 | The primal visual pathway | 11 |
| 1.2.2 | The dorsal and ventral streams | 17 |
| 1.2.3 | The feedforward and feedback connections | 20 |
| 1.3 | A few fundamental notions | 23 |
| 1.3.1 | The retinal neural code shape | 23 |
| 1.3.2 | Possible coding metrics in the retina | 25 |
| 1.4 | Discussion | 30 |

Keywords

Visual pathway, image compression, two streams hypothesis, spikes, spike trains, raster plots, neural code, rate codes, timing codes

Overview

In this chapter, we will first introduce briefly the fundamental concepts of image coding. In particular, we will raise the analogy between the classical image coding systems and the stimulus coding within the human visual system (HVS). Then, we will overview the HVS from a “workflow” point of view. For this sake, we will describe the main processing stages crossed by the visual code within the different visual pathways, from the retina to the deeper cortical areas. Finally, we will introduce the reader to some fundamental concepts about the neural code generated by the retina and its different interpretation hypotheses.

Organization

This chapter is organized into four sections:

1. Section 1.1 introduces the fundamental concepts of image coding.
2. Section 1.2 describes the visual processing pathways that the visual data follows starting from the retina based on the two streams hypothesis.
3. Section 1.3 presents some fundamental concepts about the neural code of the retina such as spikes and spike trains. We especially focus on the main spike-based coding strategies in the retina and we distinguish rate and timing codes.
4. Section 1.4 summarizes the important concepts about the visual system that we tackled. Starting from this point, we discuss the different evoked retina coding strategies and justify our working hypotheses.

*While the mind is in doubt it is driven
this way and that by a slight impulse
Terence*

During the past three decades, research in lossy compression for still images yielded several coding algorithms especially the JPEG standards. Since then, subsequent efforts for conceiving lossy image coders followed almost the same schema [Antonini 1992, Christopoulos 2000, Skodras 2001]. These coding algorithms were mostly designed in a signal processing way of thinking and do not account for the actual biological visual systems behavior. Yet, computational neuroscience made substantial progress during the same period in better understanding the internal representation of the sensory world. In particular, concerning the visual stimuli sensing, one can find many results and heuristics on how information is encoded, transmitted, and interpreted within the mammals visual system. Based on those results, it is our conviction that the mammals visual system developed efficient coding strategies that could be used as a source of inspiration to imagine novel compression algorithms.

Our goal in this work is to conceive novel image coders inspired from the existing biological systems which role is to encode the visual stimuli. To this end, we focus our study on the retina as it is the first coding device encountered by a visual stimulus within the visual chain of treatments. So that, we will begin this introductory chapter by recalling the fundamental concepts about image coding and compression that will be used in the rest of this work [Rabbani 1991, Wallace 1992, Antonini 1992, Skodras 2001]. Then, we will overview the behavior of the retina and the HVS as a whole and describe the functional role of its different cortical areas. For this sake, we will base ourselves on the so-called “two streams hypothesis” [Goodale 1992].

Yet, the retina transforms the visual stimulus into a brain-friendly signal: the *neural code* [Rieke 1997]. This neural code follows thenceforth two visual pathways across the cortex to be processed and analyzed, namely the dorsal and the ventral streams. So that, in this chapter we will briefly introduce the reader to these visual pathways and the processing stages that they involve. We then bring into focus the neural code of the retina it-

self. We tackle the fundamental concepts that will be necessary for the comprehension of the rest of this work. Especially, we detail the notion of *spikes* which are the electrical entities that convey the visual data in the cortex. As no clear evidence is established about the relevant metrics to be measured over the spikes, numerous hypotheses are discussed in the literature [Bialek 1991, Mainen 1995, VanRullen 2005, Johansson 2004]. Thus, we will make an overview of the main candidate coding strategies based on spikes that might be employed by the retina.

This chapter is organized into four sections. First, in Section 1.1, we introduce the reader briefly to the fundamental concepts of image coding and raise the possible analogy to be made with the retina functioning. Second, in Section 1.2, we describe the pathways that the visual data follows starting from the retina and the processing stages within. Third, in Section 1.3, we present some fundamental concepts about the neural code of the retina such as spikes and *raster plots*. We will also make an overview of the main retina coding strategies that are potentially relevant with regard to the mass of data conveyed by spikes. Finally, in Section 1.4, we discuss the different notions addressed and especially the different possible retinal coding strategies. This discussion will inset our main working hypothesis which is the rank order coding strategy [Thorpe 1990] to be detailed in the next chapter.

1.1 Basics about image compression

We will overview in this chapter fundamental notions about the image coding theme. First, we will introduce the image transforms and in particular the pyramid transforms, then we will recall some elements about the rate-distortion theory, and finally we will evoke the possible analogy to be made between classical image coders with the retina behavior.

1.1.1 Generalities

A characteristic that is common to most images is that their neighboring pixels are correlated. Thus images contain redundant information. The goal of a coding system is then to find the less correlated representation of the image. Two fundamental issues related to compression are the reduction of

redundancy and the reduction of irrelevancy. Reducing redundancy amounts to remove duplication from the source signal. Reducing irrelevancy aims at omitting parts of the signal that will not be distinguished by the signal receiver, namely the HVS. In general, two types of redundancy can be identified:

1. Spatial Redundancy or correlation between neighboring pixel values.
2. Spectral Redundancy or correlation between different color planes or spectral bands.

The goal of image compression algorithms is to reduce the number of bits needed to represent an image by removing as much spatial and spectral redundancies as possible.

Lossless vs. Lossy compression: In lossless compression schemes, the image that is synthesized, starting from its compressed version, is strictly identical to the original image. However lossless compression can only achieve a modest amount of compression. On the contrary, an image reconstructed following lossy compression is altered compared to the original one. Often, this is caused by the fact that lossy compression scheme discards redundant information. However, lossy schemes are able to achieve much higher compression ratios. Under normal viewing conditions, barely visible loss is perceived (visually lossless).

Predictive vs. Transform coding: In predictive coding, information already sent or available is used to predict future values, and the difference is coded. Since this is done in the image or spatial domain, it is relatively simple to implement and is readily adapted to local image characteristics. Differential Pulse Code Modulation (DPCM) is one particular example of predictive coding. Transform coding, on the other hand, first transforms the image from its spatial domain representation to a different type of representation using some well-known transform and then codes the transformed values (coefficients). This method provides greater data compression compared to predictive methods, although at the expense of greater computation. We will introduce the image transforms and especially the pyramid transforms in the next section.

1.1.2 Pyramid image transforms

In most of the coding/decoding systems, source signals are encoded in a transform domain. This transform is operated because it generally enables a suitable representation of the signal by means of few coefficients. For the sake of compression, the fewer are the coefficients needed to represent the source signal, the better is the transform. Over the years, a variety of linear transforms have been developed which include Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT) [Ahmed 1974], Discrete Wavelet Transform (DWT) [Daubechies 1990, Antonini 1992, Vetterli 1995] and many more, each with its own advantages and disadvantages. In this introduction, we will consider another type of transforms, namely the pyramid transforms, as we will be thoroughly using them in this work.

An image can be represented as a pyramid structure via a pyramid transform. The pyramid structure obtained can be described as a collection of images at different scales that together represent the original image. Among the pyramid transforms specified in the literature, two main efforts can be cited. The first one is specified in [Burt 1983]. The authors coded images using the Gaussian and Laplacian pyramids. The basis functions are low-pass kernels repeated at a series of positions, and appear at scales varying by factors of two. The pyramid transform gives an exact reconstruction, though the basis set is not orthogonal, and the number of transform coefficients exceeds the number of original pixels by a factor of $\frac{4}{3}$. The second transform is specified in [Watson 1987]. The authors has developed an oriented pyramid transform. The basis functions of this transform resemble Gabor functions and their spatial frequencies and orientations are similar to that inferred in the HVS, in a cortical area named V1. Again, the nonorthogonal basis set is overcomplete by $\frac{4}{3}$. The reconstruction is nearly exact. The pyramid transforms have two main advantages:

- Localization in space and spatial frequency: One way to achieve localization in both space and spatial frequency is to compute a DCT in a small blocks. But the linear and arbitrary edges of these blocks introduce discontinuities in the image and alter the frequency localization. In the same way as the wavelet transforms, pyramid transforms guarantee a good frequency resolution while keeping reasonable spatial resolution.



Figure 1.1: Example of a progressive transmission/reconstruction with three types of pyramid transforms. Each row shows the results for a given pyramid transform type (from [Adelson 1987]).

This is achieved by means of the multiscale representation of a pyramid structure [Adelson 1987].

- Progressive transmission: In several applications it is desirable that the decoder synthesizes a low resolution version of the coded image rapidly, and then that higher resolutions are added as time goes on. Many image coding techniques can be modified to allow such a progressive transmission. But in the case of pyramids, which are inherently multiscale representations, progressive transmission is achieved simply by sending information from successive pyramid levels in sequence. This adds no informational overhead. This property will be used in the rest of this work especially when studying rank order coding. An example of progressive transmission/reconstruction is shown in Figure 1.1.

In order to compare objectively these different transforms and their ability to improve an image coding/decoding scheme, an evaluation framework is

needed. The rate-distortion theory is a discipline providing information theory tools to accomplish this task. We introduce some elements of this theory in the next section.

1.1.3 Rate-distortion theory

Rate-distortion theory is a theoretical discipline which goal is to tell what is the maximum fidelity that a coder/decoder can achieve at a given bit-rate? This issue amounts also to find what is the minimum rate required to obtain a given distortion? Here, a distortion measure is generally a distance or a divergence between a random variable and its representation after coding/decoding. The rate is usually understood as the number of bits per data sample to be stored or transmitted. These two measures define the so-called rate-distortion function.

A usual study case is the memoryless independent Gaussian source of σ_x standard deviation. Considering the distortion D as the expected value of the square of the difference between input and output signal (i.e., the mean squared error), an analytical expression for the rate-distortion function can be found [Cover 1991]. This function is shown in Figure 1.2. The rate-distortion theory states that no compression system exists that performs outside the gray area. The closer a given compression system is to the red bound (cf. Figure 1.2), the better it performs.

While the mean squared error (MSE) remains widely used, it is to be noted that the notion of distortion is still a hot issue of discussion among the image and video compression community [Nauge 2010]. In image and video compression, the human perception models are not well developed and their inclusion is mostly limited to weighting traditional measure as the MSE.

Once we made an overview of image coding and the rate distortion theory, we will evoke, in the next section, the possible analogy to be made between classical image coders with the retina behavior.

1.1.4 Typical image coders design *vs* the retina

A typical lossy image compression system consists of three closely connected components namely:

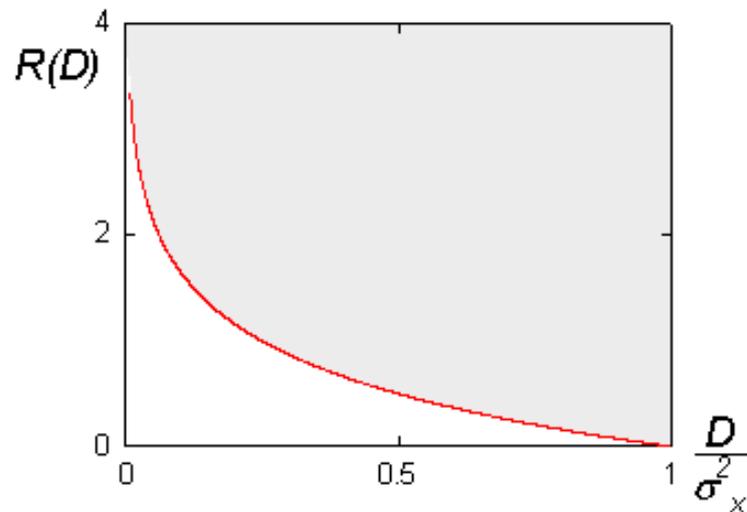


Figure 1.2: The analytical rate-distortion function in the case of a memoryless independent Gaussian source.

1. Source Encoder
2. Quantizer
3. Entropy Encoder

Compression is accomplished by applying a linear transform to decorrelate the image data, quantizing the resulting transform coefficients, and entropy coding the quantized values.

As discussed earlier, a variety of linear transforms have been developed as DFT, DCT, DWT, or the pyramid transforms. A quantizer simply reduces the number of bits needed to store the transformed coefficients by reducing the precision of those values. Since this is a many-to-one mapping, it is a lossy process and is the main source of compression in an encoder. Quantization can be performed on each individual coefficient, which is known as Scalar Quantization (SQ). Quantization can also be performed on a group of coefficients together, and this is known as Vector Quantization (VQ). Both uniform and non-uniform quantizers can be used depending on the problem at hand. For an analysis on different quantization schemes, see [Gersho 1992].

An entropy encoder further compresses the quantized values losslessly to give better overall compression. It uses a model to accurately determine the

probabilities for each quantized value and produces an appropriate code based on these probabilities so that the resultant output code stream will be smaller than the input stream. The most commonly used entropy encoders are the Huffman encoder and the arithmetic encoder, although for applications requiring fast execution, simple run-length encoding (RLE) has proven very effective. An overview on various entropy encoding techniques can be found in [Gersho 1992, Nelson 1996]. It is important to note that a properly designed quantizer and entropy encoder are absolutely necessary along with optimum signal transformation to get the best possible compression.

Figure 1.3 shows a block diagram of an image coder and raises the analogy with the retina functioning. Indeed, the retina operates a transform on the continuous input signal, before transmitting a quantized series of spikes to the brain. Besides, the outermost photosensing cells of the retina are one hundred times more numerous than the innermost cells that are wired of the optic nerve. Thus the retina might be compressing the source signal. As the HVS is at the center of our investigations, we will overview in the next section, the cortical areas that the neural code will get through once transmitted by the retina.

1.2 The visual processing pathways:

The two streams and the FF/FB connections within

The main goal of our work is to design an image coder that is inspired from the retina. In this context, our comprehension of the neural code of the retina requires to have the adequate knowledge of (i) the succession of cortical areas that the neural code will get through once transmitted by the retina, and (ii) the usage that these considered areas are supposed to make of it. This is a challenging task since the visual system is the largest in volume among all sensing systems. So that, we will overview the main cortical areas involved in the visual chain of treatments following the retinal coding and the resulting decisions to be undertaken by the cortex. According to the nature of these decisions, the visual neural code follows one of two pathways: the dorsal stream or the ventral one. Within each pathway the different cortical areas are interconnected in two directions: a feedforward (also called FF or top-

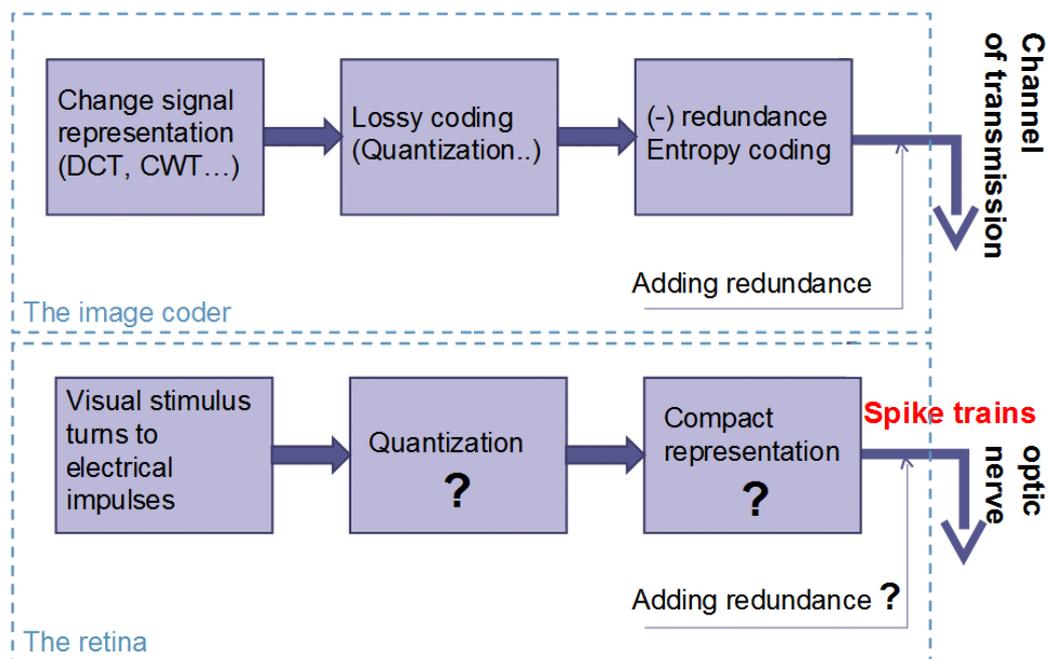


Figure 1.3: Possible analogy between a classical image coder and the retina.

down) direction then a feedback (also called FB or bottom-up) one.

In this section, we first describe in Section 1.2.1 the track that is common to all visual stimuli from the retina to the first cortical relay, the LGN. Then in Section 1.2.2, we describe the visual flux bifurcation into a ventral and a dorsal stream and the role of each of them. Finally, we describe the feedforward and feedback interconnections that exist within each stream in Section 1.2.3.

1.2.1 From the retina to the LGN

The image forming data follows a single track from the retina to the first cortical relay before bifurcating. We overview in this section the two main stages encountered by the stimulus in this primal pathway: (i) the retina and (ii) the Lateral Geniculate Nucleus (LGN). From the LGN through the primary visual cortex (V1) to higher order areas, the visual stream is divided in two as will be detailed in the next section. Figure 1.4 schematizes the connections of the retina with the cited cortical areas by means of the optic nerve.

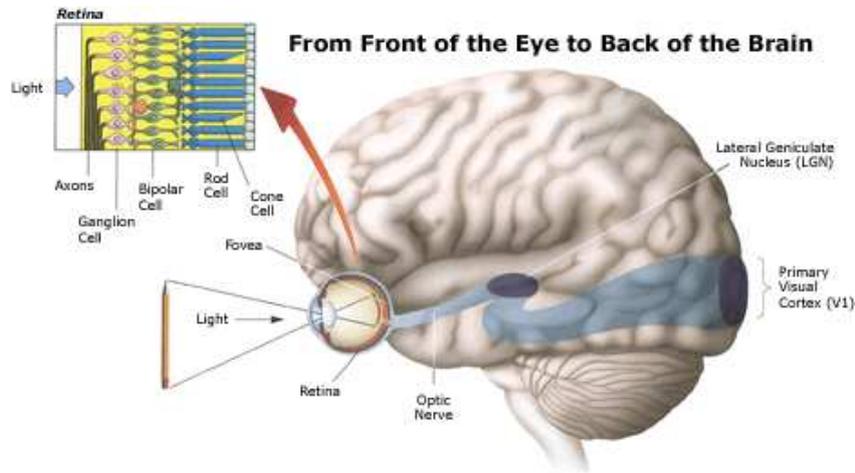


Figure 1.4: The visual pathway from the retina to the occipital region of the brain (from the Howard Hughes Medical Institute website¹¹). The top right panel shows a schematic view of the retina and its different layers. The neural code generated by the retina is transmitted to the LGN via the optic nerve. The neural code then continues its way through the chain of treatments to the V1 area.

The first organ encountered by a light stimulus is the retina. The retina is the device that acquires the visual stimuli and conditions the way high order cortical areas accomplish shape categorization, motion control and the other visual tasks that the cortex is expected to do.

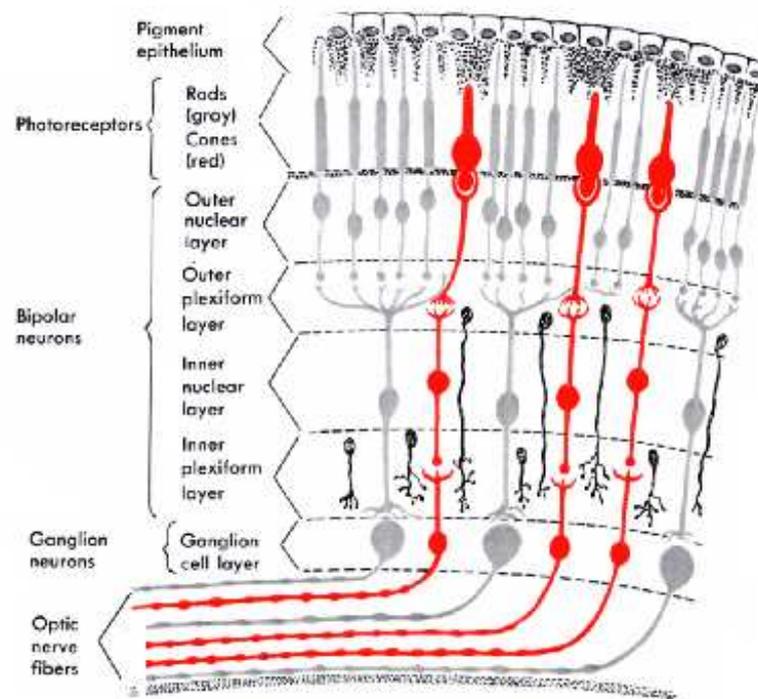
It has a layered architecture (see the right panel in Figure 1.4 and the schematic view in Figure 1.5). The successive layers of the retina perform consecutive transforms on the stimulus to finally generate the retinal neural code [Dowling 1966]. The outermost layer of the retina is tiled with a dense mosaic of photoreceptors. It is to be noted that photoreceptors are said to be outermost considering the chain of retinal treatments but the deepest ones “anatomically speaking”. Indeed, the retina is paradoxically “reversed” because the light must pass through the retina before it can reach the photoreceptors. In humans, there are several types of photoreceptors but only two are image forming: the rods and cones [Fu 2005]. The sensing topography defined by the rods and cones is at the origin of the visual process that samples the continuous stimulus passing through the ocular optics to a discrete array of signals. Rods are mainly devoted to the peripheral vision and to the motion detection, while cones are mainly devoted to color contrast perception. Though, this role

¹¹<http://www.hhmi.org/senses/b150.html>

¹²http://www.arthursclipart.org/medical/senseorgans/page_02.htm



(a)



(b)

Figure 1.5: 1.5(a): Optical coherence tomography (OCT) scan of the retina (from [Cabrera Fernández 2005]). 1.5(b) Schematic view of the successive layers of the retina (modified from Arthur's clipart website ¹²).

14 Chapter 1. Basics about the image compression and the HVS

distinction is not as strict as previously thought since rods seem to contribute to color contrast perception via the rod-cone gap junctions [Cao 2008].

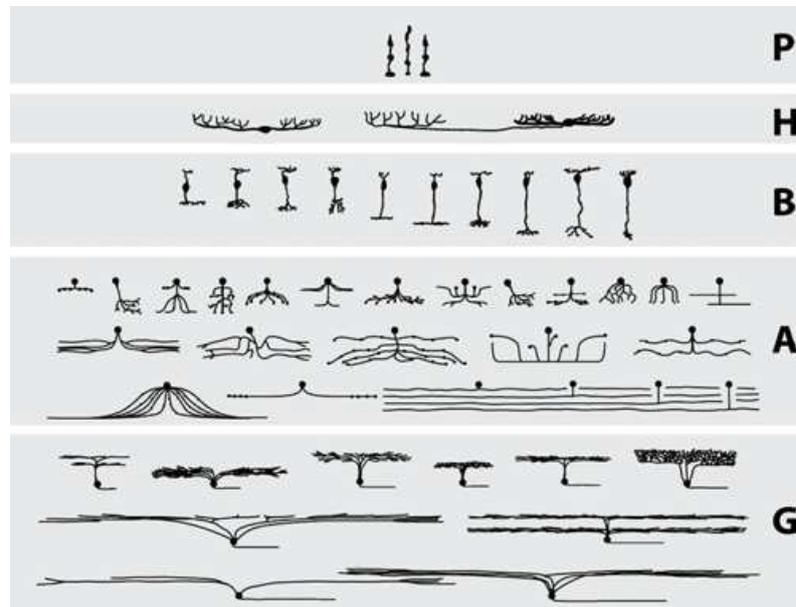
Among the subsequent layers we may cite the outer nuclear layer, inner nuclear layer, the outer plexiform layer, the inner plexiform layer, the ganglionic layer, and finally the internal limiting membrane. We will focus here on our main layers of interest, the inner nuclear layer and the ganglionic layer. These layers encompass the neurons of the retina. They are interconnected and each one has a specific function. Figure 1.6 shows the diversity of the retinal cell types and the specificity of their wiring.

The inner nuclear layer includes a variety of retinal neurons which may be classified into three categories:

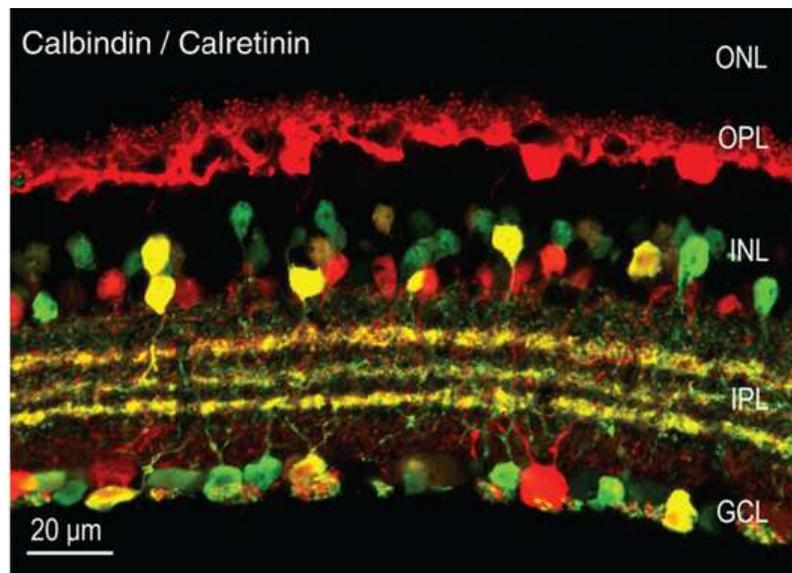
1. The bipolar cells that are thus named because they are connected to both the photoreceptor cells and ganglion cells: the two “poles”. There are several groups of bipolar cells: (i) the bipolar rods, connecting several rods to a ganglion cell, and (ii) the bipolar cones, connecting one or more cones to a ganglion cell.
2. The horizontal cells which that are thus named because of the position of their synapses. These cells modulate the information transmission laterally.
3. Amacrine cells that are in contact via their synapses with the bipolar and ganglion cells.

The last neuron layer of the retina is the ganglionic layer. The ganglion cells are terminated with wires called axons, and the set of these axons is the optic nerve. While, the human retina is populated with around 10^8 photoreceptors, there are only 10^6 ganglion cells that are responsible of the information transmission to the brain [Curcio 1990]. This lead us to conjecture that the retina compresses the visual data, and this fact provided us with a further motivation for this work.

The cells within each layer are characterized by their sensitivity region that is referred to as its receptive field. An example of such receptive fields in the case of bipolar cells is given in Figure 1.7. These retina layers encompass several processing stages including contrast detection filtering, gain control, and non-linear rectification. Interested readers may refer to [Gollisch 2010]



(a)



(b)

Figure 1.6: 1.6(a) Diversity of retinal cell types. For all five classes of retinal neurons?photoreceptors (P), horizontal cells (H), bipolar cells (B), amacrine cells (A), and ganglion cells (G) (from [Gollisch 2010]). 1.6(b) Specificity of retinal wiring (from [Gollisch 2010])

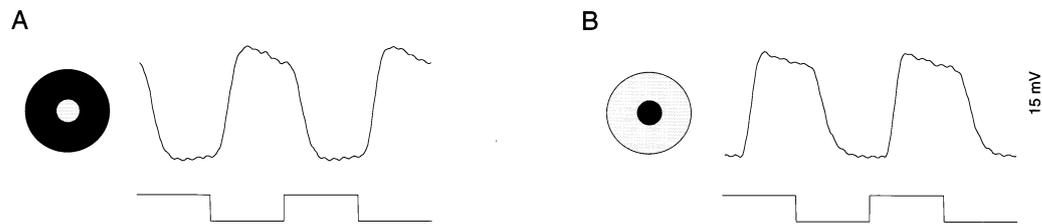


Figure 1.7: Center-surround receptive field structure of an OFF-center midget bipolar cell (from [Dacey 2000]). (A) Cell hyperpolarized to a small 150 μm diameter spot centered on the receptive field. (B) Cell depolarized to an annulus (inner diameter = 150 μm and outer diameter = 1200 μm).

for a review of the different retina layers. The behavior of these layers will also be detailed in Part II.

The ganglion cells transmit the image forming data via the optic nerve to a major cortical area: the LGN. The LGN is a sensory relay nucleus lying in the thalamus. Though the retinal ganglion cells have been categorized into five classes, only two of them project onto the LGN cells. These two classes are the M cells and the P cells. The M cells are connected to the magnocellular (or macrocellular) layers of the LGN, which are two in number for macaques. These layers initiate the magnocellular or dorsal pathway. Respectively, the P cells are connected the parvocellular layers of the LGN, which are four in number for macaques. These layers initiate the parvocellular or ventral pathway. Figure 1.8 schematizes the cited layers of the LGN. The LGN is thought to convey information mainly about spatial contrast in the stimulus, but the content-based differentiation between the two pathways remains highly unknown [Nealey 1994]. Though, several neurophysiologic studies suggest that the magnocellular pathway transmits low stimulus contrasts and contributes also in the transmission of high stimulus contrasts to V1. On the contrary, the parvocellular pathway appears to influence mainly the responses to high stimulus contrasts. Thus, the magnocellular pathway is often assumed to be the predominant conveyor of information about spatial contrast to the visual cortex [Purpura 1988]. Besides, the parvocellular pathway appears to be slower than the magnocellular one. The role of this speed differentiation is also unclear [Livingstone 1988, Maunsell 1999].

¹³<http://www.colorado.edu/intphys/Class/IPHY3730/07vision.html>

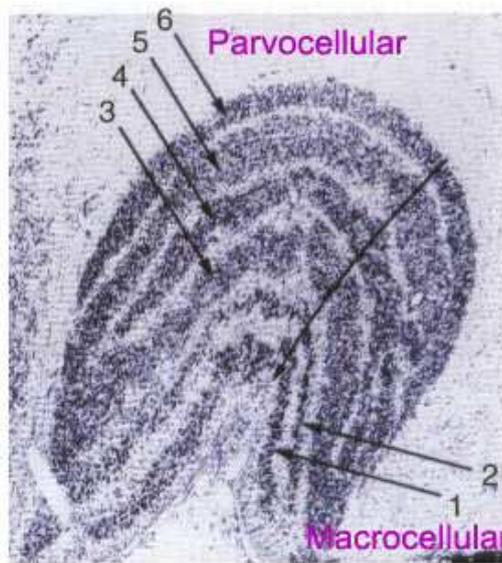


Figure 1.8: The lateral geniculate nucleus LGN (from the university of Colorado website¹³). The figure the multilayered architecture of the LGN. Four layers are parvocellular and two are magnocellular (also called macrocellular).

Another major supposed role of the LGN is to enable stereoscopy in the human visual system also referred to as binocular vision. Indeed, lesion studies have proven that the parvocellular layers of the LGN affects fine but not coarse static stereopsis [Schiller 1990]. Though, the actual involvement of LGN in stereoscopic vision is challenged in several studies [Lehky 1996].

The subsequent cortical area that receives information directly from the LGN is the primary visual cortex V1 also called the striate cortex. From that point, the visual information stream is then divided in two before flowing through the cortical hierarchy as detailed in the next section.

1.2.2 The dorsal and ventral streams

The visual system is thought to have a hierarchical architecture such that the chained cortical areas within it perform increasingly complex tasks. This chain of treatments yields decisions about (i) the cognitive operations as categorization, and about (ii) the motor actions such as grasping an object. Consequently, since [Goodale 1992] emerged the so-called *two streams hypothesis* which states that two distinct, though still communicating, visual pathways

are present within the visual cortex. One is dedicated to the perception of objects, and the other to the control of motor actions to be performed on those objects. These two pathways are termed as the dorsal and the ventral streams. The areas crossed by the dorsal stream lie in the so-called occipitoparietal region of the cortex, while those crossed by the ventral stream lie in the so-called occipitotemporal region. Figure 1.9 schematizes these two pathways in the apes brain. As described below, behavioural evidence derived from lesion studies and neuroimaging enabled the inference of these two streams functions.

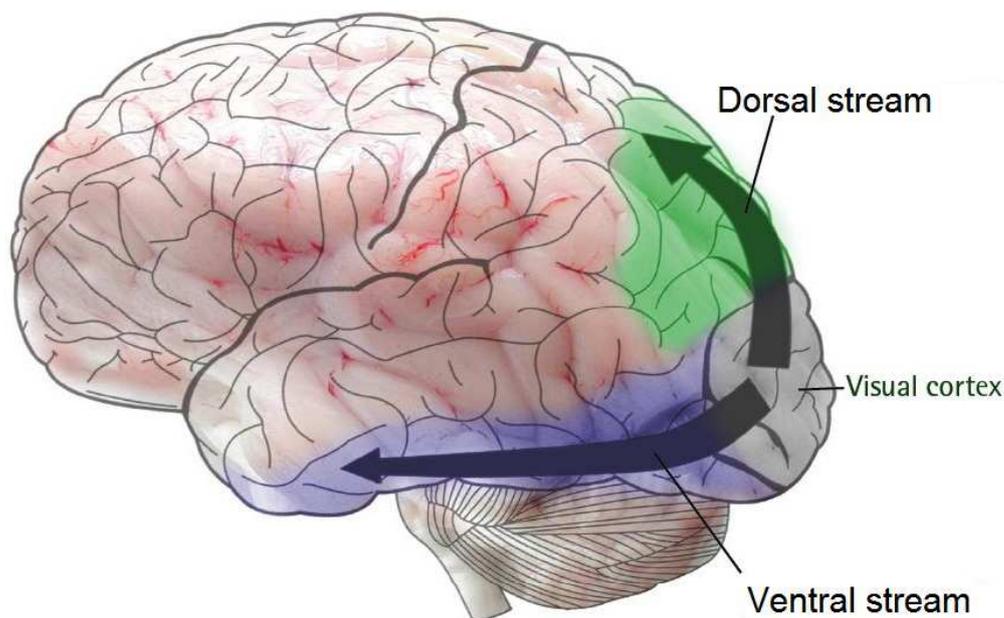


Figure 1.9: The ventral and dorsal visual pathways (modified from the Australian National University website¹⁴). The areas crossed by the dorsal stream lie in the occipitoparietal region of the cortex (in green). The areas crossed by the ventral stream lie in the occipitotemporal region (in purple).

The first stream is the dorsal “action” stream that projects to the posterior parietal cortex. It starts from the LGN magnocellular layers and encompasses a chain of consecutive areas including V1, V2, V3, MT, MST, and FEF [Schmolesky 1998]. The magnocellular pathway is responsible of the estimation of objects position in space and their movement comprehension. It helps visuomotor modules monitor the motor acts to be performed on the considered objects [Goodale 1992, Goodale 2004]. Behavioural evidence

supports this assertion about the role of the dorsal pathway. Indeed lesion studies showed that apes with posterior parietal lesions are impaired in the discrimination of rotated shapes or mirrored images, while cognitive tasks as face discrimination are not impaired [Ockleford 1977, Walsh 1996]. Besides, neuroimaging and particularly functional magnetic resonance imaging (fMRI) provided further evidence about the role of the dorsal pathway. For example, it has been shown that the anterior intraparietal (AIP) region of the brain responds more strongly during grasping than reaching towards visual objects [Binkofski 1998]. Because of the functions that the dorsal pathway is involved in, it is often referred to as the “where” pathway.

The second stream is the ventral “perception” stream that projects to the temporal lobe. It starts from the LGN parvocellular layers and encompasses a chain of consecutive areas including V1, V2, and V4 [Schmolesky 1998]. The parvocellular or ventral pathway is responsible of the elaboration of stimulus representations that are used for cognitive tasks such as the categorization of objects. The temporal lobe holds in its different areas rich and detailed maps representing the visual environment [Goodale 1992, Goodale 2004]. For example, (i) V1 contains inter alia an orientation-based representation of the stimulus contours [Li 2002, Tanaka 2009], (ii) V2 contains inter alia a map for visual horizontal disparity [Chen 2008], (iii) while V4 contains inter alia a retinotopic saliency map that guides the eye movements during a visual search task [Mazer 2003]. As for the dorsal pathway, lesion studies and neuroimaging support the hypothesis stating that the ventral pathway is responsible of the perception. Indeed, lesion studies showed that monkeys with lesions of the inferior temporal cortex were profoundly impaired in visual pattern discrimination, while they can discriminate between shapes that differ only in orientation [Gross 1973, Walsh 1996]. In this case also, fMRI brought further evidence about the role of the ventral pathway. In fact, the occipitotemporal cortex appears to be highly activated by 3D objects depicted in line drawings compared to textures or scrambled objects [Grill-Spector 2001].

Several studies state that the dorsal and ventral streams discrimination is

¹⁴<http://sciencewise.anu.edu.au/articles/dyslexia>

relatively strict at the level of V1 and in the adjacent V2. This segregation is also thought to predominate in V4 or in the middle temporal area MT. Though this hypothesis is challenged. For example, the cells of V1 seem to integrate information from both streams [Allison 2000]. Thus, the segregation between dorsal and ventral information in the cortex might not be as obvious as initially thought. The functional study of the visual pathways is further complicated by the criss-cross of the connections within. We describe in the next section this feedforward and feedback connectivity.

1.2.3 The feedforward and feedback connections

The visual system is thought to be organized into two hierarchical processing chains as described above: the first one goes through the dorsal stream and the second through the ventral one. Within each pathway there exists a dichotomous communication system with feedforward and feedback connections. The so called *top-down* processing of the neural code is done during the feedforward sweep, while the *bottom-up* processing is done during the feedback sweep. While the functional role of the top-down processing is inferred from the studies cited in the previous section, this role is less obvious for the bottom-up processing. Thus, in this section, we describe the functional role of the bottom-up processing for (i) visuomotor control in the dorsal pathway and (ii) for perception in the ventral stream.

The areas within the dorsal stream are responsible of complex tasks such as the monitoring of visuomotor actions. Such tasks require high order analysis of the position and movement of a given object. The hierarchy within the dorsal stream may suggest that a top-down scan through the feedforward connections is sufficient to accomplish the “where” tasks. We report in this section the literature results about the role of the feedback connections in this analysis process.

Yet, though V1 is a major relay within the dorsal stream, it is often considered to be a simplistic visual processor. Thus, it was thought that V1 is unable to accomplish visuomotor control, and that higher order cortical areas as MT are responsible of it. This is due to the apparently simple nature of the stimuli that activate its cells as bars and spots. In addition, the fact that

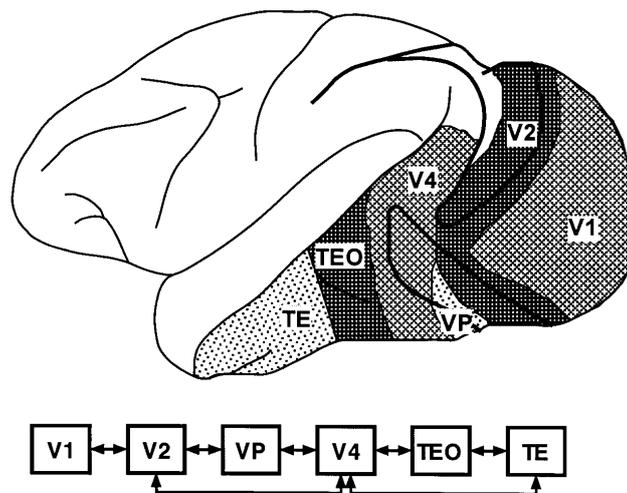


Figure 1.10: The feedback and feedforward connections in the ventral stream (from [Ungerleider 2000]). Here are shown the areas V1, V2, VP, and V4 in the occipital region of the brain and the areas TEO and TE in the temporal cortex. The connections between successive pairs of areas are reciprocal, such that feedforward projections from one area to the next are reciprocated by feedback projections from the second area back to the first.

V1 responds within 20 ms leaves a too short time for a complex analysis to be done [Bullier 2001]. Surprisingly, it appears that this assumption have to be reconsidered and that there is a need for bottom-up processing via the feedback connections. In fact, evidence has been made that V1 has a late activity that is initiated after the first top-down sweep is over, around 600 ms after stimulus onset. In particular, experiments using transcranial magnetic stimulation (TMS) has shown that this late activity of V1 was critical for awareness of visual motion [Pascual-Leone 2001]. Other results such as [Hupe 1998] concluded that V1 gets feedback information from the top of the hierarchy of the dorsal stream before the arrival of the parvocellular information. These efforts suggest that feedback connections in the dorsal stream contribute to the “where” functions but probably also to the “what” functions.

The ventral stream has also been shown to be crossed by feedforward and feedback connections. Indeed, anatomical studies reveal that almost each pair of successive areas within the ventral stream is connected back and forth [Felleman 1991]. Figure 1.10 shows the different areas crossed by the ventral stream and a diagram of the feedforward and feedback connections within. Several studies supported this assumption such as [Bullier 2001].

22 Chapter 1. Basics about the image compression and the HVS

Though, until recently, no experimental evidence confirmed the functional role of feedback in the ventral stream, whereas this has been demonstrated for the dorsal stream as described above [Koivisto 2011, Crouzet 2011]. Thus, the first feedforward sweep in the ventral stream was thought to be predominant for perception and cognitive tasks. In fact, while the visual neural code goes forward in the ventral stream, the cells of the crossed areas become selective to increasingly complex features. This progression enables categorization tasks for instance. In this context, experiments showed that simple go/no-go categorization tasks in apes and humans occur within less than 150 ms from stimulus onset [Thorpe 1996, Serre 2007]. Recent studies have also shown that more complex categorization tasks such as the recognition of valid and invalid coins occur within the same delay [Tallon-Baudry 2011]. Yet, the neural code travels through around 10 coding stages to reach the high order cortical areas and spends 10 to 20 ms per stage. Bearing in mind these facts, the categorization occurs too early to consider that the feedback pathway brings a crucial contribution for perception. The cited works lead to the conjecture that, despite the existence of feedback connections, a single forward sweep through the ventral stream is sufficient to perform complex cognitive tasks.

However a recent study in [Koivisto 2011] proved the causal role of the feedback pathway in natural scene categorization. The authors tested a go/no-go categorization task as in [Thorpe 1996]. The results showed that V1 and V2 contribute to categorization and subjective perception during a long activity period after the first feedforward sweep is over. This result suggests that the feedback connections within the ventral stream accounts for the “what” perception functions.

We introduced in this section the visual pathways that the neural code travels through starting from the retina, and the processing stages that they involve. This overview highlighted the complexity of the processes conducted within the visual cortex. Though, it has to be noted that these complex visual tasks can be done thanks to the neural code generated by the retina. So that, we expect this neural code to account for a variety of visuomotor and cognitive issues. As our main goal is to inspire ourselves from the retina to design image

coders, we will bring into focus the retina neural code in the next section.

1.3 Fundamental notions about the neural code of the retina

In this section we present some prerequisites about the neural code of the retina. In Section 1.3.1, we will especially introduce the fundamental notion of spikes which are the coding entities that convey visual information. Then, in Section 1.3.2, we will give an overview of a variety of possible spike-based coding strategies that might be employed by the retina.

1.3.1 The retinal neural code shape

The human visual system conveys information broadly by means of a set of electrical impulses termed as action potentials or also *spikes* [Rieke 1997, Gerstner 2002]. Spikes are emitted by specific neurons tiling the retina referred to as ganglion cells. As a response to a given visual stimulus each ganglion cell emits a series of spikes: the *spike train*. The set of these spike trains defined over the set of ganglion cells is the neural code of the retina.

Interestingly a spike as generated by a ganglion cell is an all-or-none event. This means that it occurs fully or do not occur at all. As a consequence, one can consider that the retina quantizes the visual data. Furthermore, neurophysiologists noticed that all spikes have almost the same characteristic shape and amplitude. So that, the spike amplitude is independent of the stimulus intensity. This observation lead neurophysiologists to conclude that the retina generates a binary-like code. Thus the neural code is often represented by a so-called *raster plot*. A raster plot is intended to represent spike occurrences on a binned time axis, this over a set of observed neurons. A schematic view of a spike, a spike train, and a raster plot is shown in Figure 1.11. This figure also shows the hierarchy relating these different notions.

The raster plot contains all the information about the acquired visual stimuli but is still a challenging code to understand. The neural code of the retina is not a simple intensity signal transform and through experiments, it appears that the code corresponding to a given stimulus encompasses many

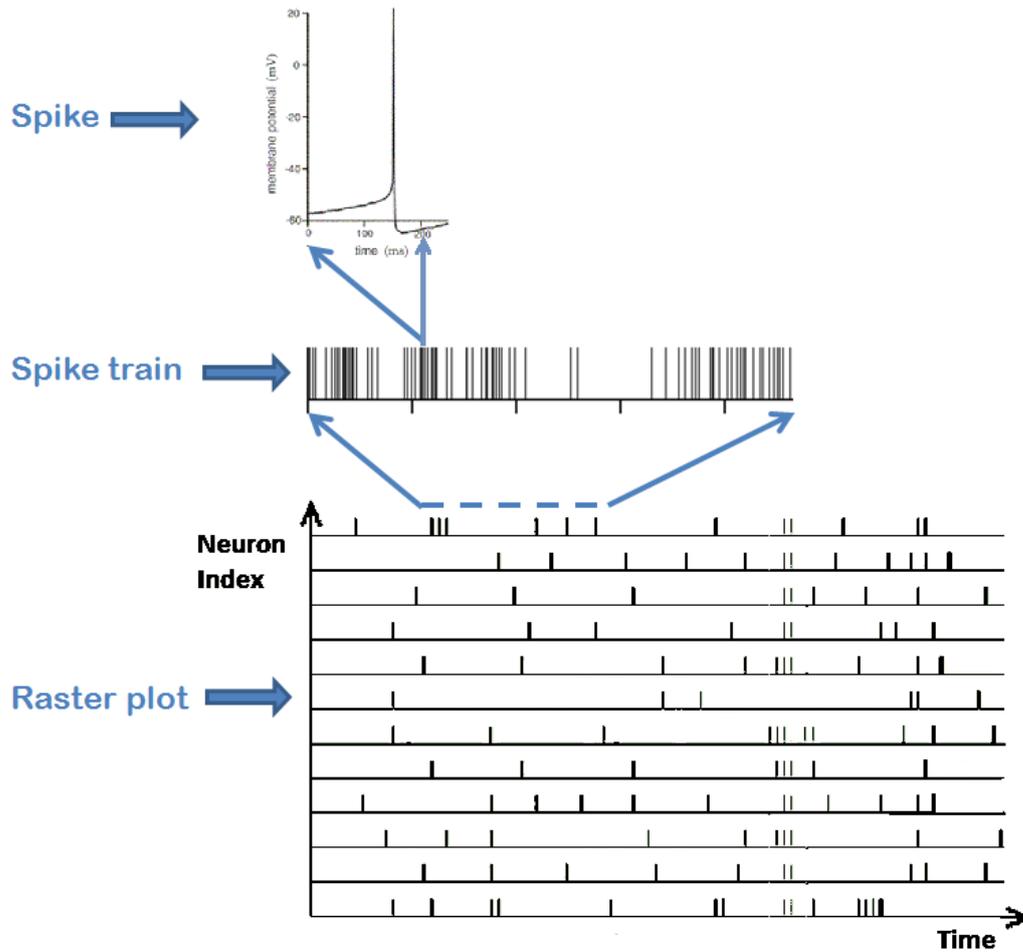


Figure 1.11: A spike, a spike train, and a raster plot. Spikes have the same shape and amplitude. A series of spikes emitted by a single neuron in a predefined time bin is a spike train. The whole set of spike trains is the neural code of the retina. The neural code is represented using a raster plot that contains all the information about the visual stimuli.

measurable characteristics, and that each one of these characteristics could convey the relevant information about the stimulus. We give an overview of the main hypotheses discussed in the literature in the subsequent section.

1.3.2 Possible coding metrics in the retina

Since [Adrian 1926] it was claimed that the relevant information would be contained mostly in the mean spike firing rate. Whereas, in [Perkell 1968a], authors suggested that the information conveyed by each single spike timing could be relevant with regard to the stimulus. Since then numerous other coding metrics were proposed as the relevant ones to be measured in the mass of data that is represented in a raster plot. Figure 1.12 outlines a few of them. Roughly two main sorts of coding schemes are proposed in the literature: rate codes and timing codes.

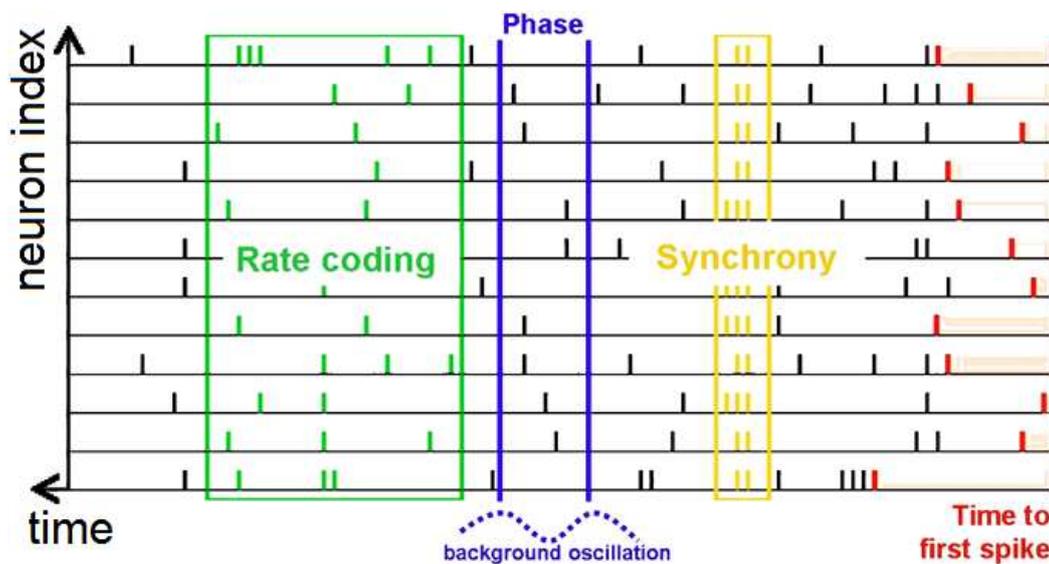


Figure 1.12: Some examples of possible relevant measures in the spiking code (from [Escobar 2009]). The vertical axis represents a range of ganglion cells. The horizontal axis is oriented from right to left and represents time. Each line in the Figure is a spike train. *In green* Rate coding : For each cell, the firing rate is the spike count over a predetermined time window, of width T , divided by T . *In blue* Phase coding : the phase of spike signals is measured with respect to a background carrier wave (here in dotted line). *In orange* Synchrony coding: ganglion cells reacting to a stimulus region "belonging the same object" fire simultaneously. Cells are thus grouped into clusters. This defines a bio-inspired segmentation algorithm. *In red* Time to first spike coding: The code output is the time to emit a first spike for each ganglion cell. In contradiction with the previous coding scheme, this assumes complete asynchrony in the encoding procedure.

Rate coding metrics (or also frequency coding metrics) assume that each ganglion cell conveys most of the valuable information about the visual stimuli in its spike firing rates. Rate coding strategies are widely used in the literature. This is due to the non-deterministic behavior of the retina as a coder [Stein 2005]. Indeed, given a single visual stimulus, the retina generates different (though still resembling) neural codes over the repeated trials. So in order to define a single output code for a given stimulus, experimenters had to consider the neural code from probabilistic point of view. The most commonly used approach is to recourse to averaging methods that are robust with the regard to the trial-to-trial code variability. These averaging methods are referred to as rate coding strategies. We draw the reader attention to the fact that rate coding strategies confusingly refer to a range of metrics. Among them, we can cite three distinct metrics depending on the experimental protocol:

- (i) *Trial-to-trial averaging.* The experimenter presents a visual stimulus to the retina neurons over several trials (cf. Figure 1.13). Each neuron is considered separately and the spike trains generated are recorded. For each neuron the experimenter averages the spikes count, over the number of trials, in a succession of relatively tight time bins. Typically, a time bin has a size of 1 to 5 *ms* but greater values could be used as in the example below where the authors used a time bin of 30 *ms*.
- (ii) *Population averaging.* The experimenter presents a visual stimulus to a population of uncorrelated retina neurons (cf. Figure 1.14). All the neurons in the population are submitted to the same stimulus. The experimenter averages their spike counts, over the number of neurons in the recorded population, in a succession of relatively tight time bins. Here also, a time bin has a size of 1 to 5 *ms*.
- (iii) *Time averaging.* The experimenter presents a visual stimulus to the retina neurons (cf. Figure 1.12 green lines). The spike trains are generally recorded using an array of electrodes. The experimenter averages the spike count over a relatively wide time window for each single neuron. Typically the time window has a size of 50 to 500 *ms*.

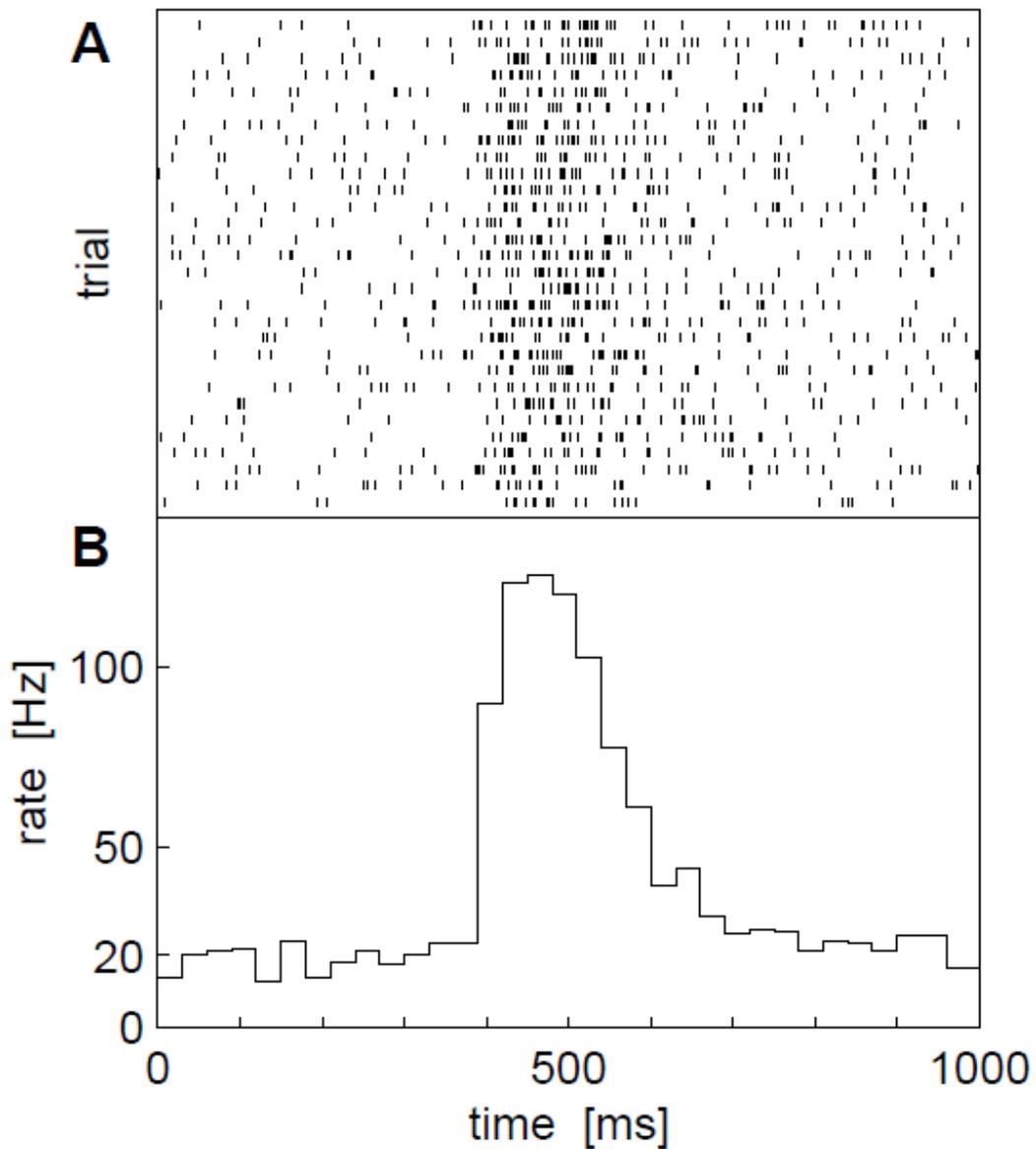


Figure 1.13: Rate estimation by means of trial-averaging (from [Nawrot 1999]). (A) Raster display of spike events for 30 trial repetitions. (B) Peristimulus time histogram (PSTH¹⁵) of average spike response, constructed from all 30 trials using a bin size of 30 ms.

The latter time averaging assumption is the one that is the most commonly used in the literature (Figure 1.12 green lines). For a more detailed discussion about the above cited rate codes see [Rieke 1997]. In addition to these simple rate codes, more sophisticated frequency-based approaches emerged. For example in [Freeman 2001], the authors consider that a stimulus alters a pre-defined carrier wave signal. The retina could then encode information in the

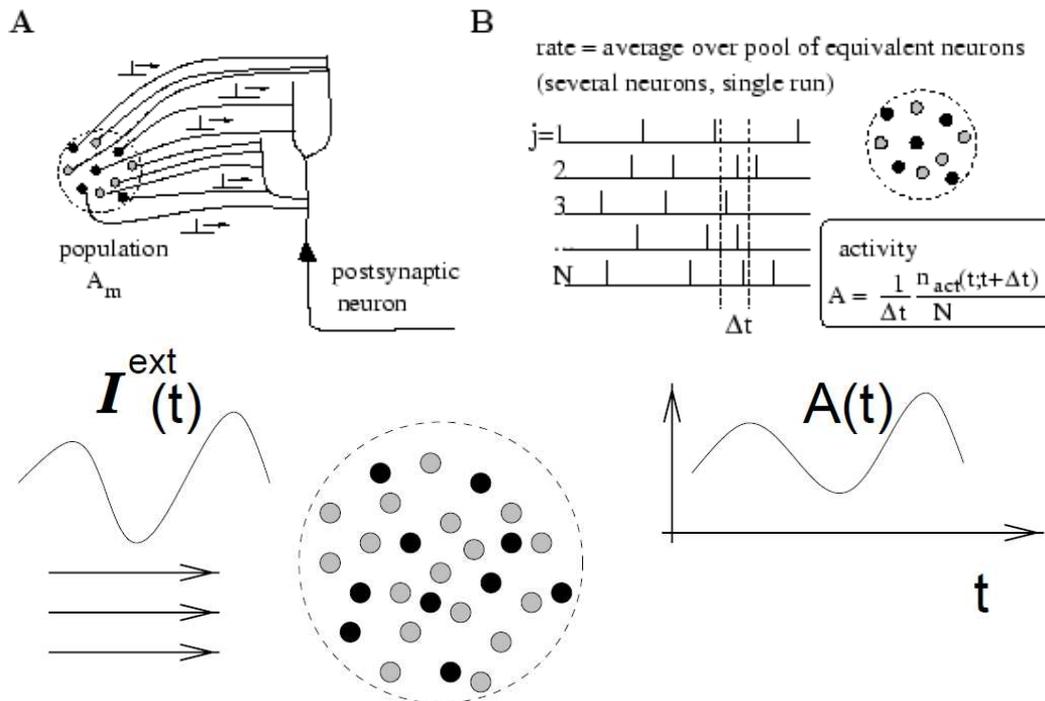


Figure 1.14: Population of neurons (from [Gerstner 2002, Gerstner 2000]). All neurons receive the same input $I^{ext}(t)$ (left), which results in a time-dependent population activity $A(t)$ (right).

phase of spikes with respect to the background wave oscillation (Figure 1.12 blue lines).

Though largely adopted, several studies such as [Gautrais 1998, Stein 2005] criticized rate coding strategies. Several timing codes were proposed as an alternative.

Time coding metrics emerged in the last twenty years to offer an alternative to rate codes. Indeed, there is some biological evidence stating that rate codes are unable to discriminate rapidly-varying visual stimuli. Besides, decisions made in the visual cortex occur in such a rapid manner that the underlying coding process could not rely on the estimation of a neural firing rate over a large time window.

One first approach is the *time-to-first-spike* coding strategy (Figure 1.12 red lines). Many authors proposed that the first wave of spikes that the retina neurons generate contains most of the information about the stim-

¹⁵The PSTH shows neuron firing times wrapped to one cycle of the stimulus. In this case, the stimulus is periodic.

ulus [Johansson 2004, VanRullen 2005, Zohar 2011]. Clearly, neurophysiologists demonstrated that the more a neuron is excited the shorter is the latency time of its first spike. So that, one can assume that considering only the first spike per each neuron in the retina enables the visual cortex to accomplish high order visual tasks. In addition, there is conjecture confirming this hypothesis for other sensory modes as the auditory system [Phillips 1998] and somatic nervous system [Panzeri 2001]. This approach was mainly used to explain the fast categorization ability in the visual cortex [Thorpe 1996, Johansson 2004]. Interestingly, the time-to-first-spike coding scheme found also technical applications in bio-inspired electronic sensory networks. For instance, this strategy has been applied to devise a low-consumption image sensor [Guo 2007]. We will reconsider this approach in the next chapter when describing the rank order coding scheme.

One limitation of the time-to-first-spike scheme is the fact that the stimulus onset time is not always clearly identified in natural situations, ie. when recording is not made in laboratory conditions. By using relative spike times one can avoid this issue. For example, authors in [Gollisch 2008] reported that the retina neurons encode the spatial structure of a visual stimulus in the relative timing of their first spikes. Still, these approaches discard all the spikes following the first one per each neuron. So that, one alternative approach inspired from the efforts conducted in the auditory system study is to measure the *interspike intervals* (ISI) [Shih 2011]. Obviously, this metric accounts for the spikes fired after the first one. Several experiments about this metric suggest there exist a relationship between the ISI measures and the stimulus amplitude, and thus ISI might be exploited as a part of the neural code [Oswald 2007].

The coding schemes above are mainly used for categorization tasks. When considering different visual tasks such as segmentation, other metrics appear to be more relevant. One possible coding strategy, that is suitable for segmentation, is to group neurons into clusters which elements fire simultaneously. The underlying idea is that simultaneous firing could mean “belonging to the same region” in the image perceived (Figure 1.12 orange lines).

The coding strategies cited above, though not exhaustive, show that the spike-based possible coding strategies are numerous. Readers interested in a detailed overview of possible coding strategies in the retina may refer to [Van Rullen 2002, Gerstner 2002, VanRullen 2005]. Among all these strategies we chose the rank order coding (ROC) as a basis for our work, where the ROC is an extension for the time-to-first-spike hypothesis. The reasons behind this choice are exposed in the following section.

1.4 Discussion

On a first step in this chapter, we based ourselves on the two streams hypothesis and overviewed the visual processing pathways within the cortex. We described the functional behavior of the dorsal and the ventral streams. We reported the neurophysiologic evidence about the contribution of the dorsal pathway in the monitoring of visuomotor tasks. We also reported literature results that proved the contribution of the ventral stream in the cognitive visual tasks as the categorization. Then, we described the feedforward and feedback interconnections that exist within each stream. In particular, we described several behavioral experiments showing that: (i) during the feedforward scan a rapid top-down processing pass is done in less than 150 ms from stimulus onset, (ii) and that during the feedback scan, a slower bottom-up processing is accomplished 600 ms after stimulus onset.

On a second step in this chapter, we presented some fundamental concepts about the neural code of the retina such as spikes. Then, we discussed a variety of measurable metrics that could be used by the retina to encode the information. Among these spike-based coding strategies, we distinguished rate and timing codes. Whether the retina employs rate or timing codes is still an open issue in the neuroscience community. The measurement of firing rates are the most common. This due to the experimental ease of these measurements and their resilience to the trial-to-trial variability of the retina neural code. However, we reported some experimental evidence suggesting that rate may be too simplistic to account for the rich information content conveyed by spikes. Yet, we evoked several studies demonstrating that coding

strategies based on the exact spike timings might be relevant.

The goal of this work is to conceive an image coder that is inspired from the retina. So that, our coder will be necessarily based on a model of the retina. Ideally, this retina-based coder should be optimized for perception, ie. when compared to other coders at equal rate we expect it to have a better perceptual quality. Furthermore, our coder should be simple. Thus, considering late feedbacks in the process that generates the neural code might be inappropriate for this sake. A fast top-down analysis of the stimulus appears to be more relevant. Finally, our coder should be rapid. Thus considering a coding strategy that integrates spikes during a long period of time before averaging might be inappropriate. For the reasons cited above, we will focus our work on the neural code generated by the retina that circulates through the feedforward connections of the ventral stream.

Fortunately, in the literature there exist perception-oriented retina models that could account for the rapidity of processing observed in the primate visual system as the one introduced in [Van Rullen 2001b]. The main working hypothesis behind this model is the rank order coding assumption which states that: “*The first wave of spikes surfing forth through the ventral stream encodes the perception-relevant information about the stimulus*” [Van Rullen 2002]. Part I will introduce the retina model under study and the fundamental concepts behind it.

Part I

FROM THE RANK ORDER CODE TO A BIO-INSPIRED SCALABLE IMAGE CODER

THE RANK ORDER CODE

Contents

| | | |
|------------|---|-----------|
| 2.1 | The rank order coding | 38 |
| 2.1.1 | The rank order coding principles | 38 |
| 2.1.2 | The experimental evidence behind | 38 |
| 2.2 | A mathematical formalism for the rank order coding | 42 |
| 2.3 | A retina model based on rank order coding | 43 |
| 2.3.1 | The retina cells: model and architecture | 43 |
| 2.3.2 | Image analysis | 47 |
| 2.3.3 | Image synthesis | 50 |
| 2.4 | Discussion | 51 |

Keywords

Rank order coding, spikes, retina model, difference of Gaussians, dyadic grid, permutations.

Overview

In this chapter, we will focus on our main working hypothesis for the first part of this work: the rank order coding. First, we will describe the guiding principles of this hypothesis as stated in [Thorpe 1990] and we will report the neurophysiologic results supporting it. Second, we will give a new mathematical formalism that rigorously defines the rank order code. Finally, we present the simplified model of the retina in [Van Rullen 2001b] that is based on the rank order coding.

Contribution

We bring a new mathematical formalism for the rank order coding. This formalization of the hypothesis statement is intended to provide a rigorous basis for future works. Details are described in Section 2.2.

Organization

This chapter is organized into three sections:

1. Section 2.1 is divided into two sections. Section 2.1.1 enumerates the guiding principles of the rank order coding. Section 2.1.2 reports several neurophysiologic results that sustains this hypothesis.
2. Section 2.2 gives a mathematical formalism for the rank order coding.
3. Section 2.3 presents a retina model defined in [Van Rullen 2001b] that is based on the rank order coding.

There is no logical way to the discovery of these elemental laws. There is only the way of intuition, which is helped by a feeling for the order lying behind the appearance.

Albert Einstein

As discussed in the previous chapter, several hypotheses about the coding features within the neural code of the retina were proposed. Among these hypotheses, we will focus on the latency time of the first spike given the stimulus onset. In particular, we will consider in the first part of this work an extension of this hypothesis: the rank order coding. This choice was motivated by compelling neurophysiologic evidence and technical reasons as well. So that, this chapter will be structured into three sections. The first one will introduce the rank order coding founding bases, the second will formalize mathematically the concept of rank ordering, and the third will detail a retina model design intended to support it. First, in Section 2.1 we will recall the principles of the rank order coding hypothesis and the experimental evidence that sustains it since the Thorpe founding effort [Thorpe 1990]. Second, we will give a new mathematical formalism for the rank order coding strategy in Section 2.2. This formalism is intended to offer a rigorous framework for our present effort and also for future extensions to be made. Finally, in Section 2.3 we will detail the specification of a retina model proposed in [Van Rullen 2001b] to support the rank order coding. The model under study transforms an image into a wave of spikes in a bio-inspired manner. In addition, this retina model provides both a coding and a decoding algorithm which makes it a further interesting choice for a bio-inspired coder basis. The adaptation and usage of the retina model in [Van Rullen 2001b] in the context image coders design will be detailed in the next two chapters.

2.1 The rank order coding:

Principles and neurophysiologic evidence

The rank order coding scheme will be the basis of the first part of this work. So that, this section will be structured into two parts. First, in Section 2.1.1, we enumerate the guiding principles of the rank order coding hypothesis. Second, in Section 2.1.2, we give an overview of the neurophysiologic experimental results that supports it.

2.1.1 The rank order coding principles

As an explanation for the apes extraordinary performance in cognitive tasks, Thorpe proposed in its founding effort [Thorpe 1990] the rank order coding hypothesis. This hypothesis states that the order in which the retina emits its first wave of spikes encodes the stimulus. The rank order coding followed a review of experimental results conducted in [Thorpe 1989] which conjecture that the visual system accomplishes its main cognitive tasks in a single feedforward pass through the ventral stream. The hypothesis relies on the following simplifying assumptions:

- (i) From stimulus onset only the first spike emitted by each neuron is considered in the retina response. All subsequent spikes are discarded.
- (ii) The latency time before a neuron fires its first spike is a decreasing function on its degree of excitation.
- (iii) Only the order in which the neurons fire is the neural code of a given stimulus. The exact latency time of each single neuron is discarded.

Since then, accumulating neurophysiologic results came as a support for this hypothesis and more generally to approaches based on the time-to-first spike metric. These results further motivated our choice of it. In the following section, we overview these results.

2.1.2 The experimental evidence behind

Since the sixties, neurophysiologists has been interested in the coding features within the retina neural code. This issue is fundamental since the comprehen-

sion of the retina code is determinant for the comprehension of the functional role of the visual cortical areas, and vice versa. A widely admitted assumption states that the retina neurons convey the information by means of their firing rate. Though as described in Section 1.3.2, alternatives to rate coding emerged such as the time-to-first spike metric. The latter hypothesis relies on the experimental evidence that the more a neuron is excited the shorter is its first spike firing latency. Figure 2.1 schematizes this phenomenon. One major

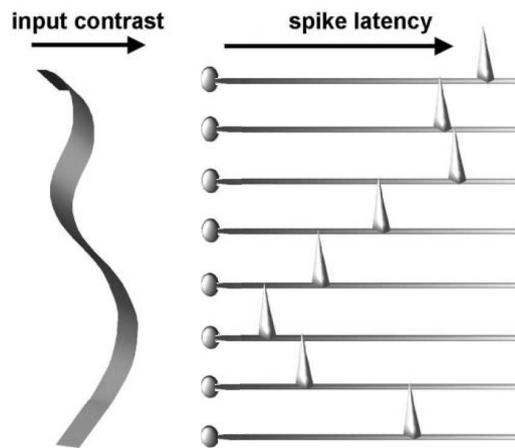


Figure 2.1: The first wave of spikes (from [Van Rullen 2002]). The more a neuron is excited the shorter is its first spike firing latency time. All spikes following the first one, at the level of each neuron, are discarded.

result that lead in this direction is that the image coding performed by the retina allows face recognition tasks to be undertaken within short latencies between 80 and 160 ms [Perrett 1982]. Since then, several experimental studies corroborated approaches based on the time-to-first spike metric. Among these approaches, the rank order coding [Thorpe 1990] got a specific attention. In the following we report some of these experimental results.

Oram et al further studied face recognition capabilities of primates in their 1992 effort [Oram 1992]. The evidence was made that the decision occurs quickly though information has to be processed through several stages in the brain. Typically, the discrimination between static head views stimuli occurs within 25 *ms* from response onset. The authors stated that this measures could make sense if the visual cortex considered only the first spike emitted by each neuron for face recognition tasks. As a consequence authors raised the idea that the first wave of spikes encodes the stimulus. Obviously, this

coding scheme relies on the time-to-first-spike metric.

The previous result was generalized by Thorpe et al in [Thorpe 1996]. Authors confronted the rapid categorization capabilities of the visual cortex to stimuli that are not specific to face recognition. Stimuli considered for the experiment are natural complex scenes. The ape subject was given a go/no-go task, namely it had to answer the question whether there is an animal in the scene or not. Pictures were flashed during 20 *ms* and the generated retina code (the event related potentials in this case) was recorded in the frontal sites of the brain. The study of the output code showed that decision occurs within 150 *ms* from stimulus onset. Later this result was extended to non biologically relevant targets in [Van Rullen 2001a]. As an explanation the authors claimed that the first wave of spikes conveys the code used for rapid categorization tasks.

Later, Meister et al validated the relevance of the time-to-first-spike metric in the neural code of the retina [Gollisch 2008]. This effort relied on quantitative measures. Authors investigated the neural output of the salamander retina. The stimuli considered were flashed still images of gratings. Spike trains were recorded at the level of the retina ganglion cells. Two metrics were then measured over spike trains. The first one was the time-to-first-spike and the second was the spike firing rate within a time window of fixed size. The experimental results confirmed that the time-to-first-spike code is more discriminative than the rate code with regard to the input stimuli. Besides, the amount of information measured by the entropy (see [Cover 1991] for a definition) showed that the time-to-first-spike code conveys twice as much information as the rate code [Gollisch 2008]. These measures validated the assumption that time-to-first-spike metric is predominant in the visual neural code.

Recent studies have also shown that more complex categorization tasks occur within extremely short delays [Tallon-Baudry 2011]. The authors considered two stimuli categories: (i) valid coins such as Euros or Australian Dollars and (ii) invalid coins that lost their change value such as French Francs or Finnish Marks. The authors then recorded magneto-encephalographic responses to visual stimuli from the two categories. It appeared that the ventral visual pathway discriminates between valid and invalid coins within only 150

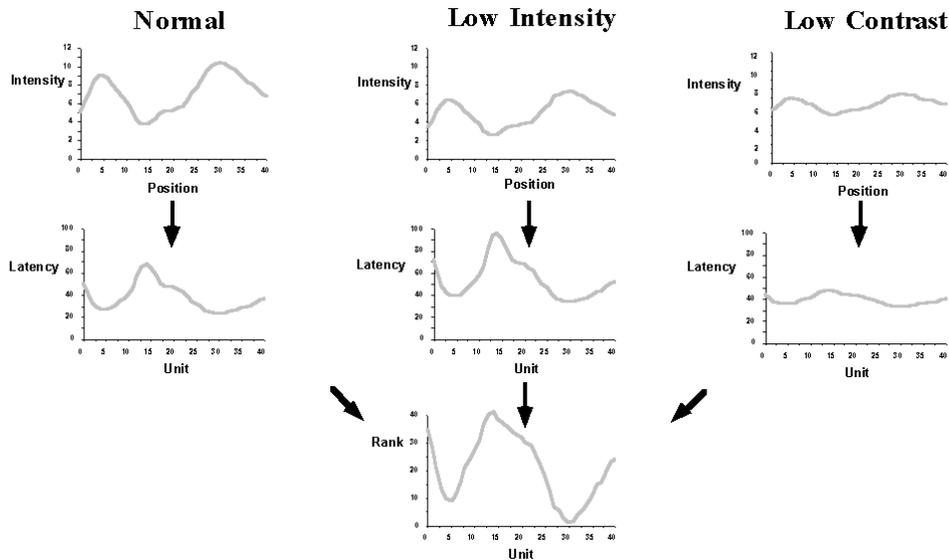


Figure 2.2: Invariance of ROC to stimuli contrast and dynamic range (from Spikenet documentation²¹). The three graphs on the left illustrate how an input intensity profile can be converted first into a latency profile, and then into a rank profile. The center and right-hand panels show that changes in either intensity (center) or contrast (right) will have no effect on the rank ordering of the units, although latencies will change.

ms. The categorization was successful whether the coins were familiar or not.

As discussed above neurophysiologic evidence supports the idea that only the first wave of spikes encodes the visual stimulus in the retina. This idea was extended with the proposition of the rank order coding [Thorpe 1990, Thorpe 1996, Van Rullen 2001b, Van Rullen 2002, Delorme 2003]. The rank order coding is based on the assumption that the visual stimulus is not encoded by the exact time-to-first-spike of the retina neurons but rather by their order of occurrence. One major advantage of order coding over time-to-first-spike coding is its invariance to the contrast and dynamic range of the stimulus. Authors in [Delorme 2003] demonstrated this property (see Figure 2.2). This invariance property is important because it accounts for the dynamic range and contrast invariance observed in some areas of the visual cortex. Especially in V1, the neurons should discriminate orientations independently from the contrast and illumination of the stimulus image. The authors then concluded that coding the visual stimulus by means of the neurons firing rank is more coherent with the subsequent visual cortical areas behavior.

We described in this section the guiding principles and the biological evidence behind it. Though the idea at the basis of this code is simple, the

literature lacks a formal mathematical definition for it. So that, we formalize in the following the coding function underlying this hypothesis.

2.2 A mathematical formalism for the rank order coding

On a first step in this section, we give a mathematical formalism that defines the time-to-first-spike coding scheme within a rigorous framework. Then on a second step, we extend this definition to the rank order coding scheme.

As described previously many authors consider that only the first wave of spikes conveys the relevant information about the stimulus. Under this assumption, one can define the neural code of the retina as the set of the occurrence times of the first spikes given the stimulus onset. Because the literature lacks a formal definition for this code, let us give it in the following. If we suppose that (i) the stimulus is a regular $n \times m$ array of intensities $f \in M_{m,n}$, and that (ii) the retina is tiled with a regular $N \times M$ array of neurons $(\nu_{ij})_{(i,j) \in \llbracket 0, M-1 \rrbracket \times \llbracket 0, N-1 \rrbracket}$, we define the retina time-to-first-spike coding function ϕ_{wave} by:

$$\begin{aligned} \phi_{wave} : M_{m,n}(\mathbb{R}) &\longrightarrow M_{m,n}(\mathbb{R}^+) \\ f &\longmapsto \tau, \end{aligned} \tag{2.1}$$

where τ_{ij} is the occurrence time of the first spike emitted by the neuron ν_{ij} given the stimulus onset. According the time-to-first-spike scheme, the cortex could recover f , the original stimulus, given its code $\tau = \phi_{wave}(f)$. The rank order coding proposed in [Thorpe 1990] is a biologically plausible extension of this coding strategy. As described in Section 2.1.1, this coding scheme supposes that the order in which the retina neurons emit their first spike is the only metric that encodes a given stimulus. Let us recall the rank order coding assumptions:

- (i) Only the first spike emitted by each neuron is considered.

- (ii) The more a neuron is excited the shorter is its firing latency time.
- (iii) Only the order in which the neurons fire encodes a given stimulus.

Thus, we define the retina rank order coding function ϕ_{roc} by:

$$\begin{aligned} \phi_{roc} : M_{m,n}(\mathbb{R}) &\longrightarrow \mathfrak{S}(\llbracket 0, M-1 \rrbracket \times \llbracket 0, N-1 \rrbracket) \\ f &\longmapsto \sigma, \end{aligned} \quad (2.2)$$

where $\mathfrak{S}(\llbracket 0, M-1 \rrbracket \times \llbracket 0, N-1 \rrbracket)$ is the group of possible permutations over the elements of $\llbracket 0, M-1 \rrbracket \times \llbracket 0, N-1 \rrbracket$, and σ is the permutation such that:

$$\tau_{\sigma(0,0)} < \tau_{\sigma(1,0)} < \dots < \tau_{\sigma(i,j)} < \dots < \tau_{\sigma(M-1,N-1)}, \quad (2.3)$$

where $\tau = \phi_{wave}(f)$. This means that σ is the permutation that sorts the neurons $\nu(i, j)$ in the increasing order of their time-to-first-spike $\tau(i, j)$ (cf. Equation (2.1)).

In order to test the relevance of rank order coding, a model of the retina was proposed in [Van Rullen 2001b]. With the aid of the formal definition in Equation (2.1), we describe this spike-based model in the next section. The model described will be at the heart of the following chapters of this work.

2.3 A retina model based on rank order coding

In this section, we describe the retina cells model and architecture according to the bio-inspired retina model specified in [Van Rullen 2001b]. We then show how this modelled retina analyzes then synthesizes a still image stimulus.

2.3.1 The retina cells: model and architecture

Neurophysiologic experiments have shown that, as for classical image coders, the retina encodes the stimulus representation in a transform domain. The

²¹<http://sccn.ucsd.edu/~arno/spikenet/order.html>

retinal stimulus transform is performed in the cells of the outer layers. Quantitative studies such as [Field 1994, Rodieck 1965] have proven that the outer cells processing can be approximated by a linear filtering. In particular, neurophysiologic experiments conducted in [Field 1994] yielded the largely adopted DoG filter which is a weighted difference of spatial Gaussians that is defined as follows:

$$DoG(x, y) = w^c G_{\sigma^c}(x, y) - w^s G_{\sigma^s}(x, y), \quad (2.4)$$

where w^c and w^s are the respective positive weights of the center and surround components of the receptive fields, σ^c and σ^s are the standard deviations of the Gaussian kernels G_{σ^c} and G_{σ^s} , such that $\sigma^c < \sigma^s$. Based on [Field 1994], the retina model specification in [Van Rullen 2001b] chose the DoG filter to model the retina cells and set the weights to $w^c = w^s = 1$ and the standard deviations ratio to $\frac{\sigma^s}{\sigma^c} = 3$. The latter values are biologically realistic. A DoG filter thus defined is a contour detector as shown in Figure 2.3.

Having these parameters, the value of σ^c (cf. Equation (2.4)) determines the spatial resolution of the retina cell considered. In the actual retina, one can observe a wide range of cells with different scales. Besides, the retina cells density and scale are inversely proportional. The authors in [Van Rullen 2001b] modelled this architecture by spreading the retina cells over a hierarchical grid Γ that is said to be *dyadic*. Formally a finite dyadic grid is defined as in [Cardoso 2006]:

Définition 1. *Finite dyadic grids* : *A finite dyadic grid Γ has an obvious unique representation as a finite 0-4 tree, a quadtree where every node (which represents a cell of Γ) has either 0 or 4 children nodes. Conversely, every 0-4 tree defines a finite dyadic grid.*

Such a grid is represented in Figure 2.4. In this architecture, all filters of the same level Γ_k in Γ have the same scale. For each scale the receptive fields of the retina cells have to tile the stimulus image space. Thus each layer Γ_k in the grid Γ (with $0 \leq k < K$) is tiled with filtering cells denoted by DoG_k , having a scale k , and generating a transform subband B_k such that:

$$DoG_k(x, y) = w^c G_{\sigma_k^c}(x, y) - w^s G_{\sigma_k^s}(x, y), \quad (2.5)$$

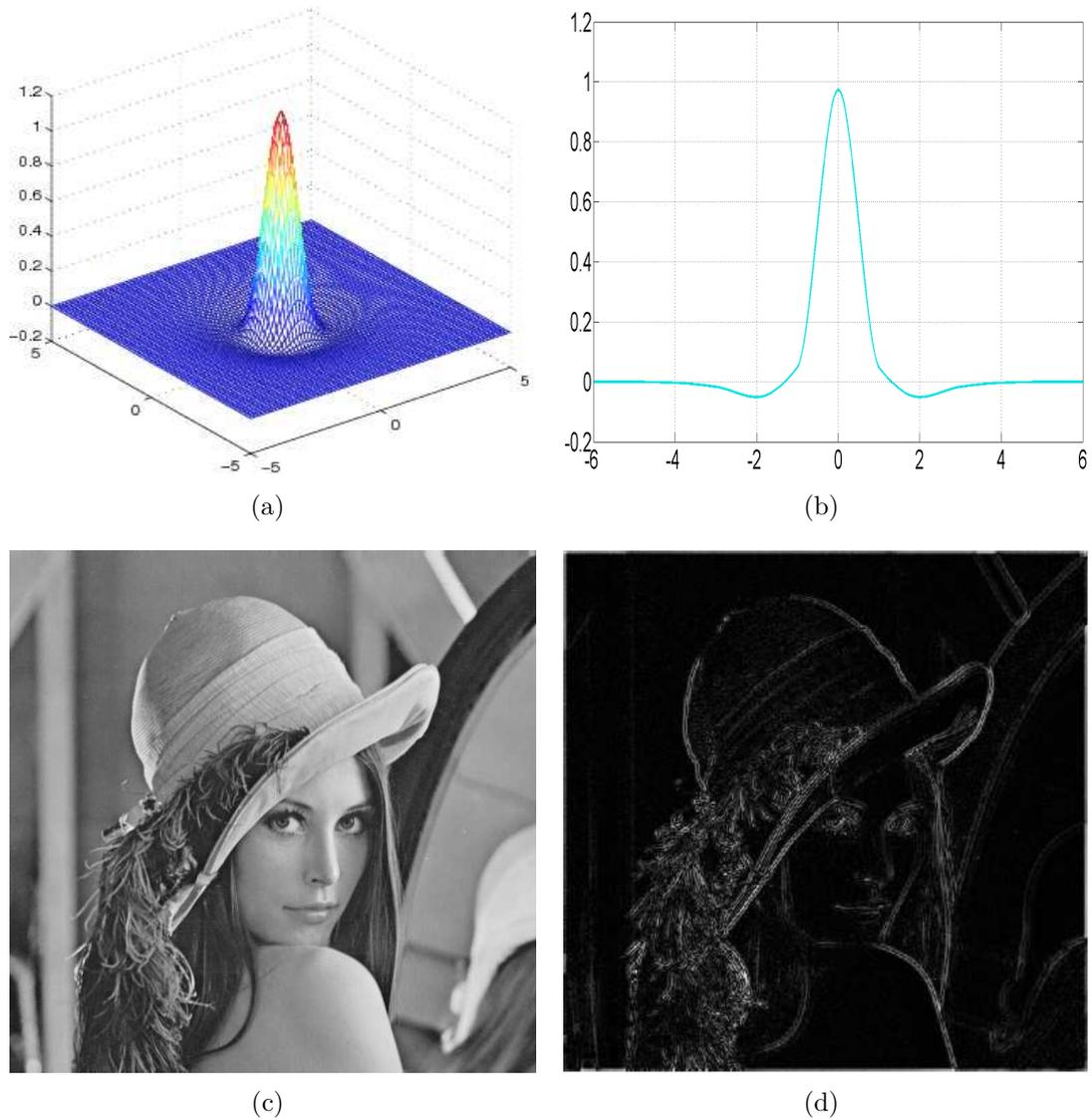
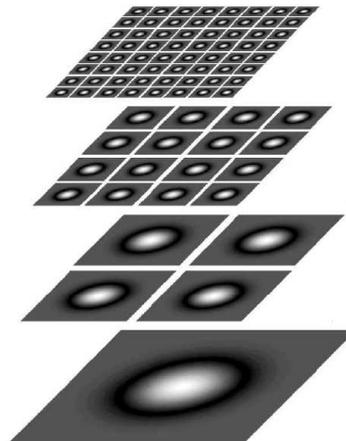


Figure 2.3: Behavior of the DoG filter when applied to Lena with biologically plausible parameters. 2.3(a)-2.3(b): the DoG filter with parameters $w_c = w_s = 1$, $\sigma^c = 0.5 \text{ pixel}$, and $\sigma^s = 1.5 \text{ pixel}$. 2.3(c): the classical test image Lena. 2.3(d): the absolute value of the image resulting from the convolution of Lena with the DoG filter in 2.3(a)-2.3(b).

Figure 2.4: Dyadic repartition of the DoG filters considered for the image analysis. Here 4 scales are represented. As we get from the tightest scale (at the top layer) to the largest one (at the bottom layer) the number of filters considered per scale is divided by 4 at each step (from [Van Rullen 2002]).



where $\sigma_{k+1}^c = \frac{1}{2}\sigma_k^c$, $\sigma_{k+1}^s = \frac{1}{2}\sigma_k^s$, and $\sigma_{K-1}^c = 0.5$ pixel. Each DoG_k filter has a size of $(2M_k + 1) \times (2M_k + 1)$ such that M_k is proportional to σ_k^s (We will consider in the rest of this work that $M_k = 3\sigma_k^s$). The authors considered eight possible scales (ie, $K = 8$) for the model. The choice made for the number of scales K is arbitrary but has an incidence on the quality of the image reconstruction as we will show later in this work. These scale-variable cells chop the image spectrum into different subbands. Indeed, Figure 2.5 shows that the DoG filters that model the retina cells cover a large part of the available spectrum.

Remark The presented architecture accounts, to some extent, for the actual retina topology. The cells density and scale are indeed inversely proportional. Though the authors in [Van Rullen 2001b] do not claim biological plausibility. Such an architecture is not biologically plausible for mainly two reasons. First, in [Hammond 1974] experiments made on cats retina showed that the largest receptive fields of the retina cells are, in average, only two to three times wider than the tightest ones. Whereas in the model presented the scaling factor goes up to 128 from the Γ_7 to Γ_0 . Second, for a given scale, ganglion cells are not regularly dispatched over the retina. In fact observations made on apes [Lee 1998] and cats [Hammond 1974] showed that large receptive field cells are more dense in the periphery of the retina, while tight receptive field cells are more dense in its center. Baring in mind these facts, the model is still strongly inspired from the mammals retina and could be employed to

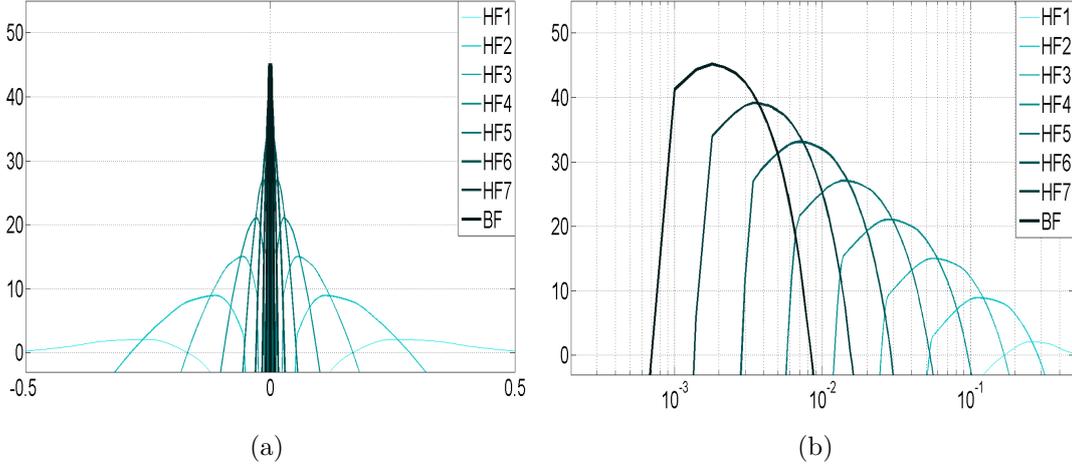


Figure 2.5: 2.5(a): Spectra of the DoG filters. The abscissa represents the frequencies ranging from (-0.5) to (0.5) sample per pixel. The ordinate axis represents the different DoG_k filters gain in dB. 2.5(b): Half of the spectrum in 2.5(a) with the abscissa having a logarithmic step. Frequencies are ranging from 0 to (0.5) sample per pixel.

devise bio-inspired coders for images.

2.3.2 Image analysis

Having the specification of the retina cells behavior and the architecture described in Section 2.3.1, the authors in [Van Rullen 2001b] had also to define the measure of excitation of a given retina cell. To measure this degree of excitation, one computes the convolution of the stimulus image f by the DoG_k filter corresponding to the cell considered. Yet each layer k in the dyadic grid Γ is undersampled with a pace of 2^{K-k-1} pixels with an original offset of $\lfloor 2^{K-k-2} \rfloor$ pixels, where $\lfloor \cdot \rfloor$ is the floor operator. Having this, we define the function u_k , such that the resulting c_{kij} coefficients are computed at the locations $(u_k(i), u_k(j))$ as follows:

$$u_k(i) = \lfloor 2^{K-k-2} \rfloor + 2^{K-k-1}i, \forall k \in \llbracket 0, K-1 \rrbracket. \quad (2.6)$$

u_k can be seen as an undersampling function. We notice that $u_{K-1}(i) = i$, and that the other functions $(u_k)_{k \in \llbracket 0, K-2 \rrbracket}$ are undersampled versions of u_{K-1} . We also define N_k :

$$N_k = \left\lfloor \frac{N - \lfloor N - 2^{K-k-2} \rfloor}{2^{K-k-1}} \right\rfloor, \quad (2.7)$$

such that N_k^2 is the number of cells in Γ_k , the k^{th} level of Γ . We also notice here that $N_{K-1} = N$. Having these definitions, c_{kij} is then computed as follows:

$$c_{kij} = \sum_{x = u_k(i) - M_k}^{u_k(i) + M_k} \sum_{y = u_k(j) - M_k}^{u_k(j) + M_k} DoG_k(u_k(i) - x, u_k(j) - y) f(x, y). \quad (2.8)$$

Thus, we obtain N_s cells responses. If N is a power of 2 then $N_s = \frac{4}{3}N^2 - 1$. Neurons responses are then sorted in the decreasing order of their amplitude, i.e., $|c_{kij}|$. According to the rank order coding principles, this is equivalent to sorting the cells according to the increasing order of their time-to-first-spike. At this level, the retina code for an N^2 -sized stimulus image is the series of N_s sorted quadruplets (k, i, j, c_{kij}) . The retina code can equivalently be determined by the sorting permutation σ (cf. Equation (2.3)) and the list of corresponding coefficients $(c_{\sigma(k,i,j)})_{kij \in \Gamma}$. In accordance with the retina model notations we re-define the permutation σ (cf. Equation (2.3)) by means of the two conditions that follows:

$$\left\{ \begin{array}{l} \sigma \in \mathfrak{S} \left(\bigcup_{k=0}^{K-1} \left(\{k\} \times u_k(\llbracket 0, N_k \rrbracket) \times u_k(\llbracket 0, N_k \rrbracket) \right) \right), \text{ the group of possible} \\ \text{permutations over the locations of the cells in the dyadic grid } \Gamma. \\ |c_{(\sigma(0,0,0))}| > |c_{(\sigma(1,0,0))}| > \dots > |c_{(\sigma(k,i,j))}| > \dots > |c_{(\sigma(K-1, N_{K-1}-1, N_{K-1}-1))}|. \end{array} \right. \quad (2.9)$$

As mentioned in Section 2.1.1, the retina-based code that we consider discards the exact values of the coefficients $(|c_{kij}|)_{kij \in \Gamma}$. Indeed, in [Van Rullen 2001b, Van Rullen 2002, Perrinet 2003, Perrinet 2004], the authors showed that there exists a one-to-one map f^{rank} between the firing rank $\sigma(k, i, j)$ and the amplitude $|c_{kij}|$. Interestingly this map has approximately the same shape across natural images. Thus the authors supposed that the values of $(|c_{kij}|)_{kij \in \Gamma}$ are known a priori at the level of the visual cortex. The authors also supposed that this knowledge has been probably acquired by learning. This assumption was justified by the neurophysiologic experiments conducted in [Delorme 2003]. Furthermore, the loss of the exact values of $(|c_{kij}|)_{kij \in \Gamma}$ diminishes the amount of information that is necessary to encode the stimulus image, and this will be of some interest for the conception of

bio-inspired image coders.

From an implementation point of view, the characteristic one-to-one-map f^{rank} is constructed off-line and stored in a look-up table. For instance, in [Van Rullen 2001b] this look-up table is obtained as the average of the possible f^{rank} maps over a set of natural images. Another example of possible construction of f^{rank} is given in [Perrinet 2003]. The author used a parametric function of the rank r_{kij} corresponding to a given coefficient $(|c_{kij}|)$ that is defined as follows:

$$f^{rank}(r_{kij}) = C (r_{\sigma^{-1}(k,i,j)})^{-\gamma}, \quad (2.10)$$

where C is an arbitrary constant positive value, γ is the parameter of the function that is learned over a set of natural images, and r_{kij} the rank of the considered c_{kij} coefficient. The rank r_{kij} is defined as follows:

$$\begin{cases} (k', i', j') = \sigma((k, i, j)) \\ r_{kij} = 0, & \text{if } k' = 0, \\ r_{kij} = k' N_{k'-1}^2 + i' N_{k'} + j', & \text{if } k' > 0, \end{cases} \quad (2.11)$$

where σ is the permutation defined in Equation (2.9). Figure 2.6 shows some examples of f^{rank} maps that could be used to recover the $(|c_{kij}|)_{kij \in \Gamma}$ coefficients.

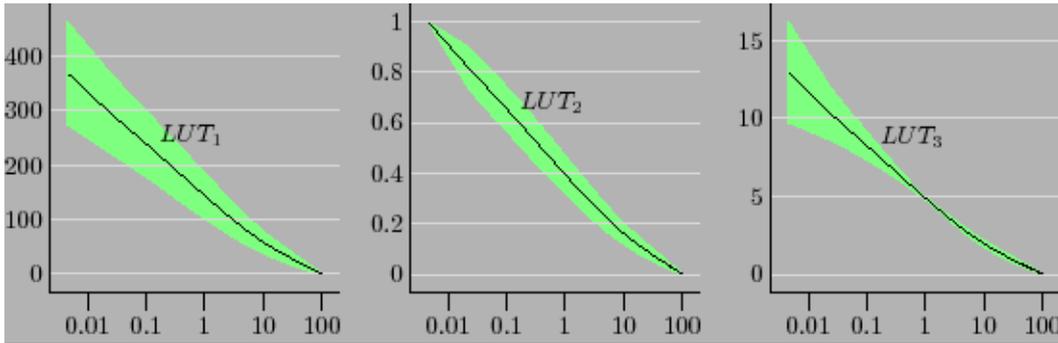


Figure 2.6: Three examples of possible look-up tables to recover the excitation coefficients $(|c_{kij}|)_{kij \in \Gamma}$. The vertical axis represents estimated values of the coefficients $(|c_{kij}|)_{kij \in \Gamma}$. The horizontal axis represents the cell rank normalized between 0 and 100. The green area around the curve represents the standard deviation over the set of tested images (from [Perrinet 2003]).

As the look-up table infers only the absolute values of the $(|c_{kij}|)_{kij \in \Gamma}$

coefficients, we obviously need to store the series $(\text{sign}(c_{kij}))$ to recover the exact values of the coefficients lost. At the end of this stage, this retina model described generates a sorted series of N_s quadruplets $(k, i, j, \text{sign}(c_{\sigma^{-1}(k,i,j)}))_{kij \in \Gamma}$ of retina cells locations and their corresponding activation signs or simply the N_s sorted couples $(r_{kij}, \text{sign}(c_{\sigma^{-1}(k,i,j)}))_{kij \in \Gamma}$. The retina code can equivalently be determined by the sorting permutation σ (cf. Equation (2.9)) and the list of corresponding activation signs $(\text{sign}(c_{\sigma^{-1}(k,i,j)}))_{kij \in \Gamma}$. The output of this retina model can be referred to as the *rank profile* of a given stimulus image.

We present in the next section the specification of the stimulus synthesis algorithm in [Van Rullen 2001b]. The authors aim is to reconstruct the original stimulus with no other data than its rank profile, ie. the series of the N_s sorted quadruplets $(k, i, j, \text{sign}(c_{\sigma^{-1}(k,i,j)}))_{kij \in \Gamma}$ or equivalently the N_s sorted couples $(r_{kij}, \text{sign}(c_{\sigma^{-1}(k,i,j)}))_{kij \in \Gamma}$.

2.3.3 Image synthesis

The retina model in [Van Rullen 2001b] is provided with a decoding algorithm. This allows the model users to recover the stimulus image starting from the retina output consisting in a rank profile. The decoding capability is a major advantage as it enables the conception of an image coder/decoder as discussed later in this work. The specification of the retina model decoder is simple and we detail it in the following.

First let us suppose that the exact values of the analysis coefficients $(c_{kij})_{kij \in \Gamma}$ are known. The authors in [Van Rullen 2001b] and later implementations assumed that all the DoG filters in the dyadic grid Γ are orthonormal. Then, having a set of N_s coefficients, the reconstruction estimate \tilde{f}_{N_s} of the stimulus image f is obtained as follows:

$$\tilde{f}_{N_s}(x, y) = \sum_{r_{kij}=0}^{N_s-1} c_{kij} \text{DoG}_k(u_k(i) - x, u_k(j) - y), \quad (2.12)$$

where r_{kij} is the rank of the considered c_{kij} coefficient according to the sorting permutation σ (cf. Equation (2.9)) as defined in Equation (2.11).

Now let us reconsider the estimation of \tilde{f}_{N_s} bearing in mind that the only data coding a stimulus is its rank profile, ie. discarding the exact values of the analysis coefficients. Recovering c_{kij} by means of the look-up table f^{rank} and replacing it in (2.12) the stimulus reconstruction estimate \tilde{f}_{N_s} becomes:

$$\tilde{f}_{N_s}(x, y) = \sum_{r_{kij}=0}^{N_s-1} \text{sign}(c_{kij}) f^{rank}(r_{kij}) \text{DoG}_k(u_k(i) - x, u_k(j) - y).$$

This formula allows progressive reconstruction as one can stop at a given maximum rank value, N_s . Thereby $\tilde{f}_0, \tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_{N_s}$ are different but increasingly accurate reconstruction of the stimulus image. An example of progressive reconstruction is shown in Figure 2.7. This property allows the coder to be scalable.

It is to be noted also that the reconstruction evolution maps actual retina behavior, as low frequencies are first transmitted, then details progressively added.

2.4 Discussion

As discussed in the previous chapter, there exists a variety of coding metrics that can be measured over the spikes population, but no clear evidence establishes which one is employed by the retina to encode the visual stimuli. Among these metrics, we focused in this chapter on the rank order coding which is our main working hypothesis. To that end, we described the guiding principles and the biological evidence behind it. We attached a particular attention to the introduction of a mathematical formalism for the concepts used which might be lacking in the literature. Then, we detailed the specification of a retina model introduced in [Van Rullen 2001b] to test the the rank order coding for image categorization applications. This model supports the analysis and the synthesis of images, so that we can use it as a basis for a lossy image coder/decoder. However, considering the synthesis stage, we note that the orthonormality assumption of the DoG filters is false and leads to approximations in the reconstruction. We tackle this issue in the next chapter and resolve it in an original fashion by using the frames theory.

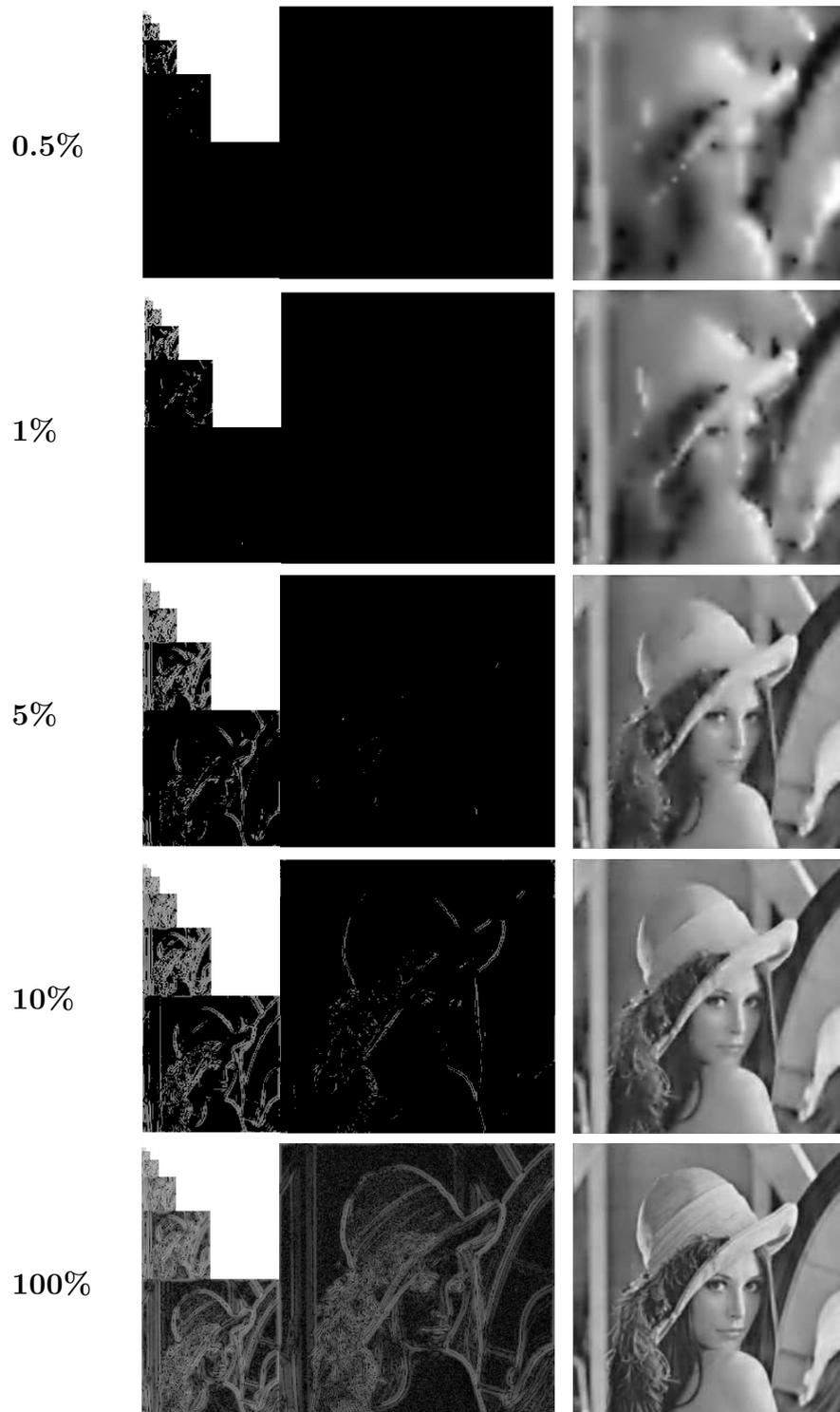


Figure 2.7: Progressive stimulus synthesis of Lena (cf. Figure 2.3(c)) using the retina model in [Van Rullen 2001b]. Left column: The percentage of coefficients taken into account in the dyadic grid Γ . Center column: The coefficients taken into account in the dyadic grid Γ . All coefficients above the considered rank are discarded and represented in black. Right column: The reconstruction obtained with the considered coefficients. As the percentage of spikes decoded increases, the visual quality of the coded/decoded images increases.

FRAMES FOR THE EXACT INVERSION OF THE RANK ORDER CODER

Contents

| | | |
|------------|---|-----------|
| 3.1 | The ROC limitations | 56 |
| 3.1.1 | Matrix notations for ROC | 56 |
| 3.1.2 | Limitations and straightforward solution | 60 |
| 3.2 | Inverting the retina model | 63 |
| 3.2.1 | Introduction of a low-pass scaling function | 65 |
| 3.2.2 | The DoG transform is a frame operator | 66 |
| 3.2.3 | Synthesis using the dual DoG frame | 69 |
| 3.3 | Results | 74 |
| 3.4 | Discussion | 75 |

keywords

Rank order code, bio-inspired image coding, frame theory, dual frames, scalability, out-of-core algorithms.

Overview

We revisit in this chapter the retina model specified in [Van Rullen 2001b] by proposing an original and exact synthesis procedure for it. The solu-

tion that we propose relies on the frames theory. Our contribution encompasses a theoretical and a technical aspect. We published these results in [Masmoudi 2012b, Masmoudi 2012a]

Contribution

Our contribution is threefold.

1. We add to the original retinal filter bank an adequate scaling function that enables the invertibility of the transform.
2. We prove mathematically that the filter bank that we augmented with a scaling function is a frame. Then, we propose the computation of dual frame to eliminate the reconstruction errors yielded by the classical synthesis.
3. We solve the technical issue related to memory overhead when computing the dual frame with a novel recursive out-of-core algorithm.

Organization

This chapter is organized into four sections:

1. Section 3.1 shows the limitations of the retina model specified in the previous chapter.
2. Section 3.2 gives the specification of an exact decoding algorithm through the construction of a dual frame. The solution encompasses the addition of a low pass scaling function (Section 3.2.1), the demonstration that the analyzing filter bank is a frame (Section 3.2.2) and the out-of-core construction of the dual frame (Section 3.2.3).
3. Section 3.3 summarizes the results and shows the gain obtained by our solution when compared to the classical decoder in terms of reconstruction quality at equal rate.
4. Section 3.4 discusses our approach with regard to state of the art.

The person who can combine frames of reference and draw connections between ostensibly unrelated points of view is likely to be the one who makes the creative breakthrough.

Denise Shekerjian

Our goal in this chapter is to revisit the retina model specified in [Van Rullen 2001b] by proposing an original and exact synthesis procedure for it. This model was designed to support the rank order coding (ROC) which is based on the hypothesis that the retina represents the visual stimulus by the order in which its cells are activated. The retina model under study encompasses an analyzing (or coding) stage and a synthesizing (or decoding) one. It was implemented to accomplish recognition tasks with a manner that is inspired from the retina behavior. The classical model analysis/synthesis procedure was detailed in the last chapter. It involves: (i) A model of the stimulus transform in the retina consisting in a DoG filter bank analysis; (ii) A sorting stage of the filters according to their activation degree; (iii) A straightforward decoding procedure that consists in a weighted sum of the most activated filters. Focusing on this last stage, it appears that the decoding procedure employed yields reconstruction errors that limit the model Rate/Quality performances when used as an image codec. Attempts made in the literature to overcome this issue are time consuming and alter the coding procedure or are infeasible for standard size images and lacking mathematical support. Here we solve this problem in an original fashion by using the *frames theory*, where a frame of a vector space designates an extension for the notion of basis. Our contribution is threefold. (i) We add to the original retinal filter bank an adequate scaling function, and (ii) we give an original mathematical demonstration that the filter bank thus defined is a frame. Then we propose an algorithm that guarantees an errorless reconstruction of the stimulus through the computation of a so-called *dual frame*; (iii) We solve the technical issue related to memory overhead that prevented the use of frames for high dimension spaces, with a novel out-of-core algorithm that computes the dual frame. Furthermore, the framework presented here can be extended to several models of the visual cortical areas using redundant representations.

This chapter is organized as follows: In Section 3.1, we discuss the coding capabilities of the bio-inspired retina model under study and show the limitations of it. Then in Section 3.2 we define an exact decoding scheme through the construction of a dual frame. Finally in Section 3.3 and Section 3.4, we summarize and discuss the results and show the gain that we obtain in terms of Rate/Quality tradeoff.

3.1 The retina model behind the ROC: Limitations and a first imperfect solution

We discuss in this section the behavior of the retina model under study when used as an image coder/decoder. To this end, we will divide the section into two parts. First, in Section 3.1.1, we give a brief reminder of the retina model specification and introduce a matrix-based formalism for the analysis and synthesis procedures of it. This formalism simplifies the expression of the analysis as well as the synthesis procedures and will be necessary for the subsequent sections. Second, in Section 3.1.2, we show the strong limitations of the retina model when used as an image coder/decoder. We then give a first straightforward solution to the issues raised and test the ability of the model to code and decode stimuli images with sufficient fidelity.

3.1.1 Matrix notations for ROC

As detailed in Section 2.3.2, the retina cells are modelled by DoG filters. These cells are arranged in a dyadic grid Γ of K layers to sweep all the stimulus spectrum as shown in Figure 3.1(b) [Van Rullen 2001b, Perrinet 2004, Masmoudi 2010d]. As in the retina topology, the cells density and scale are inversely proportional. So that, each layer $0 \leq k < K$ in the grid Γ , is tiled with filtering cells, denoted by DoG_k , having a scale k and generating a transform subband B_k . We recall the definition of the filter DoG_k (cf. Equation (2.5)):

$$DoG_k(x, y) = w^c G_{\sigma_k^c}(x, y) - w^s G_{\sigma_k^s}(x, y),$$

where $\sigma_{k+1}^c = \frac{1}{2}\sigma_k^c$ and $\sigma_{k+1}^s = \frac{1}{2}\sigma_k^s$. Each DoG_k filter has a size of $(2M_k + 1)^2$, with $M_k = 3\sigma_k^s$. The biologically plausible parameters chosen are $w^c = w^s = 1$, $\sigma_k^c = \frac{1}{3}\sigma_k^s \forall k$, and $\sigma_{K-1}^c = 0.5$ pixel. Then the activation coefficient c_{kij} of each cell in the layer Γ_k is computed at the location $(u_k(i), u_k(j))$ (cf. Equation (2.6)). The coefficient c_{kij} is the result of the convolution of the original image f by the corresponding DoG_k filter. We recall the definition of c_{kij} (cf. Equation (2.8)):

$$c_{kij} = \sum_{x=u_k(i)-M_k}^{u_k(i)+M_k} \sum_{y=u_k(j)-M_k}^{u_k(j)+M_k} DoG_k(u_k(i)-x, u_k(j)-y) f(x, y).$$

This architecture is similar to a Laplacian pyramid [Burt 1983]. Thus, if f is an N^2 -sized image and if N is a power of 2, this transform generates a vector c of $(\frac{4}{3}N^2 - 1)$ coefficients c_{kij} . Unlike the implementations in [Van Rullen 2001b, Perrinet 2004, Masmoudi 2010d], we use, in the rest of this work, a matrix Φ to compute the DoG transform through the modelled retina. The circular convolution can indeed be modelled by a Toeplitz matrix. The lines of Φ are the different analyzing DoG_k filters shifted to map the different $(u_k(i), u_k(j))$ locations. Baring in mind that u_k is a dyadic undersampling function, the construction of Φ yields an “undersampled Toeplitz-bloc” sparse matrix. Each bloc in the matrix is 4 times less dense than the subsequent one as shown in Figure 3.2(a). The DoG transform is then outlined in the following simple and concise equation:

$$c = \Phi f. \tag{3.1}$$

Such an implementation allows fast computation of this multi-scale DoG transform through sparse matrix specific algorithms [Golub 1996]. This will in addition help us to construct an exact inversion algorithm for this retina model as we will see in the next section. An example of such a transform performed on the *cameraman* test image is shown in Figure 3.1.

As mentioned earlier in this work, the retina model in [Van Rullen 2001b] specifies also a decoding algorithm. We recall that if the decoder is provided with the exact values of the N_s analysis coefficients in c , the straightforward reconstruction \tilde{f}_{N_s} of the stimulus image f is obtained by the following formula

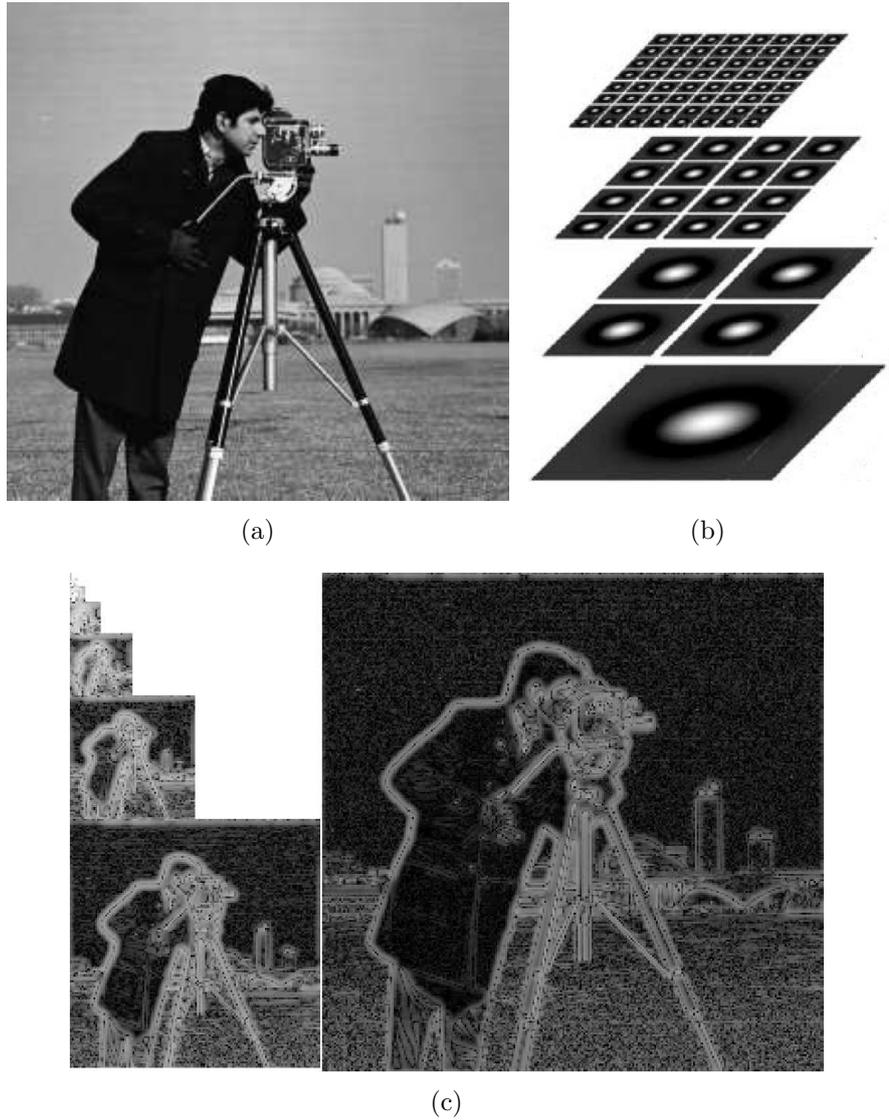


Figure 3.1: 3.1(a): The *cameraman* test image. The image size is 257×257 pixels. 3.1(b): Example of a dyadic grid of DoG's used for the image analysis (from [Van Rullen 2001b]). 3.1(c): The transform result c re-arranged as a pyramid of images with the different generated subbands B_k (here shown in a logarithmic scale).

(cf. Equation (2.12)):

$$\tilde{f}_{N_s}(x, y) = \sum_{r_{kij}=0}^{N_s-1} c_{kij} \text{DoG}_k(u_k(i) - x, u_k(j) - y),$$

where r_{kij} is the rank of the considered c_{kij} coefficient according to the sorting permutation σ (cf. Equation (2.9)) as defined in Equation (2.11). We remind

the reader that σ sorts the c_{kij} coefficients in the decreasing order of their absolute values. Interestingly, this straightforward synthesis amounts to the multiplication of the coefficients vector c by Φ^* the Hermitian transpose of Φ . Figure 3.2(b) shows the template of the synthesis application Φ^* . The reconstruction procedure is then outlined in the following equation:

$$\tilde{f}_{N_s} = \Phi^* c. \quad (3.2)$$

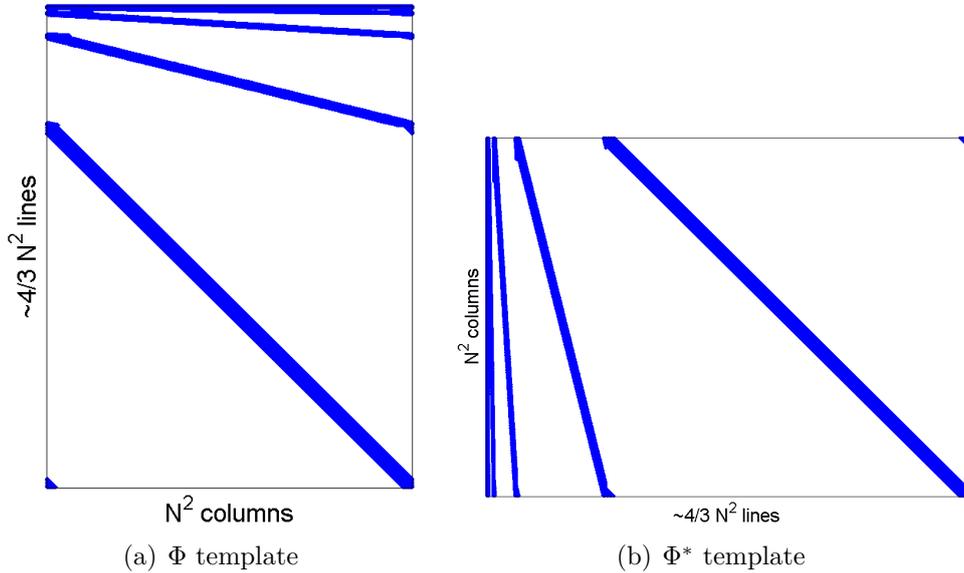


Figure 3.2: Template of the DoG analysis and synthesis matrices, Φ and Φ^* . In this work, Φ and Φ^* are represented as matrices where blue dots correspond to non-zero elements. Note here that Φ and Φ^* are highly sparse matrices.

The Equation (2.12) that we recalled above enables the scalability feature in the image decoder [Masmoudi 2010d]. This feature is of great importance in the context of image coding as a progressive reconstruction can be defined depending on N_s . Yet, $\tilde{f}_0, \tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_{N_s}$ are different but increasingly accurate estimations of the stimulus image. Our aim is to recover this behavior with the matrix-based implementation we just defined. To this end, we must threshold the coefficients in c to get the vector c^m , such that \tilde{f}_m is the image of c^m by the reconstruction function. The coefficients c_i^m of c^m are the m most valuable

coefficients in c . Thus, the vector c^m is defined as follows:

$$\begin{cases} p_{kij} = k N_k^2 + i N_k + j, \\ c_{p_{kij}}^m = c_{kij}, \text{ if } r_{kij} \leq m, \\ c_{p_{kij}}^m = 0, \text{ if } r_{kij} > m, \end{cases} \quad (3.3)$$

where r_{kij} is the rank of the considered coefficient c_{kij} as defined in Equation (2.11). So that the progressive reconstruction formula becomes:

$$\tilde{f}_m = \Phi^* c^m. \quad (3.4)$$

The matrix-based implementation that we defined in this section is mathematically equivalent to the original implementation specified in [Van Rullen 2001b]. If one wants to discard the exact values of the coefficients in c^m , a look-up-table f^{rank} can be used as described in Section 2.3.3. The look-up table infers the absolute values of the coefficients, and the transmitted series $(\text{sign}(c_{r_{kij}}))_{0 \leq r_{kij} < m}$ enables the recovery of the exact values of the coefficients. Obviously Equation (3.4) still holds when using the look-up-table to guess the values of c_i^m .

We suppose in the next section that the decoder is provided with a perfect look-up-table, so that no prediction error is introduced in the synthesis procedure. We then experience, under this condition, the ability of the retina model to code and decode a stimulus image with sufficient fidelity.

3.1.2 Limitations of the retina model and a first straightforward solution to code images

The retina model under study, as well as most of the retina models available in the literature, was designed to accomplish fast categorization tasks. Our aim in this work is to design an image coder/decoder based on the retina behavior. So that, the coded/decoded image \tilde{f}_m defined in Equation (3.4) have to match as accurately as possible the original image f .

To this end, we experience the result of the coding/decoding procedure in the optimal case when all the retina cells have fired, ie. $m = N_s$. Furthermore,

we suppose that the decoder is provided with a perfect look-up-table f^{rank} , so that the exact values of the coefficients $c_i^{N_s}$ are known. Unfortunately, the mean squared error (MSE) of the reconstruction $\|f - \tilde{f}_{N_s}\|_2$ is poor. This leads to a PSNR quality of 1.9 dB if we consider the test image Lena for f . Thus, the retina model, as it is, is unsuitable for image coding applications. We can analyze the reasons through the result shown in Figure 3.3(a). Figure 3.3(a) compares the grayscale histograms of f and the corresponding \tilde{f}_{N_s} . We observe that the histograms resemble each other in shape. However, we also notice that the dynamic range and the energy values of f are completely lost in its estimation \tilde{f}_{N_s} . For these reasons we cannot use this retina model as such in an image coder/decoder. Note that we considered the test image Lena, but the observations that we make still hold for any other image. We give in the following the main reasons behind this erratic behavior and propose a first solution.

The first issue behind the errors in the reconstruction \tilde{f}_{N_s} is that the authors in [Van Rullen 2001b] supposed that the DoG filters used for the transform of the stimulus image are orthonormal. The DoG vectors used for the analysis could then be used for the synthesis. As mentioned in [Bhattacharya 2007], this assertion is false. Yet, we can easily verify that any pairwise scalar product of different DoG filters is small but not nil, so that we have:

$$\left\{ \begin{array}{l} \langle DoG_k(u_k(i) - \cdot, u_k(j) - \cdot), DoG_k(u_k(i) - \cdot, u_k(j) - \cdot) \rangle = 1, \\ \forall (k, i, j) \in \left(\cup_{k=0}^{K-1} \left(\{k\} \times u_k(\llbracket 0, N_k \rrbracket) \times u_k(\llbracket 0, N_k \rrbracket) \right) \right) \\ \langle DoG_k(u_k(i) - \cdot, u_k(j) - \cdot), DoG_{k'}(u_{k'}(i') - \cdot, u_{k'}(j') - \cdot) \rangle = \varepsilon, \\ \forall (k, i, j) \neq (k', i', j') \in \left(\cup_{k=0}^{K-1} \left(\{k\} \times u_k(\llbracket 0, N_k \rrbracket) \times u_k(\llbracket 0, N_k \rrbracket) \right) \right) \\ \text{with } 0 < \varepsilon \ll 1 \end{array} \right. \quad (3.5)$$

Another issue is that the number of DoG filters used to analyze an N^2 -sized image is equal to $(\frac{4}{3}N^2 - 1)$ and thus exceeds the space dimension of the stimulus image. Thus the DoG filters do not form a basis. As a consequence, the reconstruction estimation \tilde{f}_{N_s} could not reach the expected optimal

reconstruction f if we use the formula in Equation (3.4).

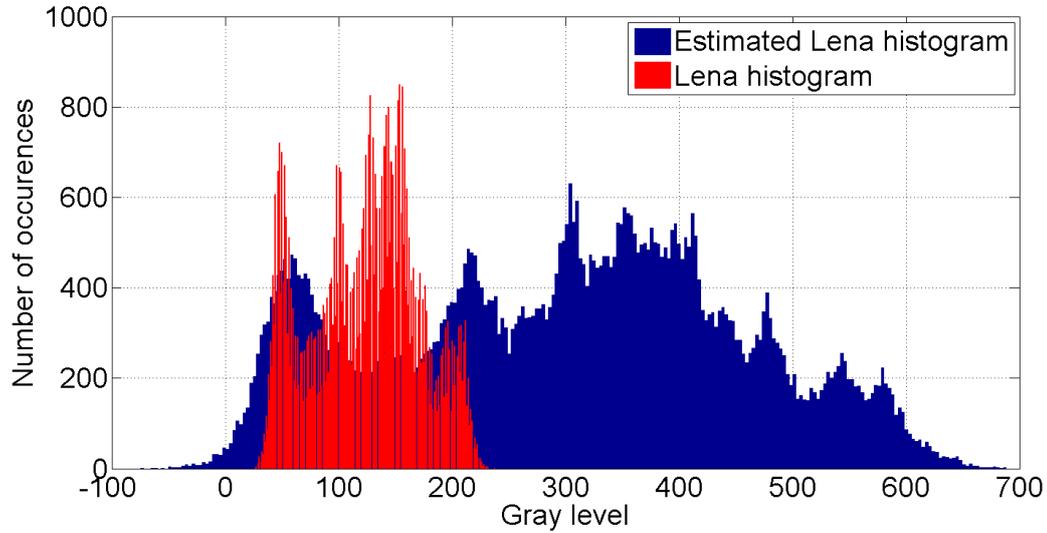
A first solution that might be employed is to equalize the two first moments of f and its estimation \tilde{f}_m . In the following, we denote the image obtained from \tilde{f}_m after the moment equalization by \tilde{f}_m^{corr} . This method is implemented easily and resolves the problems of energy and dynamic range conservation. Indeed, the first moment of f is a measure of its mean and the second moment a measure of its energy. Furthermore, from a coding point of view, adding two scalar values in the data transmitted to the decoder has a negligible bit-cost. We can outline the process employed in the following equations.

$$\left\{ \begin{array}{l} \mu_f = \mathbb{E}(f), \\ \sigma_f = \mathbb{E}\left((f - \mu_f)^2\right), \\ \mu_{\tilde{f}_m} = \mathbb{E}(f), \\ \sigma_{\tilde{f}_m} = \mathbb{E}\left((f - \mu_f)^2\right), \\ \tilde{f}_m^{corr} = \sigma_f \frac{\tilde{f}_m - \mu_{\tilde{f}_m}}{\sigma_{\tilde{f}_m}} + \mu_f, \end{array} \right. \quad (3.6)$$

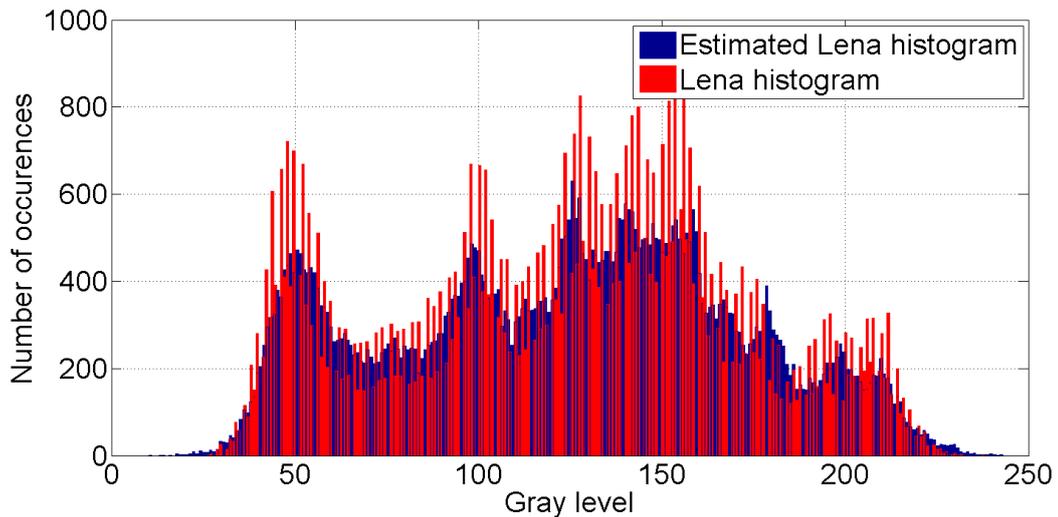
where $\mathbb{E}(f)$ is the mathematical expectation of f . Figure 3.3(b) shows the comparison between the histogram of f and the obtained histogram of $\tilde{f}_{N_s}^{corr}$. We notice that the problem of histogram matching is mostly resolved, but also that a slight residual distortion remains between the two histograms.

An example of such a reconstruction is given in Figure 3.4, with all the retina cells taken into account. Figure 3.4 also shows that the retina model decoding procedure, though giving a good approximation of the stimulus, is still inaccurate. In this example, reconstruction quality is evaluated to 26.8 dB/27dB of PSNR for *Lena/cameraman*.

We tested this slightly modified retina model as part of an image coder decoder in [Masmoudi 2010d]. The results obtained are encouraging but the residual errors still limit the rate/quality performances compared to the JPEG standards. We give a mathematical and exact solution to the issues of orthonormality and over-dimensionality of the DoG transform in the following section.



(a) The original retina model



(b) The model with corrected moments

Figure 3.3: Comparison between the histograms of a given stimulus image f and the corresponding coded/decoded image. We considered in this case the test image Lena. Figure 3.3(a) shows the histograms of f and the corresponding \tilde{f}_{N_s} . The dynamic range and the energy values of f are completely lost in its estimation \tilde{f}_{N_s} . Figure 3.3(b) shows the histograms of the test image Lena f and the corresponding $\tilde{f}_{N_s}^{corr}$. The dynamic range and the energy values of f are recovered in $\tilde{f}_{N_s}^{corr}$ through the equalization of their first two moments.

3.2 Inverting the retina model

In this section, we define an original and exact image reconstruction algorithm starting from the ROC. First, we introduce in Section 3.2.1 a low-pass scaling



Figure 3.4: Result of the decoding procedure with the equalized moments using the totality of the retina cells. 3.4(a)-3.4(c): Reconstructed images *Lena* and *cameraman*. The PSNR quality measure of $\tilde{f}_{N_s}^{corr}$ yields 26.8 dB (for *Lena*)/ 27 dB (for *cameraman*). 3.4(d)-3.4(b): Error image shown in a logarithmic scale: high frequencies are the ones that are the most affected by this approach.

function in the analyzing filter bank. This modification will be shown to be necessary for the transform invertibility. Then, in Section 3.2.2, we use the matrix-based formalism given above to prove that our filter bank is a frame. Finally, in Section 3.2.3, we show the exact reconstruction results using the dual frame and introduce an out-of-core algorithm to construct it.

3.2.1 Introduction of a low-pass scaling function

We introduce a low-pass scaling function in the filter bank used for image analysis. This modification does not alter the ROC coder architecture and has both a mathematical and a biological justification.

Indeed, the Fourier transform of a Gaussian is another Gaussian, so that $\mathcal{F}(DoG_k)$ is a difference of Gaussians. Therefore, with $w^c = w^s = 1$ (cf. Equation (2.5)), we have:

$$\mathcal{F}(DoG_k) = 2\pi (\sigma_k^c)^2 G_{(\sigma_k^c)^{-1}} - 2\pi (\sigma_k^s)^2 G_{(\sigma_k^s)^{-1}}. \quad (3.7)$$

We can easily verify that the central Fourier coefficient $\mathcal{F}(DoG_k)(u_0(0), u_0(0)) = 0 \quad \forall k$, and that $\mathcal{F}(DoG_k)(i, j) > 0 \quad \forall (i, j) \neq (u_0(0), u_0(0))$.

In order to cover up the centre of the spectrum, we propose to replace the DoG_0 filter, with no change in the notation, by a Gaussian low-pass scaling function consisting in its central component, such that:

$$DoG_0(x, y) = w^c G_{\sigma_0^c}(x, y). \quad (3.8)$$

Figures 3.5(a) and 3.5(b) show the spectrum partitioning with the different DoG_k filters ($k \geq 1$, in blue) and the spectrum of the new scaling function DoG_0 (in red dashed line) which covers low frequencies. With no scaling function, all constant images would be mapped into the null image 0 and this would make the transform be non-invertible. Here we overcome this problem as the central Fourier coefficient $\mathcal{F}(DoG_0)(u_0(0), u_0(0)) > 0$.

The scaling function introduction is further justified by the actual retina behavior. Indeed, the surround $G_{\sigma_k^s}$ in Equation (2.5) appears progressively across time driving the filter passband from low frequencies to higher ones. So that, the Gaussian scaling function represents the very early state of the retina cells.

Having the DoG_0 scaling function, and in order to define an inverse for the new transform, let us demonstrate that it is a “frame”.

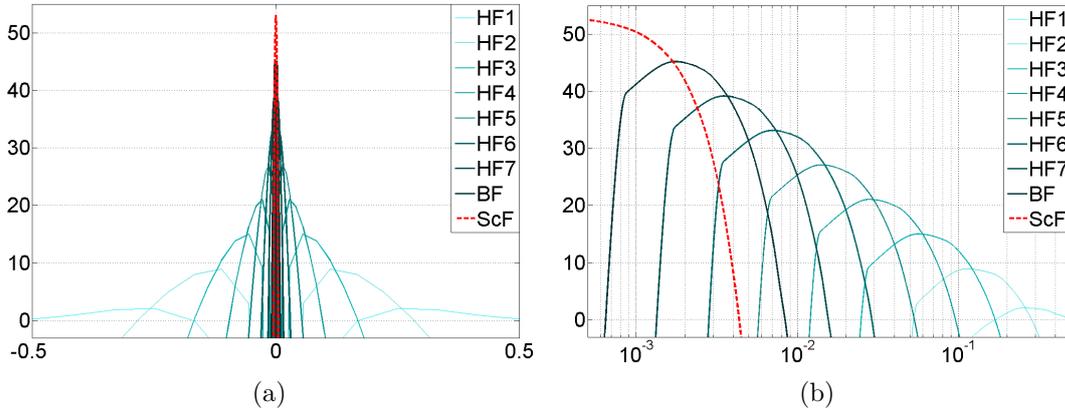


Figure 3.5: 3.5(a): Spectra of the DoG filters. The abscissa represents the frequencies. The ordinate axis represents the different DoG_k filters gain in dB. 3.5(b): Half of the spectrum in 3.5(a) with the abscissa having a logarithmic step. The scaling function DoG_0 is plotted in red dashed line.

3.2.2 The DoG transform is a frame operator

Our aim is to prove that the bio-inspired retina image transform presented, if augmented with a low-pass scaling, amounts to a projection of the input image f onto a frame of a vector space. This proof is crucial to invert the transform exactly as we will show later in this section.

The frame is a generalization of the idea of a basis to sets which may be linearly dependent [Duffin 1952, Kovacevic 2008]. These frames allow a redundant signal representation which, for instance, can be employed for coding with error resilience. By proving the frame nature of this transform, we will be able to achieve an exact reverse transform through the construction of a dual frame.

A set of vectors is a frame if it verifies the so-called “frame condition” [Duffin 1952, Kovacevic 2008] which states that $\exists \beta \geq \alpha > 0$ such that:

$$\alpha \|f\|^2 \leq \sum_{kij \in \Gamma} (c_{kij})^2 \leq \beta \|f\|^2, \forall f. \quad (3.9)$$

Positioning with respect to the state of the art Pyramid architectures are very common in signal processing and involve a wide range of filters [Van Rullen 2001b, Do 2003, Rakshit 1995]. For example, in [Rakshit 1995] the authors proved experimentally that the classical Lapla-

cian pyramid is a frame. However, in our case, we prove that the pyramid introduced in [Van Rullen 2001b] -which is not Laplacian- is a frame. We showed this mathematically through an original demonstration. Also, in [Do 2003] the authors proposed the design of a set of orthogonal vectors inspired from the Laplacian pyramid to conceive a new orthogonal and tight ($\alpha = \beta$ cf. Eq. (3.9)) frame. The filter bank defined from [Van Rullen 2001b] form a frame that is neither orthogonal nor tight.

Proposition 3.2.1. *Let $\Psi(\phi) = \{\phi_{kij}, (k, i, j) \in \Gamma\}$ be the set of vectors ϕ_{kij} , such that $\phi_{kij}(x, y) = DoG_k(u_k(i) - x, u_k(j) - y)$ as defined in Eq. (2.5) for $k > 0$, and in Eq. (3.8) for $k = 0$. Then $\Psi(\phi)$ is a frame of the $N \times N$ images vector space.*

Proof:

Upper bounding: Let us prove the upper bounding in Equation (3.9). We have:

$$\sum_{kij \in \Gamma} (c_{kij})^2 = \sum_{k=0}^{K-1} \|B_k\|^2, \quad (3.10)$$

where B_k is the subband of scale k generated by the image transform with:

$$B_k(i, j) = \sum_{x=u_k(i)-M_k}^{u_k(i)+M_k} \sum_{y=u_k(j)-M_k}^{u_k(j)+M_k} DoG_k(u_k(i) - x, u_k(j) - y) f(x, y).$$

If we denote by U_k the undersampling operator corresponding to the function u_k (cf. Equation (2.6)), we can write the following:

$$B_k = U_k(DoG_k * f). \quad (3.11)$$

Then, we have the following obvious inequalities:

$$\begin{aligned} \|B_k\| &= \|U_k(DoG_k * f)\| \\ &\leq \|U_k(|DoG_k| * |f|)\| \\ &\leq \| |DoG_k| * |f| \| \\ &\leq \|DoG_k\| \|f\|. \end{aligned}$$

Then, we get back to (3.10) and infer the following bounding:

$$\begin{aligned}
\sum_{kij \in \Gamma} (c_{kij})^2 &= \sum_{k=0}^{K-1} \|B_k\|^2 \\
&\leq \left(\sum_{k=0}^{K-1} \|DoG_k\|^2 \right) \|f\|^2 \\
&= \beta \|f\|^2.
\end{aligned} \tag{3.12}$$

Lower bounding: Now, let us demonstrate the lower bounding in Equation (3.9). We start from the fact that:

$$\sum_{k=0}^{K-1} \|B_k\|^2 \geq \|B_{K-1}\|^2 + \|B_0\|^2, \tag{3.13}$$

which amounts to write the following inequalities:

$$\begin{aligned}
\sum_{kij \in \Gamma} (c_{kij})^2 &= \sum_{k=0}^{K-1} \|B_k\|^2 \\
&\geq \|DoG_{K-1} * f\|^2 + \|(DoG_0 * f)(u_0(0), u_0(0))\|^2 \\
&= \|\mathcal{F}(DoG_{K-1}) \mathcal{F}(f)\|^2 + \|\mathcal{F}(DoG_0) \mathcal{F}(f)(u_0(0), u_0(0))\|^2 \\
&= \sum_{i,j=0}^{N-1} \left(\mathcal{F}(DoG_{K-1})(i, j) \mathcal{F}(f)(i, j) \right)^2 \\
&\quad + \|\mathcal{F}(DoG_0)(u_0(0), u_0(0)) \mathcal{F}(f)(u_0(0), u_0(0))\|^2,
\end{aligned}$$

where \mathcal{F} designates the discrete Fourier transform. We know that $\mathcal{F}(DoG_{K-1})(i, j) > 0$, $\forall (i, j) \neq (u_0(0), u_0(0))$ and that $\mathcal{F}(DoG_{K-1})(u_0(0), u_0(0)) = 0$. We also have $\mathcal{F}(DoG_0)(u_0(0), u_0(0)) > 0$. So, if we define a set S_{K-1} by $S_{K-1} = \llbracket 0, N-1 \rrbracket^2 \setminus (u_0(0), u_0(0))$ and α by:

$$\alpha = \min \left\{ \mathcal{F}(DoG_0)^2(u_0(0), u_0(0)), \left\{ \mathcal{F}(DoG_{K-1})^2(i, j), (i, j) \in S_{K-1} \right\} \right\} > 0,$$

then we get the following:

$$\begin{aligned}
& \sum_{i,j=0}^{N-1} \left(\mathcal{F}(DoG_{K-1})(i,j) \mathcal{F}(f)(i,j) \right)^2 \\
& + \left\| \mathcal{F}(DoG_0)(u_0(0), u_0(0)) \mathcal{F}(f)(u_0(0), u_0(0)) \right\|^2 \\
& = \sum_{i,j \in S_{K-1}} \left(\mathcal{F}(DoG_{K-1})(i,j) \mathcal{F}(f)(i,j) \right)^2 \\
& + \left\| \mathcal{F}(DoG_0)(u_0(0), u_0(0)) \mathcal{F}(f)(u_0(0), u_0(0)) \right\|^2 \\
& \geq \alpha \sum_{i,j \in \llbracket 0, N-1 \rrbracket^2} \left(\mathcal{F}(f)(i,j) \right)^2 \\
& = \alpha \|f\|^2,
\end{aligned}$$

and Finally, $\sum_{kij \in \Gamma} (c_{kij})^2 \geq \alpha \|f\|^2$. Thus, the set of DoG filters satisfies the frame condition (3.9).

3.2.3 Synthesis using the dual DoG frame

We introduce in this section a correction means for the reconstruction error in the retina model presented through the frame theory.

The straightforward analysis/synthesis procedure can be outlined in the relation between the input image and the reconstruction estimate:

$$\tilde{f}_{N_s} = \Phi^* \Phi f, \quad (3.14)$$

where Φ^* and Φ are the matrices specified in Equation (3.2) and (3.1). As we already demonstrated that the DoG transform is a frame, $\Phi^* \Phi$ is said to be the frame operator. To have an exact reconstruction of f , one must construct the dual DoG vectors. A preliminary step is to compute $(\Phi^* \Phi)^{-1}$, the inverse frame operator. We then get a corrected reconstruction $f_{N_s}^*$, defined by: $f_{N_s}^* = (\Phi^* \Phi)^{-1} \tilde{f}_{N_s}$. If N_s is the total number of the retina model cells, we have:

$$\begin{aligned}
f_{N_s}^* & = (\Phi^* \Phi)^{-1} \tilde{f}_{N_s} \\
& = (\Phi^* \Phi)^{-1} \Phi^* c \\
& = (\Phi^* \Phi)^{-1} \Phi^* \Phi f \\
f_{N_s}^* & = f.
\end{aligned} \quad (3.15)$$

As made clear through Equation (3.15), the dual vectors are the lines of $(\Phi^*\Phi)^{-1}\Phi^*$. Besides, $(\Phi^*\Phi)$ is a square, definite positive invertible matrix [Kovacevic 2008]. This is a crucial issue as it ensures the exactness of the reverse frame operator.

We compute the reverse frame operator $(\Phi^*\Phi)^{-1}$ and get the results shown in Figure 3.6. The reconstruction obtained by the means of the dual frame operator is accurate and requires only a simple matrix multiplication. In this example, reconstruction quality is evaluated to 295 dB of PSNR.

Dual vectors resemble the DoG analyzing filters. This is obvious as the straightforward image reconstruction \tilde{f}_{N_s} is already close to f , which means that $\Phi^*\Phi$ is close to identity. However, the dual filters lose the symmetry property of the primal ones. An example of dual vectors constructed as the rows of $(\Phi^*\Phi)^{-1}\Phi^*$ is shown in Figure 3.7. Figure 3.7 shows also that the exact reconstruction of f is obtained by a relaxation in the symmetry constraint of the DoG filters

3.2.3.1 The recursive out-of-core blockwise inversion algorithm

Though the mathematical fundamentals underlying this work are simple, the implementation of such a process is a hard problem. In spite of the sparsity of Φ and Φ^* , the frame operator $\Phi^*\Phi$ is an N^4 -sized dense matrix for an N^2 -sized image f . For instance, if $N = 257$, $\Phi^*\Phi$ holds in 16 Gbytes, and 258 Gbytes if $N = 513$. As noticed in (Do & Vetterli 2003): *A key observation is that one should use the dual frame operator for the reconstruction. While this seems a somewhat trivial observation, it has not been used in practice, probably because the usual reconstruction, while suboptimal, is very simple.* Indeed due to its technical difficulty, there is no solution in the literature that computes explicitly the dual of a frame in a general case like ours. In our work, we tackled this technical issue and resolved it with success by designing an original “out-of-core” inversion algorithm.

The frame operator $\Phi^*\Phi$ is constructed block by block, and each bloc is stored separately on disk. The inversion is then performed using a recursive



Figure 3.6: Result of the decoding procedure with the dual DoG frame using the totality of the retina cells. 3.6(a)-3.6(c): Reconstructed image. The PSNR quality measure of $f_{N_s}^*$ yields 295 dB. 3.6(b)-3.6(d): Error image in logarithmic scale. This shows that the reconstruction using the dual frame is very precise.

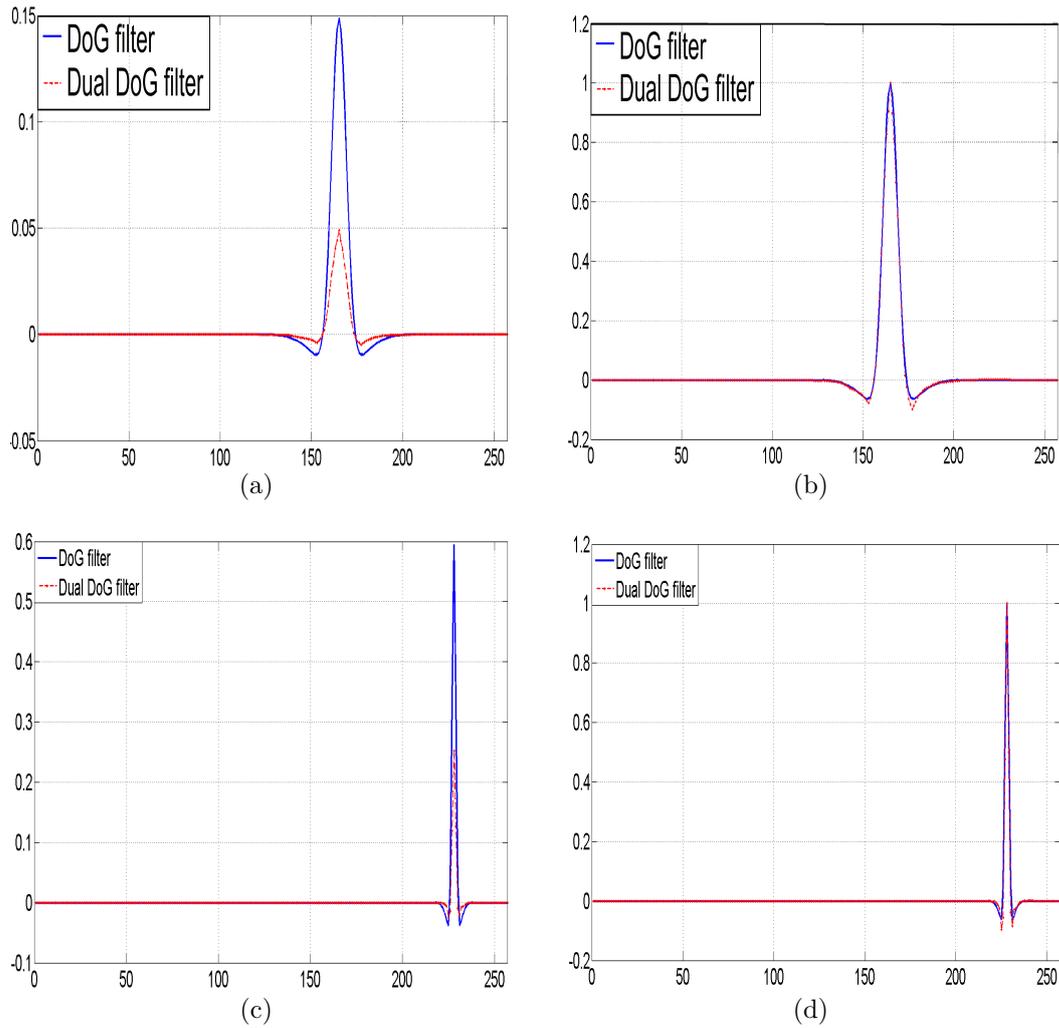


Figure 3.7: 3.7(a)- 3.7(c): Comparison between the DoG filters (in blue line) and their duals (in red dashed line). 3.7(b)- 3.7(d): Same as previous with the highest value of each filter normalized to 1. Though close in shape to the DoG filters, the dual DoG filters are asymmetric.

algorithm that relies on the blockwise matrix inversion formula that follows:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1} B Q^{-1} C A^{-1} & -A^{-1} B Q^{-1} \\ -Q^{-1} C A^{-1} & Q^{-1} \end{pmatrix},$$

where Q is the Schur complement of A , such that:

$$Q = D - C A^{-1} B.$$

Thus, inverting a matrix amounts to the inversion of two matrices that are 4 times smaller. The inversion consists then in subdividing the problem by a factor 4 at each recursion level until we reach a single bloc problem. Obviously, this algorithm requires out-of-core blockwise matrix routines for multiplication, subtraction and addition, that we implemented in a “multi-threaded” fashion to accelerate the computation. A pseudo-code description of this algorithm is available in Appendix A.

Advantages of our approach $(\Phi^* \Phi)$ is a square, definite positive, and invertible matrix [Kovacevic 2008]. Thus $(\Phi^* \Phi)^{-1}$ exists and obviously the exact reverse transform of Φ too. Another advantage of our method is that $(\Phi^* \Phi)$ is well conditioned, with a conditioning number estimated to around 16, so that its inversion is stable. This is a crucial issue as previous work aimed at conceiving the DoG reverse transform tried to invert the original filter bank with no scaling function DoG_0 [Bhattacharya 2007, Bhattacharya 2010]. This is obviously mathematically incorrect as the filter bank thus defined is not a frame and thus its pseudo inverse $(\Phi^* \Phi)^{-1} \Phi^*$ does not exist. The solution proposed by the authors of [Bhattacharya 2007, Bhattacharya 2010] gives only a least squares solution to an ill-conditioned problem. Our method instead is stable. Besides through the out-of-core algorithm that we designed we can invert $(\Phi^* \Phi)$ even for large images whereas authors in [Bhattacharya 2007, Bhattacharya 2010] are restricted to a maximum size of 32×32 . Indeed, we were able to reconstruct 257×257 and 513×513 images through our new approach.

The exactness of our decoding schema is confirmed when applied on several classical test images. For example *cameraman* reconstruction quality increases

from 27 dB with the classical decoder to 296 dB with ours (see Figure 3.9). The same test on *Lena* leads to an increase from 31 to 300 dB of PSNR (see Figure 3.8). We also confirm these results by using quality metrics that are more consistent than PSNR with the human eye perception. Here we show the mean structural similarity measure (SSIM) [Wang 2004b] (an index between 0 and 1) which also confirms the precision of our new decoder with an increase in quality from 0.9 to 1 when all the retina cells have fired (Figures 3.9 and 3.8 captions)

Furthermore, correcting the reconstruction errors using the adequate dual frame does not alter the coding procedure. Indeed, methods introduced in [Perrinet 2004, Bhattacharya 2010] are based on the matching pursuit (MP) algorithm. MP is time consuming and depends on the order in which the “match and update” mechanism is performed. Our method keeps the coding procedure straightforward, multi-threadable and order-independent.

3.3 Results

We experiment our new decoder in the context of scalable image decoding. We reconstruct the test image using an increasing number m of significant coefficients (cf. Equation (3.4)). We then compare the results when using the original DoG filters in Φ and their dual DoG filters in $(\Phi^*\Phi)^{-1}\Phi^*$ for the decoding procedure.

Figures 3.9 and 3.8 show two example results obtained for *cameraman* and *Lena*. In both figures the left column shows the progressive straightforward reconstruction \tilde{f}_{N_s} and the right column shows the corrected progressive reconstruction $f_{N_s}^*$ using the dual frame.

Qualitatively speaking, the high frequencies which were the most altered by the straightforward synthesis are now well rendered. This is obvious even for low rates and we can verify it in the camera contours (Figure 3.9) or in *Lena* hair details (Figure 3.8). This is an important issue as contours and salient points are the most important features used for categorization tasks. Bearing in mind that the retina model under study was first designed for fast categorization, our results become crucial.

Quantitatively speaking, the gain in PSNR is significant for low rates

(around 0.2 dB for both cases) and very high when we consider the totality of the retina cells in the reconstruction (over 260 dB for both cases). Figure 3.10 compares the Rate/quality curves of the two methods and shows the high improvement we obtain. Here the rate is implicitly related to the number of neurons taken into account for the image reconstruction. In all the curves shown the abscissa represents the percentage of the highest responses used for the reconstruction and the ordinate represents the reconstruction quality in terms of PSNR. This figure shows that the PSNR gain grows exponentially with the number of neurons taken into account for the reconstruction. This means that the dual frame correction, though already significant for low rates, becomes extremely important for high rates.

3.4 Discussion

We proposed in this work an original exact decoding procedure for the classical rank order coder defined in [Van Rullen 2001b]. The authors in [Van Rullen 2001b] has then proposed the design a bio-inspired retina model for the image transform and reconstruction. Our contribution encompasses a theoretical and a technical aspect.

Regarding the theoretical aspect, (i) we proved that the bio-inspired transform used to model the retina in [Van Rullen 2001b] is non-invertible as it is, and (ii) we gave an original mathematical proof that this transform if augmented with an adequate scaling function is a frame. We also showed that the scaling function besides its mathematical justification has a biological one. We then defined the corresponding dual frame that is necessary for the exact image reconstruction.

Regarding the technical aspect, we overcame the problem of memory overhead encountered while computing this dual frame. Up to our knowledge, no work in the literature concerned with high dimensionality frame inversion tackled explicitly this problem [Rakshit 1995, Do 2003]. Indeed usual reconstruction algorithms avoid such a calculation by using inaccurate, though very simple, methods. Thus we designed an original recursive out-of-core block-wise algorithm. Our algorithm is general and could be used in a variety of applications requiring a high dimension matrix inversion.

Furthermore, the method presented in this work does not alter the coding procedure and keeps it straightforward unlike the stimulus-dependent MP methods in [Perrinet 2004, Bhattacharya 2010]. In fact, MP-like algorithms require the computation of a specific reconstruction filter bank for each specific image. On the contrary, in our case the dual vectors used for the reconstruction (i) are computed once for all and (ii) are the same for all images. This keeps the decoding procedure committed to the rank order coding philosophy. Yet, the rank ordering supposes that the retina cells are independent and fire asynchronously.

One last major advantage is that our algorithm is multi-threadable. Indeed, there is no possible data hazard in the decoding procedure. Each value of $f_{N_s}^*$ is independent from the others and the concurrent reading in $\Phi^*\Phi$ does not alter the data.

Though it is to be noted that, in some implementations, the rank order decoder inaccuracy is enhanced for a supplemental reason: the inexactness of the look-up-table that might be used to re-generate the transform coefficients c_p . In this work and for a sake of clarity, we considered only the filters overlap as a source of error. Otherwise the reader could not distinguish the part of error due to the filters overlap and the other part that is due to the look-up-table. In our case, the decoder is supposed to be provided with an optimal look-up table. Still the approach presented remains (i) relevant because the inaccuracy of any look-up-table that might be used will affect both the "classical" reconstruction and the "dual frame", (ii) novel through the introduction of the frames theory and (iii) general and thus could be extended to several models of cortical areas using redundant representations.

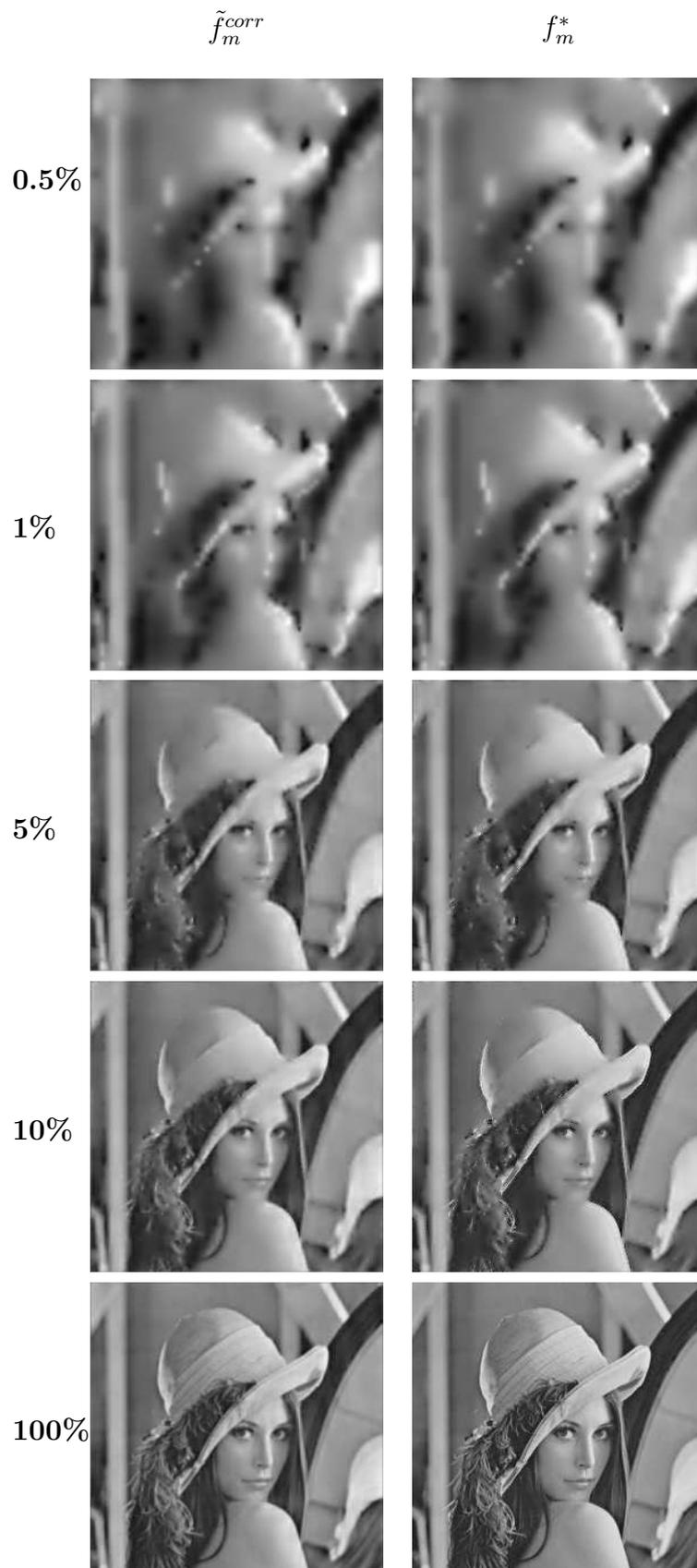


Figure 3.8: Reconstruction of the *Lena* image f using different percentages of significant coefficients. To get this reconstruction coefficients are set to 0 if under a descending threshold. The left column shows the progressive $\tilde{f}_{N_s}^{corr}$ synthesis. The right column shows $f_{N_s}^*$.

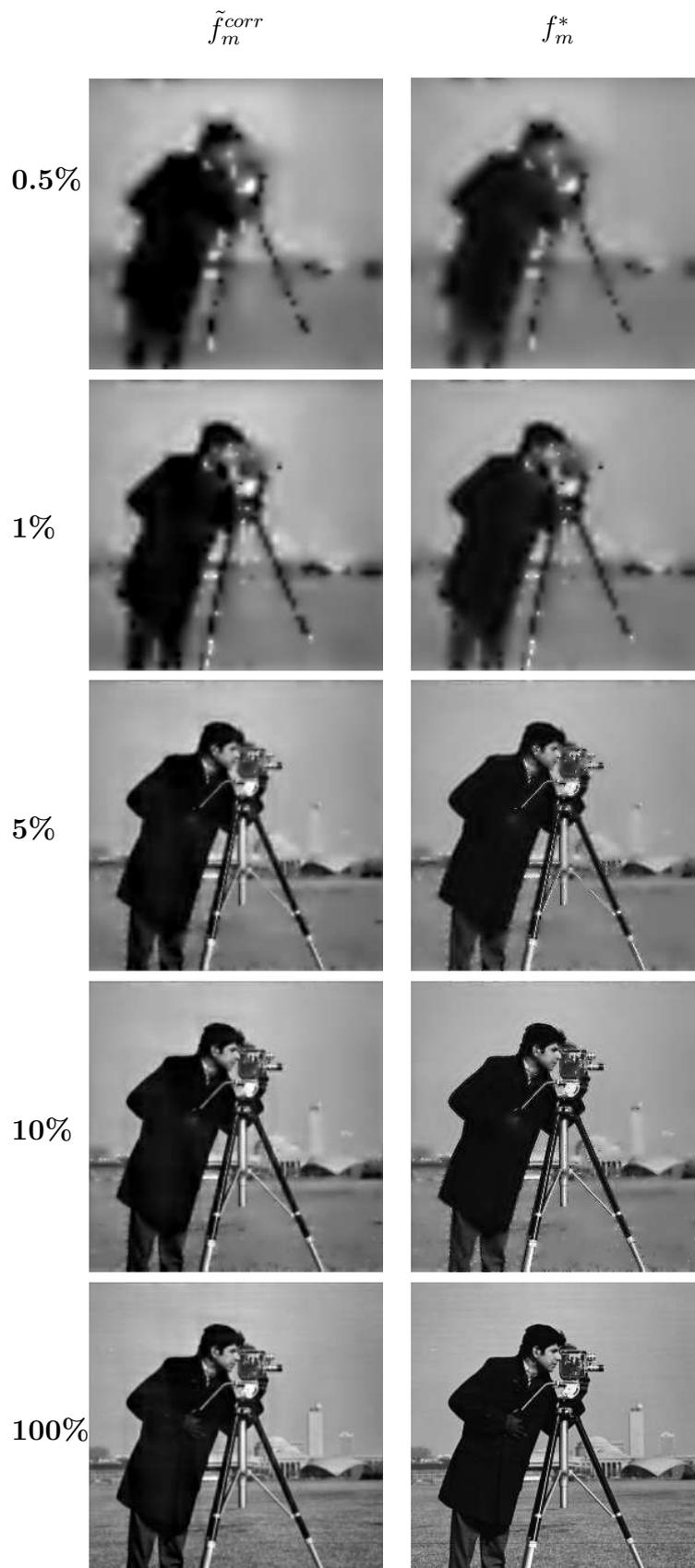


Figure 3.9: Reconstruction of the cameraman image f using different percentages of significant coefficients. The left column shows the progressive \tilde{f}_m^{corr} synthesis. The right column shows f_m^* .

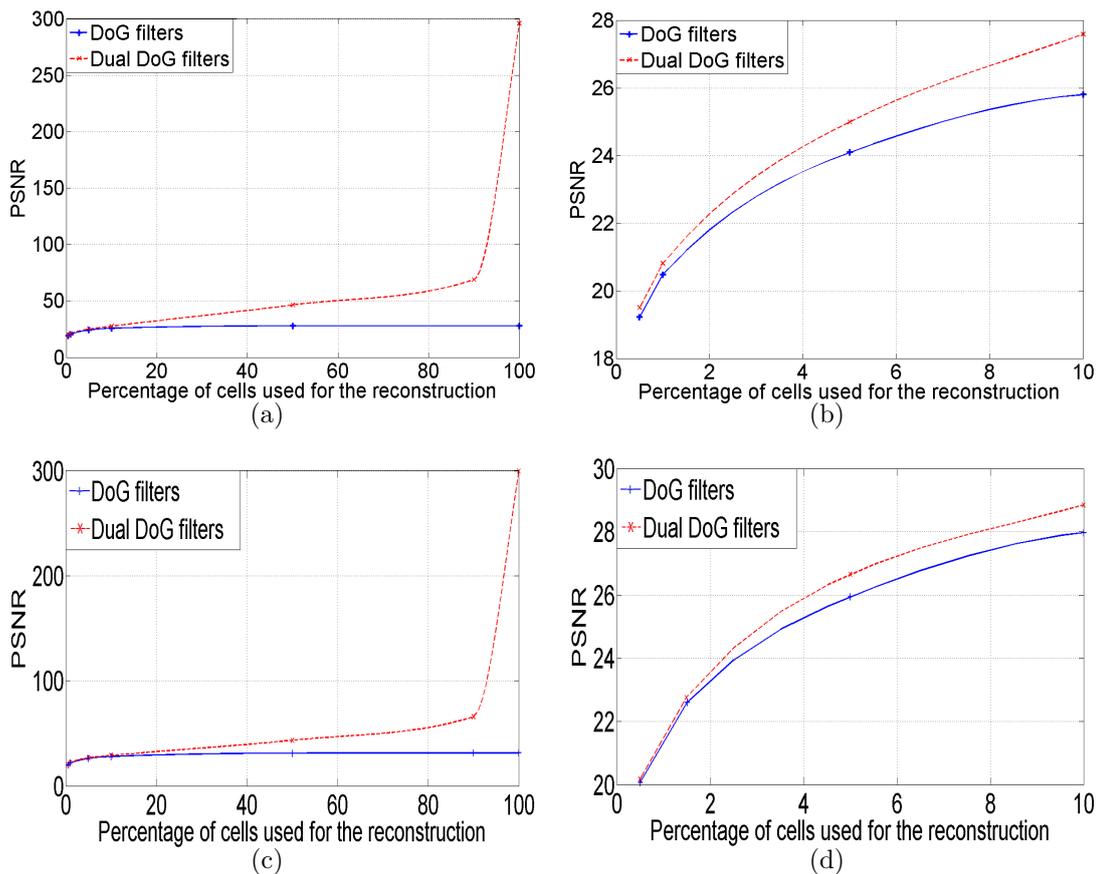


Figure 3.10: PSNR quality of the reconstruction using the DoG filters [Van Rullen 2001b] (in blue solid line) and the dual DoG filters (in red dashed line). Results for *cameraman* are shown in the first line and results for *Lena* are shown in the second line. The abscissa represents the percentage of the highest responses used for the reconstruction. The ordinate represents the reconstruction quality in terms of PSNR. 3.10(a)- 3.10(c): Results shown for percentages between 0% and 100%. 3.10(b)- 3.10(d): Results shown for percentages between 0% and 10%.

A NOVEL BIO-INSPIRED CODING SCHEME BASED ON THE ROC

Contents

| | | |
|------------|---|-----------|
| 4.1 | The coding pathway | 85 |
| 4.1.1 | System overview | 85 |
| 4.1.2 | The image transform in the retina model | 85 |
| 4.1.3 | Spikes coding using the stack-run algorithm | 88 |
| 4.1.4 | Arithmetic coding | 92 |
| 4.2 | Results | 96 |
| 4.2.1 | Quality metrics | 96 |
| 4.2.2 | Comparison to JPEG standards | 97 |
| 4.3 | Discussion | 98 |

Keywords

Static image compression, spiking retina model, rank order coding, stack run, context-based arithmetic coder.

Overview

In this chapter, we present an original retina-inspired coding scheme for static images. Our coder has three stages that combine the spiking retina model

presented in Chapter 2 and data compression techniques from the literature. In the following, we detail the specifications of our new coder stages, then we compare its performances to state of the art coders. This chapter is based mainly on the results that we published in [Masmoudi 2010d].

Contribution

We designed a novel image coding algorithm that uses the visual neural code as generated by a model of the retina. Thanks to this design basis, our coder goes beyond the current standards. Besides, image compression is an original application of neural-based codes. The contribution of our coder is the combination that we made of (i) a model of a biological coding device, and (ii) data compression techniques. These techniques include stack-run coding and context-based arithmetic coding. In this context, we attached a particular attention to the representation formalism of the code generated at the level of each stage. We showed that our coder has encouraging performances compared to state of the art coders under very strong bandwidth restrictions. Besides, we showed that our coder acts as a coder coupled to a denoiser if the data to encode is highly corrupted with noise. This feature is of great interest in the visual systems with noisy sensors and restricted transmission channel.

Organization

This chapter is organized into three sections:

1. Section 4.1 details the three stages of the coding pathway. First in Section 4.1.1, we make a quick overview of the whole system. Then, we describe separately each stage:
 - i. Section 4.1.2 recalls the retina-inspired image transform used.
 - ii. Section 4.1.3 details the stack-run coder.
 - iii. Section 4.1.4 details the context-based arithmetic coder that compresses the bit-stream.
2. Section 4.2 we summarize the results obtained. We considered two cases:

- i. The case of noiseless data to be transmitted at low bit-rates.
 - ii. The case of noisy data to be transmitted at average bit-rates.
3. Section 4.3 discusses the results obtained and introduces ideas of research efforts to be made. We especially highlight the potential of our approach based on biology mimicking for compression applications.

When you take something extremely broad, then it is not a work of expansion or work of compression. It's hard because you have to decide what to throw out.

Iris Chang

We present, in this chapter, a novel retina-inspired coding scheme for static images. We aim at using this coder to transmit the compressed visual information over low-bandwidth channels. As discussed in Chapter 1, several measurable features of the neural code of the retina could convey this visual information. In this work, we focused on the rank order coding hypothesis that we detailed in Chapter 2. The rank order coding hypothesis states that there is information in the order in which the retina cells fire spikes [Thorpe 1998]. This choice was motivated by several neurophysiologic results obtained mainly in the context of stimulus categorization in the ventral stream. The rank order coding was further supported by the conception of a bio-inspired retina model [Van Rullen 2001b].

Our coder combines this spiking retina model and well known data compression techniques. The fundamental hypothesis behind this work is that the retina generates a neural code for the visual stimulus that has a suitable rate/quality tradeoff. So that, the main novelty introduced in this chapter is to show how could the neural code of the retina be exploited in the context of still image compression. Our coder has three main stages. The first stage is the enhanced version of the retina model studied as we detailed it in Chapter 3 and in [Masmoudi 2012b, Masmoudi 2012a]. This model transforms an image into a rank profile based on the first wave of spikes. In the second stage, we re-express the rank profile of a given stimulus image using a 4-ary dictionary. This is done through the stack run coding technique. The third stage consists in applying a context-based arithmetic coder to the stack run coded signal. Then, we compare our results to the JPEG standards and we show that our model has comparable performances for lower computational cost under strong bandwidth restrictions. In particular, we discuss an interesting feature of the coder/decoder that we devised: the scalability. We also study the case when the visual data is contaminated with noise and show that

our coder acts as a coder coupled to a denoiser.

This chapter is organized as follows. In section 4.1, we detail the stages of our static image coding scheme. Then in Section 4.2, we summarize the results obtained, and in Section 4.3, we highlight many potential avenues for future research efforts in the conception of novel bio-inspired compression schemes.

4.1 The coding pathway

In this chapter, we would like to investigate if the retina model under study could be useful for image compression. So that, in Section 4.1.1 we give a proposal for a coding scheme based on rank ordering, then we detail its three stages from Section 4.1.2 to Section 4.1.4.

4.1.1 System overview

Our coding scheme encompasses three stages. First, the enhanced spiking retina model presented in the previous chapter transforms the image into a rank profile using a DoG filter bank. The exact values of the transform coefficients are recovered from the rank profile through a parametric look-up-table, so that the considered parameter is estimated at this level and is added to the bit-stream. Second, a zero-run length coder, namely a stack run coder, re-expresses the rank order code using a 4-ary dictionary. At this level the code is divided into two different files. The first one contains run lengths and the second contains non-zero coefficients. Finally, a context-based arithmetic coder compresses the bit-stream before its transmission to the decoder. These stages are summarized in the block diagram shown in Figure 4.1. The following sections detail each one of these stages.

4.1.2 The image transform in the retina model

The retina model described in Chapter 2 encompasses an analyzing stage and a synthesizing one. The analyzing stage transforms the stimulus image into a rank profile as the retina is supposed to do. It involves (i) a model of the stimulus transform in the retina consisting in a DoG filter bank analysis, and (ii) a sorting stage of the filters according to their activation degree.

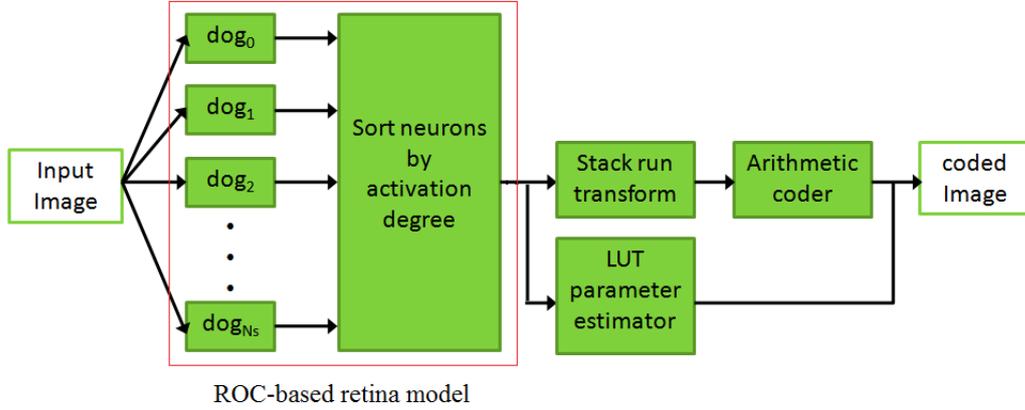


Figure 4.1: Block diagram of the compression scheme. First, a retina model based on rank ordering analyses the stimulus image. The image is analyzed through a dyadic grid of DoG filters to get a series of activation coefficients. Then these coefficients are sorted in the decreasing order of their response energy. Starting from these coefficients, the parameter of the look-up-table is estimated and added to the bit-stream. Second, the rank order code is zero-run length encoded by the stack run coder. Finally a context-based arithmetic coder is applied to get the compressed image file. The decoding process goes exactly the opposite way.

As described in Section 3.2.1, the original filter bank was augmented with a Gaussian scaling function. With this enhancement, the retina-inspired transform defined amounts to an image projection onto a frame of vectors Φ . From an implementation point of view, this transform will be performed by a single matrix computation as detailed in Section 3.1.1. This implementation has the advantage of being simple and rapid. We outlined this transform for a given image f in Equation (3.1) that we recall below:

$$c = \Phi f.$$

Once this transform is performed, we generate the rank profile of the image which is the only data encoding it. The rank profile is the sorted series of cells indexes coupled to their sign of activation: $(r_{kij}, \text{sign}(c_{\sigma^{-1}(k,i,j)}))$ (cf. Section 2.3.2). In the following, we recall the definition of the ranked index r_{kij} (cf. Equation (2.11)):

$$\begin{cases} (k', i', j') = \sigma((k, i, j)) \\ r_{kij} = 0, & \text{if } k' = 0, \\ r_{kij} = k' N_{k'-1}^2 + i' N_{k'} + j', & \text{if } k' > 0, \end{cases}$$

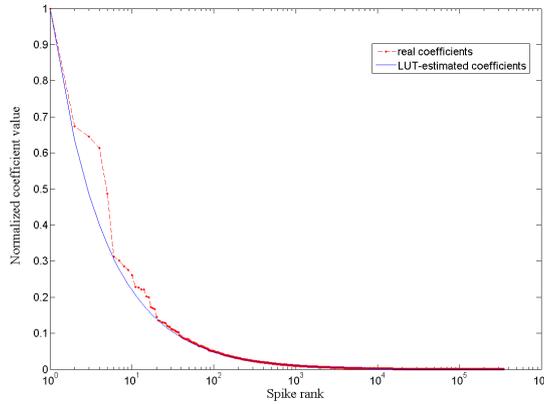


Figure 4.2: The normalized look-up-table estimated for Lena. The real ($|c_{kij}|$) coefficients values (dotted red line) are compared to the parametric functional f^{rank} (continuous blue line). The parameter of the function γ is estimated through a gradient descent.

where σ is the permutation that sorts the transform coefficients c_{kij} in the decreasing order of their energy (cf. Equation (2.9)).

Besides, a one-to-one map f^{rank} allows the recovery of the amplitudes ($|c_{kij}|$) knowing their firing rank. This one-to-one map (or also look-up-table) is known a priori by the decoder, so that there is no need to transmit it. The loss of the exact values of ($|c_{kij}|$) has the advantage of diminishing the amount of information that is necessary to encode the stimulus image. In [Perrinet 2003], f^{rank} is defined as follows (cf. Equation (2.10)):

$$f^{rank}(r_{kij}) = C (r_{\sigma^{-1}(k,i,j)})^{-\gamma}.$$

Though unlike the implementation in [Perrinet 2003], we consider that (i) $C = |c_{\sigma(0,0,0)}|$ rather than an arbitrary constant positive value, and that (ii) the parameter γ of f^{rank} is estimated using a gradient descent algorithm rather than using an average over a set of images. By this means, we get a parameter α that is specific to each image to be coded. This enables a smaller reconstruction error as the look-up-table f^{rank} is closer to the exact values. Besides, the algorithm is simple to implement. With an adequate initialization, the gradient descent algorithm converges quickly. In our case, the initialization is set to $\alpha = 0.5$. Figure 4.2 shows an example of such a map.

At the end of this stage, we get the rank order code of an image as the next stages will be using it. In this context, we redefine a spike entity, that

we denote e , as follows:

$$e_{\sigma^{-1}(k,i,j)} = \left(r_{kij}, \text{sign}(c_{\sigma^{-1}(k,i,j)}) \right). \quad (4.1)$$

If N_s is the dimension of c (cf. Equation (3.1)), then the list of N_s spikes (e_{kij}) is the complete rank order code of the stimulus image f . According to the bit-rate/quality of the image to be transmitted, one can restrict the code to the first and thus most significant N_c spikes. The message $M_{retina}^{N_c}$ received by the decoder is then a truncated rank order code that we define as follows:

$$M_{retina}^{N_c} = \underbrace{\left(e_{0,0,0}, e_{1,0,0}, \dots, e_{k,i,j}, \dots, e_{k_c,i_c,j_c} \right)}_{\text{The } N_c \text{ most significant spikes}}. \quad (4.2)$$

The message $M_{retina}^{N_c}$ in its current shape may be unsuitable for compression. Interestingly, $M_{retina}^{N_c}$ can be represented in a sparse manner to enable a zero-run length representation of it. We will detail this representation transform in the following section.

4.1.3 Spikes coding using the stack-run algorithm

As demonstrated in Figure 2.7, few spikes can reasonably represent the image to code. Indeed while the pyramid-like transform is highly populated with nil coefficients, one can reconstruct a suitable estimation of the stimulus. This property is loosely referred to as code sparseness in the literature. The sparseness of neural codes has been shown to help conceive meaningful representations of the visual data [Olshausen 1996]. Keeping sparseness as a design principle, we define in this section a sparse representation of the series $M_{retina}^{N_c}$. We then introduce a zero-run length coder that is well suited for sparse data coding, namely the stack-run coder.

Making the rank order code sparse

In general, coding the message $M_{retina}^{N_c}$ is expensive in terms of bit-rate. Indeed, the dimension of $M_{retina}^{N_c}$ is N_c but the values to encode may be large. In addition, the representation defined in Equation (4.2) does not account for the spatial neighbourhood between the indexes. For this reason, we propose

an alternative though equivalent representation that is sparse and that takes into account the relative positions of the spiking cells. So, we denote this new representation of the coded image by $M_{sparse}^{N_c}$ and we define it by:

$$M_{sparse}^{N_c} = (m_0, \dots, m_{l_{kij}}, \dots, m_{N_s-1}), \quad (4.3)$$

where l_{kij} is the lexicographic index of the considered cell defined by:

$$l_{kij} = \begin{cases} 0, & \text{if } k = 0, \\ k N_{k-1}^2 + i N_k + j, & \text{if } k > 0, \end{cases}$$

and $m_{l_{kij}}$ is the scalar value determined by the following equation:

$$m_l = \begin{cases} (1 + l_{\sigma^{-1}(k,i,j)}) \text{sign}(c_{\sigma^{-1}(k,i,j)}) & \text{if } 0 \leq l_{\sigma^{-1}(k,i,j)} < N_c, \\ 0 & \text{otherwise.} \end{cases} \quad (4.4)$$

The representation $M_{sparse}^{N_c}$ is strictly equivalent to $M_{retina}^{N_c}$ up to a one-to-one map since σ is a bijection. Let us consider an example of such a transform with $N_s = 21$ and $N_c = 4$. The message M_{retina}^4 could have the following value:

$$M_{retina}^4 = ((20, " + "), (4, " + "), (9, " - "), (15, " + ")). \quad (4.5)$$

Then, the corresponding sparse representation is given by M_{sparse}^4 such that:

$$M_{sparse}^4 = (0, 0, 0, 0, 2, 0, 0, 0, 0, -3, 0, 0, 0, 0, 0, 4, 0, 0, 0, 0, 1). \quad (4.6)$$

The dimension of $M_{sparse}^{N_c}$ is N_s , but the range of values to encode depends now on N_c . This detail is important because, if $N_c \ll N_s$, we decrease the number bits that are necessary to encode the significant values (in blue in Equation (4.6)). Besides, the runs of zeros are generally not expensive to encode (in red in Equation (4.6)). For a sufficiently small value of N_c , $M_{sparse}^{N_c}$ is a sparse data set. This sparseness is the feature we tried to enhance in our message code $M_{sparse}^{N_c}$.

A zero-run length code for the sparse representation of the image

Zero-run length coders are well suited for the compression of sparse data sets such as $M_{sparse}^{N_c}$ [Bell 1990, Sayood 2000]. In our scheme, we use an enhanced run-length coding algorithm called the stack run code [Tsai 1996]. This specific choice is motivated by both the simplicity of the algorithm and its high efficiency experienced for a variety of coding applications [Antonini 2006]. The stack run coding algorithm uses a 4 symbol dictionary D such that $D = \{0, 1, +, -\}$. As any zero-run length coder, $M_{sparse}^{N_c}$ is mapped into a series of couples (zero-run length, non-zero value). As specified in [Tsai 1996], the subsequent bit-wise operations are then applied:

- i) "-" encodes a binary bit 0 in run lengths.
- ii) "+" encodes a binary bit 1 in run lengths.
- iii) "0" encodes a binary bit 0 in non-zero values.
- iv) "1" encodes a binary bit 1 in non-zero values.

The use of 4 symbols in the alphabet removes the ambiguity between zero-run lengths and significant coefficient values. In addition to the rules above, the following rules enhance the algorithm:

- i) Every non zero value is set to its absolute value after we stored its sign.
- ii) The MSB of a non-zero value is set to "+" if the sign is positive.
- iii) The MSB of a non-zero value is set to "-" otherwise.

The resulting stuck run coded message will be denoted $M_{stack}^{N_c}$. For example, if we consider the following portion of code:

$$M_{sparse}^{N_c} = (0, 0, 0, 0, 0, 0, 0, 0, 5, 0, 0, 0, 0, 0, -18), \quad (4.7)$$

then applying the rules of stack-run encoding we obtain the stack run code $M_{stack}^{N_c}$ such that:

$$M_{stack}^{N_c} = - - - +10+ + - +0100-, \quad (4.8)$$

The string delimited by the first numeric character is "-- -+". This string must be read from the right to the left, so that we get the binary value "1000" encoding for 8 zeros. Then to the first alphabetic character included, we get the string "10+". This string encodes for the significant value "5". The rest of the code is read in the same manner. It is to be noted that the stack-run thus defined is a simplified version of the implemented algorithm. Further optimizing mechanisms are discussed in [Tsai 1996].

Context split

While reading the portion of code in Equation (4.8), the decoder switches from the *run context* to the non-zero value or *stack context* when it encounters a "0" or a "1". The decoder switches contexts in the opposite way when it encounters a "+" or a "-". Thus, we split the $M_{stack}^{N_c}$ code into two different files, one for the run context and the other for the stack context. Considering the example in Equation (4.8), we get in the first file F_{run} :

$$F_{run} = - - - + 1 + - + 0, \quad (4.9)$$

and in the second one F_{stack} :

$$F_{stack} = 0 + 100 -. \quad (4.10)$$

Thanks to this split, we get (i) a first file F_{run} mainly populated with +'s and -'s, and (ii) a second file F_{stack} mainly populated with 0's and 1's. Indeed when we separate the run context from the stack context, we skew occurrence probabilities of the symbols in each file. We can verify the histograms skewness of F_{run} and F_{stack} in Figure 4.3. This figure shows the probability distribution function of the symbols in D within each files for the test image Lena. Comparable results are obviously obtained for several other test images. Having these two files with enhanced contextual features, we can apply an arithmetic coder for compression with optimal performances. We present the principles and features of arithmetic coding in the following section.

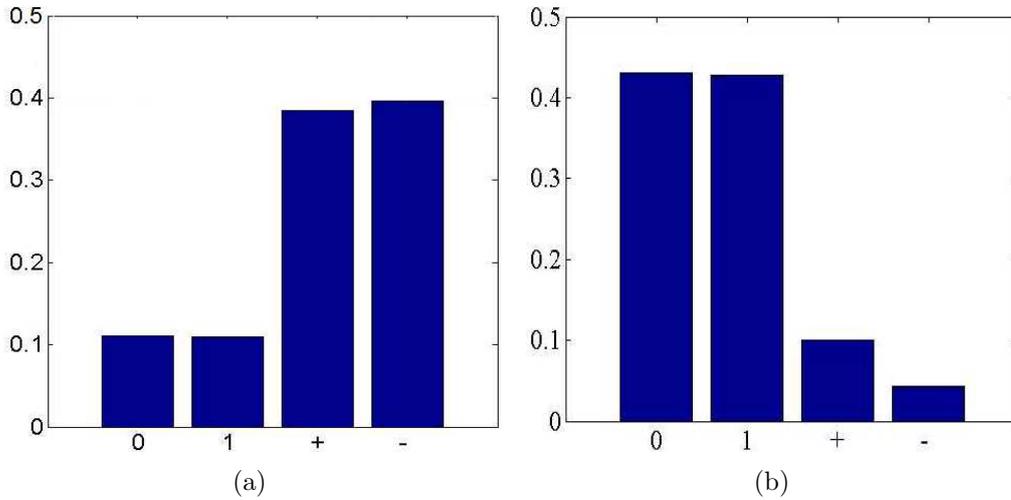


Figure 4.3: Histograms of the two files composing the stack run code for the test image Lena. 4.3(a): Histogram of the zero run length file F_{run} . 4.3(b): Histogram of the non-zero values file F_{stack} .

4.1.4 Arithmetic coding

At the end of the stack run coding stage, we get a code written with a 4 symbol dictionary D . This code is dispatched on two files, such that the probability distribution within each one of them becomes highly skewed. Interestingly arithmetic coding is an efficient lossless compression algorithm, especially when dealing with small alphabets and highly skewed probabilities [Girod 1993, Sayood 2000]. Thus, as in [Tsai 1996], we make the choice of coupling the stack run coder to a context-based arithmetic coder. The interaction between the stack run coder and the arithmetic coder stages is schematized in Figure 4.4.

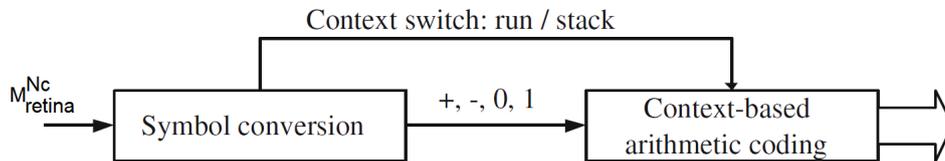


Figure 4.4: Block diagram of the stack run coder coupled to a context based arithmetic coding (inspired from [Hoang 2009]). The context is split into a run context and a stack context. This split is meant to skew the occurrence probabilities within each context and improve the arithmetic coder performances.

Arithmetic coding is a lossless compression technique that represents the

data to code as a fractional value in the scalar segment $[0, 1[$ [Langdon 1984, Sayood 2000]. Let us consider a message $M = (m_0, m_1, \dots, m_i, \dots, m_{N_s})$. Unlike entropic coders, arithmetic ones do not replace each symbol m_i in M by a predefined sequence of full bits. An arithmetic coder encodes the totality of the message in one stroke. In order to do this, the coder is provided with a model distribution P^M . This model is a map representing a plausible expectation of the probability distribution function of the symbols in M . P^M associates each symbol of the dictionary D to a predefined occurrence probability in the message M . Formally, a model P^M is a function defined as :

$$\begin{aligned} P^M : D &\longrightarrow [0, 1[: \\ d_i &\mapsto P^M(d_i), \end{aligned} \tag{4.11}$$

where $\sum_0^3 P^M(d_i) = 1$ [Bodden 2007]. Furthermore, we associate to P^M its cumulative distribution function χ^M augmented with the scalar value $\chi^M(d_{-1})$ such that:

$$\chi^M(d_i) = \begin{cases} 0, & \text{if } i = -1 \\ \sum_{k=0}^i P^M(d_k), & \text{if } 0 \leq i \leq 3. \end{cases} \tag{4.12}$$

The principle of the algorithm lies in the association of a string of symbols to an unique interval in $[0, 1[$. On a first step, the interval $[0, 1[$ is split into four subintervals $(T_i^0)_{0 \leq i \leq 3}$ such that their lengths are proportional to $(P^M(d_i))_{0 \leq i \leq 3}$. These subintervals are $T_i^0 = [\chi^M(d_{i-1}), \chi^M(d_i)[$. If $m_0 = d_i$, T_i^0 is associated to the substring message (m_0) . On a second step, the subinterval T_i^0 is split, in its turn, into four subintervals $(T_i^1)_{0 \leq i \leq 3}$ such that their lengths are proportional to $(P_M(d_i))_{0 \leq i \leq 3}$. As for the first iteration, the subinterval T_i^1 such that $m_1 = d_i$ is associated to the string message (m_0, m_1) . The same process is iterated until we reach the end of the string message M . The result of the process described is a single interval $T_i^{N_s}$ coding for the entire message. The median value of $T_i^{N_s}$ could be used as a tag encoding for the message. Figure 4.5 shows an example of the processing of the arithmetic coding algorithm.

From an implementation point of view, we will be using the adaptive arithmetic coder in [Bodden 2007]. Adaptive arithmetic coders are an extension of the classical arithmetic coder specified above. The coder used is adaptive

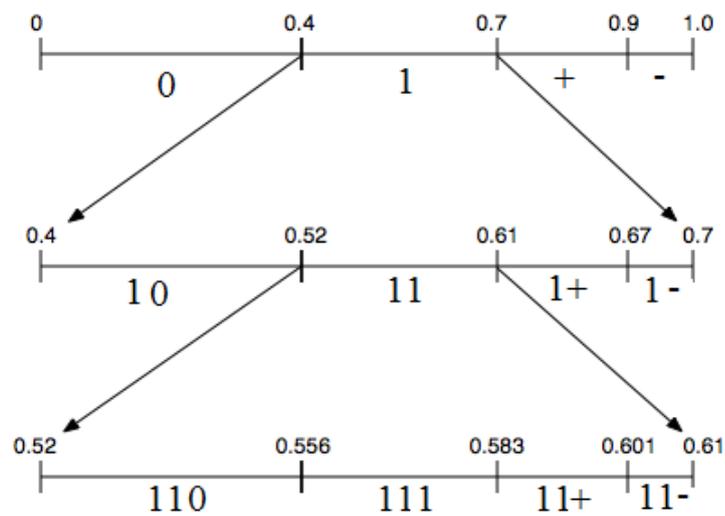


Figure 4.5: An example of arithmetic coding : The string to encode is 11+ and the model P^M used is defined by $P^M(0) = 0.4$, $P^M(1) = 0.3$, $P^M(+)$ = 0.2 and $P^M(-)$ = 0.1. On a first step, the interval $[0, 1[$ is split into subintervals which lengths are proportional to the model probabilities. The first symbol is then associated 1 to the interval $[0.4, 0.7[$. On a second step, we split $[0.4, 0.7[$ into subintervals which lengths are proportional to the model probabilities. Then, the string 11 is associated to the interval $[0.52, 0.61[$. The interval $[0.52, 0.61[$ is split, in its turn, in the same manner as the two first iterations. Finally, the string 11+ is associated to the interval $[0.583, 0.601[$.

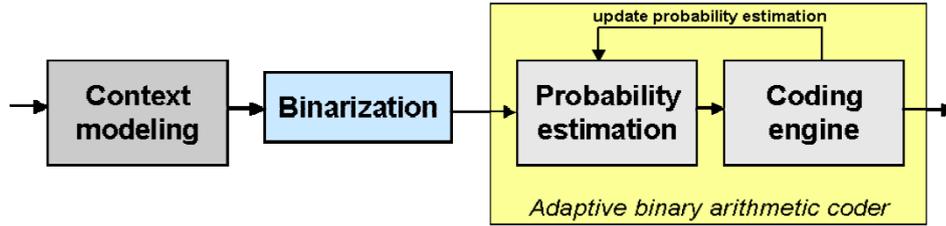


Figure 4.6: Example bloc diagram of a context based adaptive arithmetic coder (from [Marpe 2001]).

in the sense that it performs an update of the probability model P^M . Figure 4.6 shows an example of block diagram of such a coder. For a detailed specification of the algorithm, interested readers may refer to [Witten 1987]. In our specific case, we can provide our coder with two starting probability distribution functions P_{run}^M and P_{stack}^M . These models will evolve during the coding process to adapt to the file currently being processed. The subsequent empirical models can be used, (i) to encode the F_{run} file:

$$\left\{ \begin{array}{l} P_{run}^M(0) = 0.1 \\ P_{run}^M(1) = 0.1 \\ P_{run}^M(-) = 0.4 \\ P_{run}^M(+) = 0.4 \end{array} \right.$$

and (ii) to encode the F_{stack} file:

$$\left\{ \begin{array}{l} P_{stack}^M(0) = 0.4 \\ P_{stack}^M(1) = 0.4 \\ P_{stack}^M(-) = 0.1 \\ P_{stack}^M(+) = 0.1 \end{array} \right.$$

The context-based adaptive arithmetic coder is the last stage of our coding scheme.

Having our coder thus specified, we can infer in a straightforward manner the decoding scheme. The decoding process goes exactly the opposite way. First, we perform consequently the arithmetic decoding, and the stack run decoding. Then, we recover the input stimulus by means of the dual frame as described in Equation (3.15) (Chapter 3). So that in the next section, we present the resulting performances of our coder/decoder and compare them

to standard image compression algorithms.

4.2 Results

At this stage we have a complete implementation of a coder/decoder for static images. In order to do this, we first specify the quality metrics that we use. Then, we plot comparative rate-quality curves. Finally, we discuss competitiveness of our codec compared to JPEG and JPEG2000 [Skodras 2001, Christopoulos 2000].

4.2.1 Quality metrics

Our goal is to quantify the quality of the reconstructed image \tilde{f} compared to the original one f . Several quantitative measures for image quality are discussed in the literature. The most common one is the peak signal to noise ratio (PSNR) which is based on the mean squared error (MSE). PSNR is a quadratic measure and thus is well suited for optimization processes. Though, PSNR has been proved to be inconsistent with human eye perception [Girod 1993]. Other metrics were designed as alternatives to PSNR. We are interested in the metrics that quantify the reconstruction fidelity by comparing f and \tilde{f} structures. Among them the Structural SIMilarity (SSIM) index [Wang 2004b] is a measure for statistical similarity between two images \tilde{f} and f . The SSIM index is a functional of the first and second order statistical metrics. The computation of it consists in computing a similarity index between two analogous windows in \tilde{f} and f . The SSIM measure is then the average over all possible windows of the index. The resulting quantity is a coefficient between 0 and 1. Formally we define SSIM [Wang 2004b] by :

$$\begin{aligned}
 SSIM_{index}(W_x, W_y) &= \frac{(2\mu_{W_x}\mu_{W_y} + c_1)(2cov_{W_x W_y} + c_2)}{(\mu_{W_x}^2 + \mu_{W_y}^2 + c_1)(\sigma_{W_x}^2 + \sigma_{W_y}^2 + c_1)} \\
 M_{SSIM}(\tilde{f}, f) &= \frac{1}{N_w} \sum_{W_x, W_y} SSIM(W_x, W_y)
 \end{aligned} \tag{4.13}$$

where W_x is a window of \tilde{f} , W_y the corresponding window in f , μ_{W_x} the average of W_x , μ_{W_y} the average of W_y , $\sigma_{W_x}^2$ the variance of W_x , $\sigma_{W_y}^2$ the

variance of W_y , $cov_{W_x W_y}$ the covariance of W_x and W_y , N_w the number of windows.

4.2.2 Comparison to JPEG standards

We compared our results to the existing JPEG standards behavior under strong bandwidth restriction. Rates (or bit-costs) are measured in bits per pixel based on the size of the file generated by each coder. Performances are comparable until 0.15 *bpp* image rate, which shows our algorithm to have encouraging performances. The comparison curve in terms SSIM is plotted in Figure 4.7.

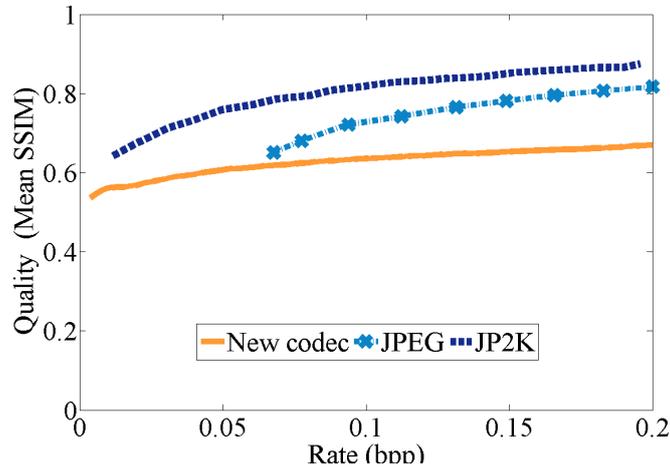


Figure 4.7: Comparison between JPEG, JPEG2000 and our new codec. The quality measure is the mean Structural SIMilarity (SSIM). The bit-cost measure is the bit per pixel (*bpp*).

Besides our codec shows good robustness to noise compared to JPEG and JPEG2000. Figure 4.8 shows the comparative performances of our codec and JPEG standards when dealing with noisy data. Indeed, the wavelet-like retina behavior in the model [Van Rullen 2001b] enables a better robustness. The rate/quality curve is plotted in Figure 4.9 and show up to 0.4 of gain in mean structural similarity measure (SSIM) for 0.25 *bpp* of image rate compared to classic JPEG. As the rate increases JPEG codecs convey more high frequency (HF) signals, which are noise. This explains the decreasing rate/quality behavior of JPEG. As HF is encoded with loss in JPEG, artifacts appear in the image decoded. On the contrary, our new codec do not show artifacts because

every spike is transmitted with no loss of information and encodes for the whole image. The scalability of our codec is monitored only by the choice of the number N_c of spikes to be encoded.



Figure 4.8: Robustness to noise: qualitative comparison between our new codec, JPEG, and JPEG2000 under the same rate restriction (here 0.27 bpp). Upper left: Lena (renormalized in the range of values from 0 to 1) with additive Gaussian noise ($mean = 0$, $variance = 0.05$). Upper right: coded/decoded image using the new codec. Lower left: coded/decoded image using JPEG. Lower right: coded/decoded image using JPEG2000.

4.3 Discussion

We have proposed a new bio-inspired codec for static images. First, the image is converted into a ROC code via a simplified retinal model, then a stack run coder is applied, followed by a first order arithmetic compressor. The

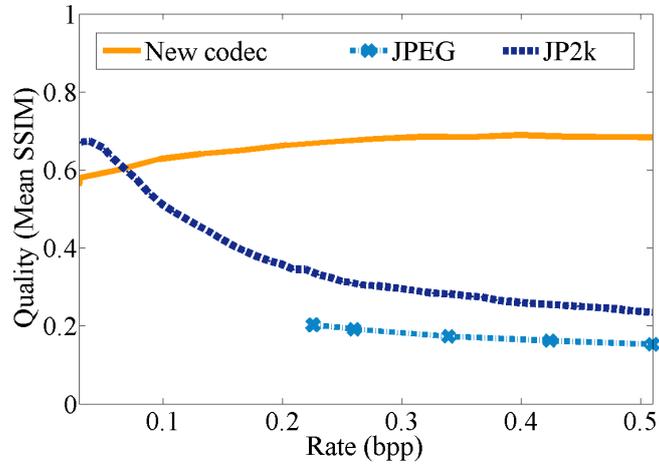


Figure 4.9: Rate/Quality behavior in the case of noisy data: quantitative comparison between our new codec, JPEG, and JPEG2000. The quality measure is the mean Structural SIMilarity (SSIM). The bit-cost measure is the bit per pixel (bpp).

performances of this coding scheme were tested against well established JPEG standards, and we obtain encouraging results for low bandwidth transmissions, especially when dealing with noisy data. This compression scheme also offers interesting features such as scalability and reasonable complexity. Limitations have been observed in terms of rate/quality, when compared to JPEG2000 for noiseless data transmissions. Beyond the proposition of a new compression scheme, we would like to highlight a variety of important issues and present potential avenues for future research efforts in this direction.

The first perspective concerns the retina coding model of our scheme. Although we focused on the latency time of the first spike, several models take into account the whole structure of spike trains. For example, it appears that burst or synchronies are features that could encode for the stimulus. This opens new perspectives to extend this model as soon as we are able to produce realistic spike trains. In particular, we will need to consider more realistic retina models converting videos into spike trains, such as [Wohrer 2009a]. The goal is then to use such models in order to reproduce some spiking pattern as observed in real cell recordings, and establish how spikes are triggered by a stimulus then decoded by the nervous system [Rieke 1997].

The second perspective concerns compressing spikes. In this work, even with a simplified representation of the spiking activity as a wave of spikes,

classical approaches as stack run coding are not optimal. In the general case, with a continuous and intense spiking activity, new ideas will have to be introduced. New bio-inspired compression schemes will have to take into account the features of the neural code that are the most relevant for the stimulus representation.

Part II

**FROM *VIRTUAL RETINA*
TO A BIO-INSPIRED AND
NON-DETERMINISTIC
IMAGE CODER WITH
TEMPORAL SCALABILITY**

Part I of this work was dedicated to the conception of a bio-inspired image coder starting from the rank order coding hypothesis. We especially considered the retina model defined in [Van Rullen 2001b] as a basis for our novel coder and made the adequate modifications to adapt it to the image coders/decoders design. Part II of this work will lead this effort towards more biological plausibility. Though the first model considered takes into account some of the retinal visual processing mechanisms, several other processing stages are ignored. So that, we will base the rest of this work on a more complete retina model called *Virtual Retina* [Wohrer 2009a]. Starting from *Virtual Retina*, our aim is to design a second bio-inspired image coder that reproduces as closely as possible the retina behavior.

Indeed the action of perception seems to be effortless, but yet neurophysiologic experiments proved that the mechanisms involved in the retina are highly complex and demanding. Recent studies such as [Gollisch 2010] confirmed that the retina is accomplishing non-trivial operations on the stimulus signal before its transmission to the visual cortex. The retina model described in [Van Rullen 2001b] and consisting in a linear filtering followed by a sorting stage is a rough simplification of the reality. Besides, the retina appears to be a non-deterministic system. The neural code of the retina fluctuates randomly, such that a single stimulus leads to different codes across trials. Various retina models from the literature propose to reproduce these phenomena. Efforts such as [Meister 1999, Wohrer 2008] made a review of a few of them. Among these models, we considered the retina simulation software named *Virtual Retina*. Our motivation to start from this model is that it keeps a strong biological plausibility and takes into account implementation constraints.

This Part is divided into two chapters. First in Chapter 5, we focus on the behavior of the inner layers of the retina solely. In this study, we isolate these inner layers to show that their behavior is similar to analog-to-digital converters. Based on that, we devise a deterministic then a non-deterministic quantizer with their corresponding decoders. In this context, we give a biologically plausible interpretation for the randomness of the neural code of the retina. Second in Chapter 6, we consider both the outer and the inner layers of the retina. Based on the behavior of this complete set of layers, we de-

sign an original scalable image coder. This novel coder accounts for the time-dependent and also non-deterministic behavior of the actual retina. Our coder mimics most of the retinal processing stages as specified in [Wohrer 2008] and its non-determinism enables interesting perceptual features.

A RETINA-INSPIRED ANALOG-TO-DIGITAL CONVERTER

Contents

| | | |
|------------|---|------------|
| 5.1 | <i>Virtual Retina: A bio-plausible model</i> | 109 |
| 5.1.1 | The outer layers | 110 |
| 5.1.2 | The inner layers | 112 |
| 5.1.3 | The ganglionic layer | 113 |
| 5.2 | A retina-inspired A/D converter | 114 |
| 5.2.1 | Study context | 114 |
| 5.2.2 | The coding pathway | 117 |
| 5.2.3 | The decoding pathway | 122 |
| 5.3 | Our A/D converter <i>vs</i> standards | 124 |
| 5.3.1 | Our A/D converter overall behavior | 124 |
| 5.3.2 | A beyond the standards A/D converter | 126 |
| 5.4 | The dithering hypothesis | 130 |
| 5.5 | Discussion | 135 |

Keywords

Analog-to-digital converter, *Virtual Retina*, spikes, rate codes, Lloyd-Max quantizer, companding, dithering.

Overview

In this chapter, we explore the behavior of the inner layers of the mammals retina. We take the *Virtual Retina* simulation software as a basis for our study. In a first step, we design a dynamic quantization scheme that relies on a rate code. This scheme encompasses three stages mimicking the inner layers of the retina. In a second step, we hypothesize that the noise observed in the retina is a dither signal. We then overview a few interesting features of our retina-inspired analog-to-digital (A/D) converter. This chapter is based mainly on the results that we published in [Masmoudi 2010a, Masmoudi 2010b, Masmoudi 2010c].

Contribution

We start from state of the art equations modelling each retina layer under study as they are resumed in [Wohrer 2009a]. Based on these efforts, our contribution in this chapter is twofold. The first contribution is the conception of a deterministic quantization/decoding system that goes beyond the standards. Indeed, our new bio-inspired quantizer has two main properties: (i) It emphasizes high magnitude signals rather than high probability ones, and (ii) it gradually changes across time from a coarse and quasi-uniform quantizer to a refined and highly non-linear one. In addition to its originality, our quantization system offers several interesting features as time scalability. The second contribution is the proposal of a novel hypothesis regarding the non-determinism observed in the neural code of the retina. This hypothesis supposes that the retina code randomness is the result of a dithering process. Under this assumption, interesting features are added to our quantization system such as (i) error whitening and (ii) error decorrelation from the input stimuli. Our proposal concerning the dither noise in the retina found an echo recently in the computational neurosciences community [Vidne 2012].

Organization

This chapter is organized into five sections:

1. Section 5.1 describes the main stages of the *Virtual Retina* simulation software.
2. Section 5.2 specifies the three stages of our quantization scheme and the corresponding decoding algorithm.
3. Section 5.3 compares our retina-inspired A/D converter to classical quantization systems.
4. Section 5.4 introduces our hypothesis about the non-determinism observed in the retinal code. Our hypothesis states that this non-determinism is the result of a dithering process.
5. Section 5.5 discusses the contributions of this chapter.

*Do what you know and perception is
converted into character.*

Ralph Waldo Emerson

Our goal in part II of this work is to design a bio-inspired image coder that keeps a strong plausibility with regard to the mammalian retina behavior. In order to do this, we first consider in this chapter the deep retina layers and inspire ourselves from their behavior to design an original and dynamic A/D converter.

Obviously, A/D signal conversion is a crucial step in the image coding chain. Interestingly as described in Chapter 1, the retina and especially its deepest layers operate a quantization process. Starting from a continuous signal, the retina generates a series of quantum electrical entities, the spikes. In order to reproduce a plausible spike-based neural code, we base our study on a biologically realistic model of the retina named *Virtual Retina* [Wohrer 2009a]. In a first step, we focus on the temporal behavior of the three last retina layers. Starting from this spike-based code, we design a dynamic quantization scheme that relies on a rate code for its simplicity and efficiency. Then, we propose a possible decoding procedure. This yields an original quantization/decoding system which evolves dynamically from coarse to fine, and from uniform to non-uniform. The bio-inspired A/D converter that we propose offers several interesting features as the scalability, and the introduction of time dependency in the coding system. These features inset implicit bit-allocation. In a second step, we model the retinal noise by a dither signal, and thenceforth we identify the retina behavior to a non-subtractive dithered quantizer. This hypothesis gives a possible interpretation for the non-determinism observed in the spike-based neural code of the retina. When introduced in our A/D converter, the dithering process enables several interesting features. Indeed as we will show later in this chapter, the dithering whitens the reconstruction error and decorrelates it from the input stimuli.

This chapter is organized according to five sections. First in Section 5.1, we describe the main stages of *Virtual Retina* and the biological phenomena that are modelled within it. Second in Section 5.2, we specify a quantization scheme that mimics the time behavior of the three deepest retina layers, and then we propose a possible decoding algorithm. Third in Section 5.3, we compare

the specified retina-inspired A/D converter to classical quantization systems. In particular, (i) we raise the analogy between the retina behavior and the general scheme of A/D conversion, and (ii) we emphasize the original features of our system as its time dynamics. Fourth in Section 5.4, we hypothesize that the non-determinism observed in the retinal code is resulting from a dithering process. In this context, we show qualitative results on the role of dithering in a coding system. Finally in Section 5.5, we discuss the main features of our A/D conversion system and give perspectives to further developments.

5.1 *Virtual Retina: A bio-plausible model*

In this work, we considered the retina model called *Virtual Retina*. This choice is motivated by the completeness of the model and its adequacy to simulation purposes. In fact, *Virtual Retina* gathers state of the art models of several retinal mechanisms into a single simulation tool. Thus, *Virtual Retina* takes into account the major phenomena involved in the retina processing [Masquelier 2011]. Besides, the authors confirmed the biological plausibility of their model by reproducing actual retina cell recordings for several experiments. *Virtual Retina* follows the organization of the mammals retina. Indeed, the considered model is composed of a stack of layers, such that each layer has a specific functional role. The architecture of *Virtual Retina* is schematized in Figure 5.1. Three main processing steps can be distinguished within this organization:

- The outer layers: The first processing step is described by non-separable spatio-temporal filters. These filters behave as time-dependent edge detectors. This is a classical step implemented in several retina models. The cells modelled have an excitatory/inhibitory structure that enables the detection of spatial contours as wells as intensity variations in time. Several studies interpret this processing stage as a wavelet analysis of the input stimulus [Mallat 1989, Gaudart 1993].
- The inner layers: The second processing step is performed by a non-linear contrast gain control mechanism. This stage models mainly bipolar cells behavior by control circuits with time-varying conductances.

This stage is implemented as a low-pass separable spatio-temporal filter. Studies such as [Beaudoin 2007] tend to confirm the model employed.

- The ganglionic layer: The third processing step is performed by leaky integrate and fire neurons. These neurons model the ganglionic layer processing that finally converts the stimulus into spikes.

5.1.1 The outer layers

The processing of the retina outer layers is mainly conducted within the so-called outer plexiform layer (OPL). The OPL is tiled with cells of finite receptive fields that are organized retinotopically. Each one of these receptive fields have a central and a surround component that are interacting across time. This results in a phenomenon referred to as center-surround differentiation. In order to reproduce it, *Virtual Retina* models an OPL cell by a non-separable spatio-temporal filter. This filter is the difference of two low-pass filters representing the two components of a given receptive field. If we denote the input light stimulus by L and the current resulting from the center (respectively surround) filtering by I^c (respectively I^s), then the output current of this stage I^{opl} is defined by:

$$\begin{cases} I^c(x, y, t) &= G_{\sigma^c}(x, y) * T_{w^u, \tau^u}(t) * E_{n^c, \tau^c}(t) * L(x, y, t), \\ I^s(x, y, t) &= G_{\sigma^s}(x, y) * E_{\tau^s} * I^c(x, y, t), \\ I^{opl}(x, y, t) &= \lambda^{opl} \left(I^c(x, y, t) - w^{opl} I^s(x, y, t) \right), \end{cases} \quad (5.1)$$

where G_{σ^c} (respectively G_{σ^s}) is the Gaussian filter representing the central (respectively surround) component, E_{τ^s} is an exponential temporal filter, E_{n^c, τ^c} is a gamma temporal filter, T_{w^u, τ^u} is a high pass temporal filter, $\{\sigma^c, \sigma^s\}$ is the set of considered Gaussian standard deviations, $\{\lambda^{opl}, w^{opl}, w^u\}$ is the set of considered weights, and $\{\tau^u, \tau^c, \tau^s\}$ is the set of considered time constants. For a detailed description of the parameters, interested readers should refer to [Wohrer 2009a]. The continuous current I^{opl} is the output of this stage and is the only data to be processed by subsequent retina layers. I^{opl} conveys information about spatial and temporal local contrasts in the stimulus. Thus, we discard the OPL in this chapter since it does not intervene in A/D conversion process. We will introduce an original model of this stage in Chapter 6.

VIRTUAL RETINA

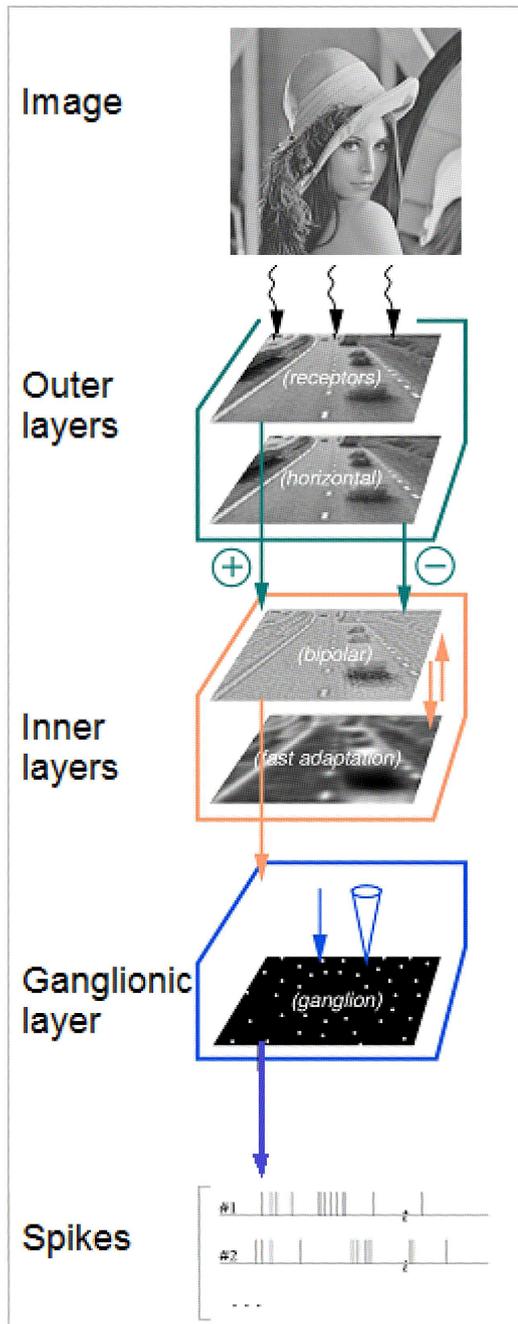


Figure 5.1: Schematic view of *Virtual Retina* architecture (from [Wohrer 2009a]). The input stimulus is processed from the topmost stage to the nethermost one. The green box delimits the outer layers processing stage. The orange box delimits the inner layers processing stage. This stage encompasses two sub-stages: (i) a fast gain controller and (ii) a non-linear rectifier. The blue box delimits the ganglionic layer stage. This stage is the one responsible for the transform of a continuous current into a countable set of spikes.

5.1.2 The inner layers

As described in the previous section, the outer layers of the retina generate a continuous current I^{opl} which magnitude is proportional to the local contrasts of the stimulus. Though, a major issue encountered by any biological system is to adjust its operational range to match the input stimuli magnitude range [Rieke 2001]. This adaptation is done in the inner layers of the retina in two steps. The first one is performed in the inner nuclear layer (INL) and the second is performed in the inner plexiform layer (IPL).

The inner nuclear layer: The retina has the capability to quickly adapt the gain of its transfer function to the input magnitude [Shapley 1978, Kim 2001, Baccus 2002]. This fast magnitude adaptation mechanism is largely observed at the level of the bipolar cells that populate the INL. The phenomenon involved is termed shunting inhibition and is both non-linear and dynamic. *Virtual Retina* includes a realistic model of it based on a feedback loop. If we denote the output potential of this stage by $V^b(x, y, t)$, then the contrast gain control is defined by the following equations:

$$\begin{cases} c^b \frac{dV^b(x, y, t)}{dt} &= I^{opl}(t) - g^b(t)V^b(x, y, t), \\ g^b(x, y, t) &= G_{\sigma^b} * E_{\tau^b} * Q(V^b(x, y, t)), \\ Q(V^b(x, y, t)) &= g_0^b + \lambda^b (V^b(x, y, t))^2, \end{cases} \quad (5.2)$$

where g^b is a variable leakage term in the membrane of a bipolar cell, Q is a quadratic activation function, E_{τ^b} is an exponential temporal filter, G_{σ^b} is a Gaussian spatial filter, and $\{c^b, g_0^b, \lambda^b\}$ is a set of constant scalar parameters defining the considered capacitance, conductance, and weight. As for the previous stage, interested readers should refer to [Wohrer 2009a] for exhaustive definitions.

As the retina has a layered architecture, the output voltage V^b is the input of the subsequent (IPL) stage.

The inner plexiform layer: We consider the signal, of voltage V^b , as generated by the bipolar cells of the retina. This potential is subject to a non-linear rectification in the IPL. Several types of cells are involved in this

rectification especially the amacrine cells. If we denote the current output of this stage by I^g , then a biologically realistic model of this rectification is given in *Virtual Retina* by:

$$I^g(x, y, t) = N \left(T_{w^g, \tau^g}(x, y, t) * V^b(x, y, t) \right), \quad (5.3)$$

where T_{w^g, τ^g} is a linear transient filter, $\{w^g, \tau^g\}$ is a set of constant scalar parameters, and N is a function defined by:

$$N(v) = \begin{cases} \frac{i_0^g}{i_0^g - \lambda^g(v - v_0^g)} & \text{if } v < v_0^g, \\ i_0^g + \lambda^g(v - v_0^g) & \text{if } v \geq v_0^g, \end{cases} \quad (5.4)$$

where $\{i_0^g, v_0^g, \lambda^g\}$ is a set of constant scalar parameters defining a current, a voltage, and a weight.

I^g is obviously the current input of the last retina stage, namely the ganglionic layer. This latter is the one that produces the neural code of the retina.

5.1.3 The ganglionic layer

The ganglionic layer is the deepest one tiling the retina. This layer is responsible of the transform of the $I^g(t)$ current into a discrete set of spikes. The cells of the ganglionic layer are modelled in *Virtual Retina* by leaky integrate and fire (LIF) neurons. Each neuron integrates a potential until it reaches a given threshold δ , fires a spike, and then restarts integrating. If we denote the potential of a ganglion cell by V^l , then we can define the spike timings (t^i) by the following equation:

$$\begin{cases} t^0 & = 0 \text{ s}, \\ t^i & < t^{i+1}, \forall i \geq 0, \\ V^l(t^i) & = \delta. \end{cases} \quad (5.5)$$

Each neuron in the ganglionic layer behaves according to the following equation:

$$\begin{cases} V^l = V_0^l & \text{if } t \in [t^i, t^i + \eta_{refr}^i[, \\ c^l \frac{dV^l(x, y, t)}{dt} = I^g(x, y, t) - g^l V^l(x, y, t) + \eta^l(x, y, t) & \text{if } t \in [t^i + \eta_{refr}^i, t^{i+1}[, \end{cases} \quad (5.6)$$

where (i) $\{V_0^l, g^l, c^l\}$ is a set of constant scalar values defining the reset voltage, the conductance and the capacitance, (ii) η^l is a noise source, and (iii) η_{refr}^i is the i^{th} realisation of a positive random variable. The variables η^l and η_{refr}^i are added to the spike generation process in order to reproduce the trial-to-trial variability of actual ganglion cells [Keat 2001, Stein 2005, Kostal 2007].

Starting from the retina model presented in this section, our aim is to devise a bio-inspired A/D converter. So that, we will specify in the next section a complete original dynamic quantization/decoding system based on a rate code.

5.2 A retina-inspired A/D converter

Our goal in this section is to revisit the last processing stages of *Virtual Retina* to design a novel bio-inspired A/D converter. The retina processing stages considered are the ones following the image transform stage in the outer layers. These stages include three layers, namely the INL, the IPL and the ganglionic layer. Thus, our A/D converter will have three stages that mimic all of the cited layers. First in Section 5.2.1, we specify our working hypotheses. Second in Section 5.2.2, we specify the stimulus coding steps through our A/D converter. Finally in Section 5.2.3, we describe the design of our decoder.

5.2.1 Study context

Up to our knowledge, there is no work that considered the retina as a basis for a complete quantization/decoding scheme. Though, few efforts have been made in the literature aiming at conceiving A/D converters that are inspired from neuron cell models. These works mainly consider the integrate and fire (IF) neuron model. For instance, authors in [Feichtinger 2011, Rastogi 2011] designed an asynchronous A/D converter based on IF neurons. The coding

process generates a family of spike timings $(t^i)_{i \geq 0}$ as a response to a time function $f(t)$. Along with the coding process, the authors designed an approximate reconstruction algorithm that assumes several constraints about the bandwidth and shape of the function to recover. The decoding algorithm is non linear and its main principle is to multiply a given kernel function (typically an exponential one) by delayed Diracs. A second algorithm is described in [Lazar 2004, Lazar 2008]. The authors designed another asynchronous A/D converter which they referred to as the time encoding machine. The neuron model considered is still the IF model, and the output is the family of spike timings $(t^i)_{i \geq 0}$. The approximate reconstruction algorithm is referred to as the time decoding machine. It assumes a minimum lag between two consecutive spikes and impose restriction on the bandwidth to be used. In this case too, the decoder performs a multiplication of delayed Diracs with a set of adequate low pass filters. We may also cite [Camuñas-Mesa 2010] as another neuron-inspired quantization scheme designed for recognition tasks rather than for the exact decoding of the stimulus.

The cited efforts assume a coding function ϕ_{if} transforming the stimulus $f(t)$ into a family of at most N^s spikes. The coding function ϕ_{if} could be defined as follows:

$$\begin{aligned} \phi_{if} : \mathcal{K}(\mathbb{R}) &\longrightarrow \mathbb{R}^{N^s} \\ f &\longmapsto (t^i)_{0 \leq i < N^s}, \end{aligned} \tag{5.7}$$

where $\mathcal{K}(\mathbb{R})$ is the space of continuous functions with compact support. Several reasons lead us to design an alternative coding scheme. The first reason is that these efforts consider integrate and fire neurons rather than the leaky integrate and fire (LIF) neurons that we manage to use (see Section 5.1.3). Besides the deep retina layers considered, and a fortiori our retina-inspired A/D converter, include two more non-linear stages prior to the LIF stage. Thus, the quantization/decoding algorithms described are unsuitable for our case. The second reason is that the cited works are directed towards the design of asynchronous devices, whereas our target application which is the image coding and communication supposes a synchronous coder. The third reason is that ϕ_{if} encodes each single value from the input stimulus into a family of

floating numbers $(t^i)_{0 \leq i < N^s}$. In the context of still image coding, the stimulus image magnitudes are integers lying in a restricted interval (typically $\llbracket 0, 255 \rrbracket$). So that, encoding each integer by the family of floating numbers $(t^i)_{0 \leq i < N^s}$ will lead to an increase in the bit-cost of the image. The latter observation makes this coding scheme incompatible with our compression purposes. Thus, we would rather consider a coding function $\phi_{lif}^{t_{obs}}$ that we define as follows:

$$\begin{aligned} \phi_{lif}^{t_{obs}} : \quad \mathcal{K}^1(\mathbb{R}) &\longrightarrow \mathbb{N} \\ f &\longmapsto |(t^i)_{t^i < t_{obs}}|, \end{aligned} \quad (5.8)$$

where \mathcal{K}^1 is the set of constant functions with compact support, $[0, t_{obs}[$ is the observation window, and $|\cdot|$ denotes the cardinality.

As made clear through Equation (5.8), we restrict the retina deep layers models to their temporal behavior as the state of the art studies did. The supplementary constraints that arise from Equation (5.8) are in concern with the input current, the temporal window of observation, and the coding feature that we assume in a spike train.

The input current: Focusing on the last stages of the mammals retina, we assume that the INL cells receive heaviside input currents that are constant during the observation window $[0, t_{obs}[$. So that the input current $I^{opl}(t)$ will be defined as follows:

$$I^{opl}(t) = \alpha \mathbb{1}_{\{t \geq 0\}}(t), \quad (5.9)$$

where α is the magnitude of the current to be encoded, and $\mathbb{1}_{\{t \geq 0\}}$ is the indicator function such that, $\mathbb{1}_{\{t \geq 0\}}(t) = 0$ if $t < 0$ and 1 otherwise. This assumption is justified by the input stimuli studied in our case which are still images. This amounts to suppose that our A/D converter encodes an image flashed during a given temporal window of observation $[0, t_{obs}[$.

The temporal window of observation: We will observe our system at different t_{obs} timings. We will assume that $0 \leq t_{obs} < T$ and that T has an order of magnitude of some tens of milliseconds. Typically in this chapter, $T = 30$ ms. There are two justifications for this choice of T . The first one is that 30 ms is approximately the duration of exposition of a frame in a

conventional video stream, though, there is no jolt while playing a sequence. Thus, we can argue that the capture and coding process through the retina takes less than 30 ms. The second justification is that the mammalian visual system is able to accomplish complex categorization tasks in durations as short as 150 ms [Van Rullen 2001a]. The authors of the latter work then concluded that the input stimulus spends approximately 10 ms per coding layer. Thus, $T = 30$ ms is a compatible duration with the number of processing layers crossed by the stimulus. In our A/D converter, these layers are in the number of three.

The coding feature in a spike train: As discussed in Chapter 1, there is a variety of possible coding features in a spike train. Among them, we chose the rate coding. Here, we assume that the number of spikes emitted within $[0, t_{obs}[$ is the relevant quantity that encodes neurons activity. Indeed, the spike count is a quantity that is assimilable to a firing rate. We denoted this quantity by $|(t^i)_{t^i < t_{obs}}|$ (cf. Equation (5.8)). In order to show our A/D converter behavior, we estimate $|(t^i)_{t^i < t_{obs}}|$ as a function of the intensity magnitude, and we do this at different observation timings t_{obs} . The spike count metric is the one that is commonly admitted for the analysis of the neural code since [Adrian 1926]. Thus, our metric choice is justified by (i) the common usage and (ii) the fact that this metric takes its values in the countable space \mathbb{N} .

5.2.2 The coding pathway

Having the assumptions specified in the previous section, we detail in the following the behavior of our A/D converter stages.

5.2.2.1 The rapid contrast gain control stage

The rapid contrast gain control stage, as described in Section 5.1.2, is a succession of two coding layers: the INL layer that ensures the contrast gain control and the IPL layer that rectifies the signal non-linearly. The time behavior of these two layers is described in the two paragraphs below.

The contrast gain control loop: First, we consider the contrast gain control stage in the retina. This stage is performed by the bipolar cells in the INL. Each bipolar cell gets a current $I^{opl}(x, y, t)$ and generates a bipolar voltage $V^b(x, y, t)$. The development of Equation (5.2) yields the following expression:

$$c^b \frac{dV^b}{dt}(x, y, t) = \frac{\lambda^b}{\tau^b} \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_0^t (V^b)^2(x-u, y-v, s) \frac{e^{-\frac{u^2+v^2}{2(\sigma^b)^2}}}{2\pi(\sigma^b)^2} e^{-\frac{t-s}{\tau^b}} ds dudv \right) V^b(x, y, t). \quad (5.10)$$

Then, we solve numerically the above ordinary differential equation through an Euler schema. In order to do this, we discretize the time and space as follows:

- G_{σ^b} has a finite spatial support, such that G_{σ^b} is an $(2m+1) \times (2m+1)$ discrete filter with an horizontal resolution of dx and a vertical resolution of dy .
- The $[0, t]$ time interval is binned according to a time resolution dt . We denote by t_i the instant such that $t_i = i dt$.

So that, we can write the following approximations:

$$\begin{aligned} \left(\frac{V^b(x, y, t_{i+1}) - V^b(x, y, t_i)}{dt} \right) &\simeq \frac{dV^b}{dt}(x, y, t_i), \\ V^b(x, y, t_{i+1}) &\simeq dt \left(\frac{dV^b}{dt}(x, y, t_i) \right) + V^b(x, y, t_i). \end{aligned}$$

This yields the following expression:

$$\begin{aligned} V^b(x, y, t_{i+1}) &= \frac{\lambda^b dt}{2\pi\tau^b c^b (\sigma^b)^2} \sum_{k=-m}^m \sum_{l=-m}^m \left(\sum_{j=0}^i (V^b)^2(x-kdx, y-ldy, t_j) e^{-\frac{t_i-t_j}{\tau^b}} dt \right) \\ &\quad e^{-\frac{(kdx)^2+(ldy)^2}{2(\sigma^b)^2}} dx dy V^b(x, y, t_i) + V^b(x, y, t_i). \end{aligned} \quad (5.11)$$

As we made a restriction considering only the time behavior of INL, we set $m = 0$ and $\sigma^b = 1$ m. For a given magnitude value α of the heaviside input $I^{opl}(t)$, let us estimate the time-varying output voltage $V^b(t)$. We do so for

a set of α values in $[0, 100]$ pA which is the range of standard current values in the retina. The model simulation leads to the results shown in Figure 5.2. Figure 5.2 proves that the gain of the INL stage increases across time. All current values in the standard range of intensities are magnified until the system saturates by 23 ms when the gain begin to decrease. The behavior of the gain control in the bipolar cells will be compared to standard signal processing devices in Section 5.3.

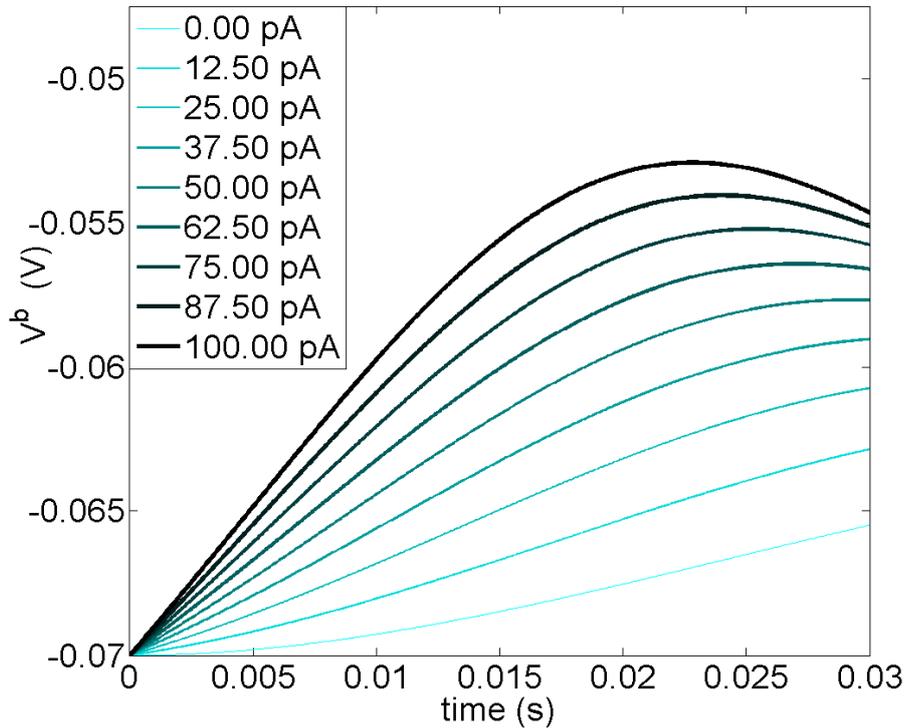


Figure 5.2: $V^b(t)$: Input/output map of the INL layer for a heaviside input $I^{opl}(t)$. Figure shows the time-varying output voltage $V^b(t)$ of the bipolar cells layer for different values of magnitude α , such that $I^{opl}(t) = \alpha \mathbb{1}_{\{t \geq 0\}}(t)$. The model parameters are set to biologically realistic values: $g_0^b = 5 \cdot 10^{-10} S$, $\tau^b = 20 \cdot 10^{-3} s$, $\lambda^b = 10^{-6} AV^{-2}$, $c^b = 10^{-10} F$.

The non-linear rectification Second, we consider the non-linear rectification stage which is ensured by the cells of the IPL layer of the retina. Each IPL cell gets a bipolar voltage $V^b(t)$ as an input and generates a ganglionic current $I^g(t)$. The effects of the non-linear rectification as defined in Equation (5.3) are shown in Figure 5.3. For a time-varying voltage $V^b(t)$, we notice that the behavior of $I^g(t)$ is different according to the values of α , where α is the magnitude of the input current. Indeed when compared to Figure 5.2,

Figure 5.3 shows that the the high magnitudes are differentiated with higher precision (cf. $20 < t_{obs} < 25$ ms) than the low magnitudes (cf. $t_{obs} < 10$ ms). This issue will be discussed with further details in Section 5.3.

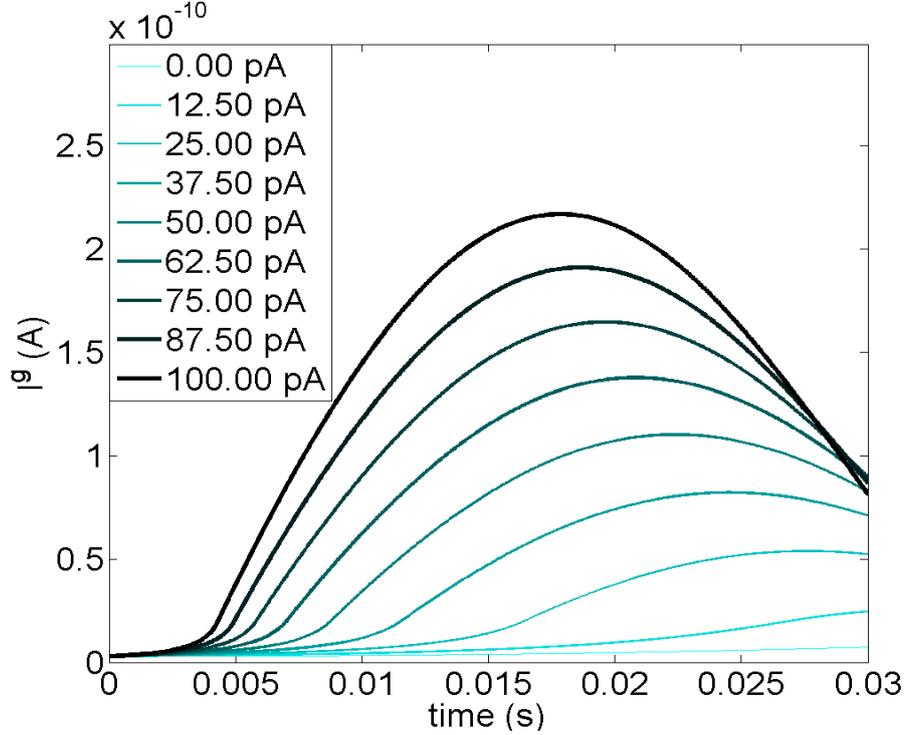


Figure 5.3: $I^g(t)$: Input/output map of the IPL layer cascaded to the INL layer. The non-linear rectification is cascaded to the contrast gain control for several heaviside inputs $I^{opl}(t)$. The figure shows the time-varying output intensity $I^g(t)$ of the IPL layer for different values of magnitude α , such that $I^{opl}(t) = \alpha \mathbb{1}_{\{t \geq 0\}}(t)$. The model parameters are set to biologically realistic values: $v_0^g = 4 \cdot 10^{-3}$ V, $i_0^g = 20 \cdot 10^{-12}$ A, $\omega^g = 7.5 \cdot 10^{-1}$, $\tau^g = 15 \cdot 10^{-3}$ s, $\lambda^g = 30 \cdot 10^{-9}$.

5.2.2.2 The spike generation stage

Finally given the initial input current $I^{opl}(t)$, our goal is to estimate the generated spike count within a given temporal window $[0, t_{obs}[$. We do this estimation for every possible input current magnitude α (cf. Equation (6.2)). So that, we start from the current $I^g(t)$ which curves are shown in Figure 5.3, then we compute the time-varying ganglion cell voltage $V^l(t)$ according to Equation (5.6). As for the gain control stage, we recourse to an Euler schema for the implementation of the LIF neurons. Examples of the resulting voltages $V^l(t)$, for different current magnitudes α , are shown in Figure 5.4. We can

notice that the more a neuron is stimulated, the quicker it fires and with a higher frequency [Adrian 1926, Van Rullen 2001b].

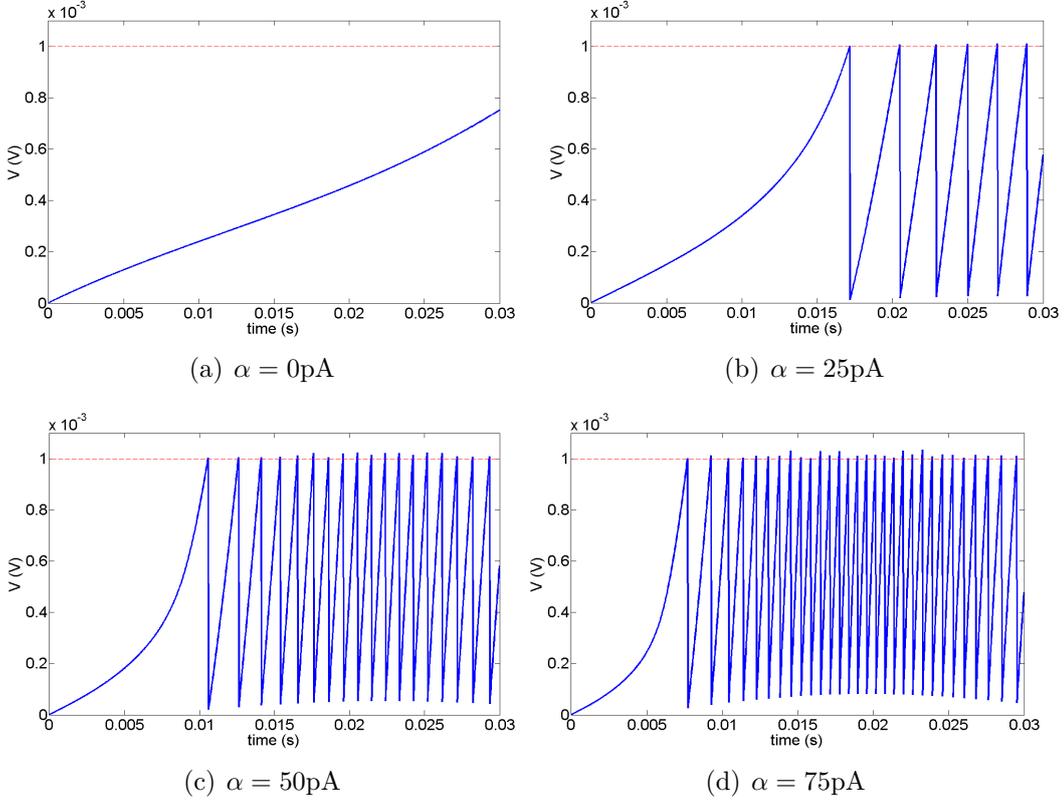


Figure 5.4: $V^l(t)$: Cascading the rapid contrast gain control and the spike generation stages for different values of magnitude α , such that $I^{opl}(t) = \alpha \mathbb{1}_{\{t \geq 0\}}(t)$. In solid line, the time behavior of the ganglion cell voltage $V^l(t)$. In dotted line, the spike emission threshold δ . $\delta = 110^{-3}V$, $g^l = 5 \cdot 10^{-9}S$, $c^l = 10^{-10}F$, $V_0^l = 0V$.

Based on the above results, we can infer the temporal maps associating a spike count to an observation time for a given value of α . We denote these maps by $n^\alpha(t)$. Example maps are shown in Figure 5.5(a). Our goal is to get an instantaneous picture of the bio-inspired system implemented at different observation timings. Given an observation time t_{obs} , this amounts to estimate the function $n^{t_{obs}}(\alpha)$. We can obtain this function starting from the curves of $n^\alpha(t)$ for different values of α , by means of a transversal cut in Figure 5.5(a). The spike count map $n^{t_{obs}}(\alpha)$ defined over the set of possible input magnitudes α is shown in Figure 5.5(b). It appears that, across time, the bio-inspired quantizer is evolving (i) from coarse to fine and (ii) from uniform to non-uniform.

The A/D converter that we end up with is time-dependent. Indeed, to each observation time t_{obs} we associate a quantizer $n^{t_{obs}}$ that maps a real magnitude value α into a quantum value $n^{t_{obs}}(\alpha) = |(t^i)_{t_i < t_{obs}}|$. Once we defined our time-dependent A/D converter, we have to design the corresponding decoding procedure. This is detailed in the following section.

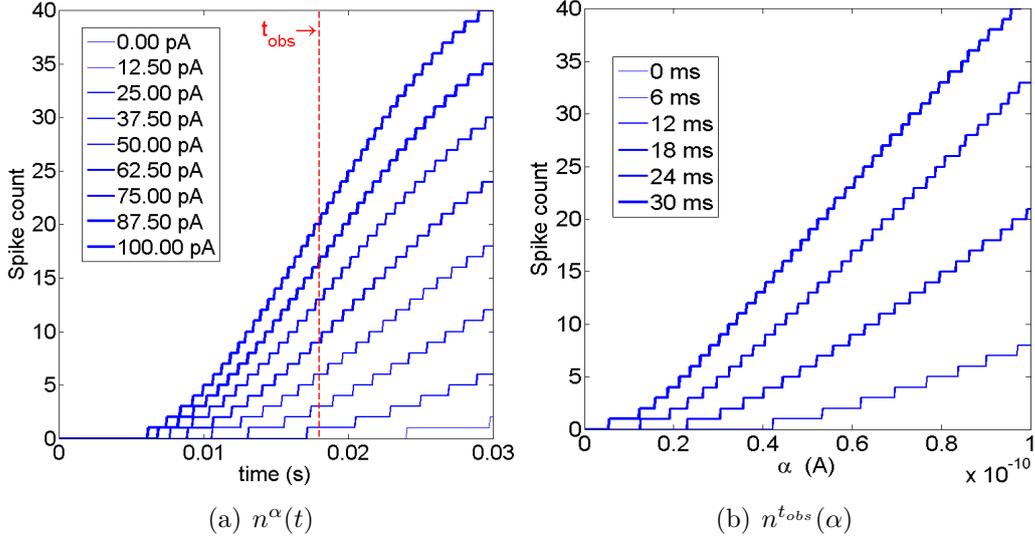


Figure 5.5: Behavior evolution of our retina-inspired A/D converter. 5.5(a): The abscissa axis represents the time. The ordinate axis represents the spike count $n^\alpha(t)$. 5.5(b): Transversal cuts in 5.5(a). The abscissa axis represents the magnitude of the step input α . The ordinate axis represents the spike count $n^{t_{obs}}(\alpha)$. $\delta = 1 \cdot 10^{-3} V$, $g^l = 5 \cdot 10^{-9} S$, $c^l = 10^{-10} F$, $V_0^l = 0V$.

5.2.3 The decoding pathway

Starting from the quantization scheme defined above, we describe in this section a straightforward decoding scheme. Our aim is to recover $\hat{\alpha}^{t_{obs}}$, the estimation of the step input magnitude α at the observation time t_{obs} . For a given observation timing t_{obs} , the only data that is transmitted to the decoder is the generated spikes count $n^{t_{obs}}(\alpha)$. All the model parameters are known a priori by both the coder and the decoder.

The decoding map, which we denote $\hat{n}^{t_{obs}}$ is a direct reverse look up table of the coding map. In order to be able to decode the input signal, we first compute the coding maps off-line as specified in 5.2. Then the coding maps axis are permuted. Figure 5.6 shows examples of decoding functions $\hat{\alpha}^{t_{obs}}$ at different observation timings from 0 to 30 ms. The reader will notice that the

decoding map $\hat{n}^{t_{obs}}$ is a set of unconnected points as the decoding map starts from \mathbb{N} , the set of possible spike counts, to \mathbb{R} , the set of decoded current values.

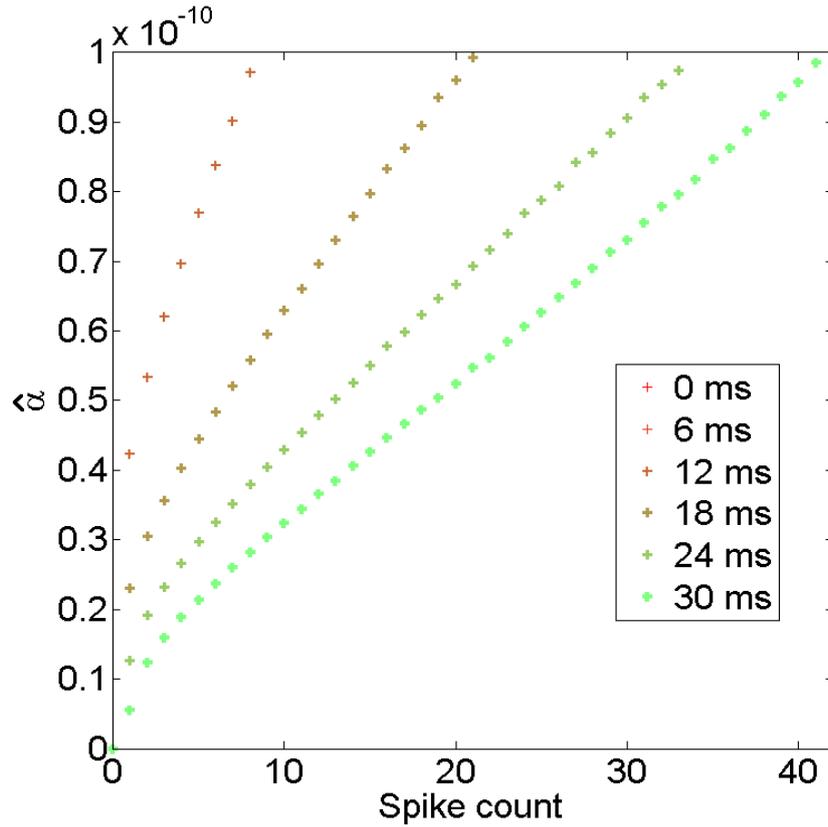


Figure 5.6: Time evolution of the decoding map. The decoder is time-dependent and evolves in the same manner as the coding map. The resulting function is shown at 0 ms, 6 ms, 12 ms, 18 ms, 24 ms, and 30 ms.

Though the coding scheme in Section 5.2 is strongly related to actual biological retina behavior, we do not claim that the proposed decoding algorithm is the one that is actually employed in the visual cortex. This being said, we have now a complete specification of a bio-inspired quantization scheme and the corresponding decoding scheme. So that in the next section, we will study our scheme behavior and compare it to standard A/D converters.

5.3 Our retina-inspired A/D converter *vs* standard A/D converters

The A/D converter that we presented relies on a biologically realistic model of the retina. The coding procedure assumes the spike count to be the coding feature in a spike train. Our retina-inspired A/D converter offers interesting properties as (i) the time-dependent gain amplification, (ii) the time-dependent non-linear rectification and (iii) the scalability of the quantizer/decoder. In the following, we first exhibit the overall behavior of our system and then we compare it to the standard quantizers and A/D converters.

5.3.1 Our A/D converter overall behavior

In order to show the behavior our bio-inspired quantization/decoding scheme, we compute the characteristic function of it, ie the map that associates $\hat{\alpha}^{t_{obs}}$ to α for a given observation timing t_{obs} . Figure 5.7 shows the evolution of our A/D converter characteristic function across time. The resulting map is a mid-tread quantizer with a central dead zone. As expected, the behavior of the quantizer evolves according to the two laws below.

- The retina inspired A/D converter evolves from coarse to fine: The number of quantizing steps for the range of magnitude values in $[0, 100]$ pA increases across time (see also Figure 5.5). This denotes a time-dependent refining process in the retina. The refining is intuitive as the visual cortex perceives roughly the stimulus first, then as time goes, reconstruct it more accurately: This mechanism enables scalability.
- The retina inspired A/D converter evolves from uniform to non-uniform: The retinal quantizer is non-uniform. High magnitude signals are mapped accurately, by a small quantization step, while small magnitude signals are coarsely rendered. This is due to the non-linear rectification in the IPL stage. Indeed, this rectification compresses the dynamic range of small magnitude signals around zero and spans higher ones in a linear fashion, this before the generation of spikes in the ganglion cells. This tendency to non-uniformity is accentuated as the gain control gets

higher across time: This phenomenon results from the time-dependent gain control and non-linear rectification.

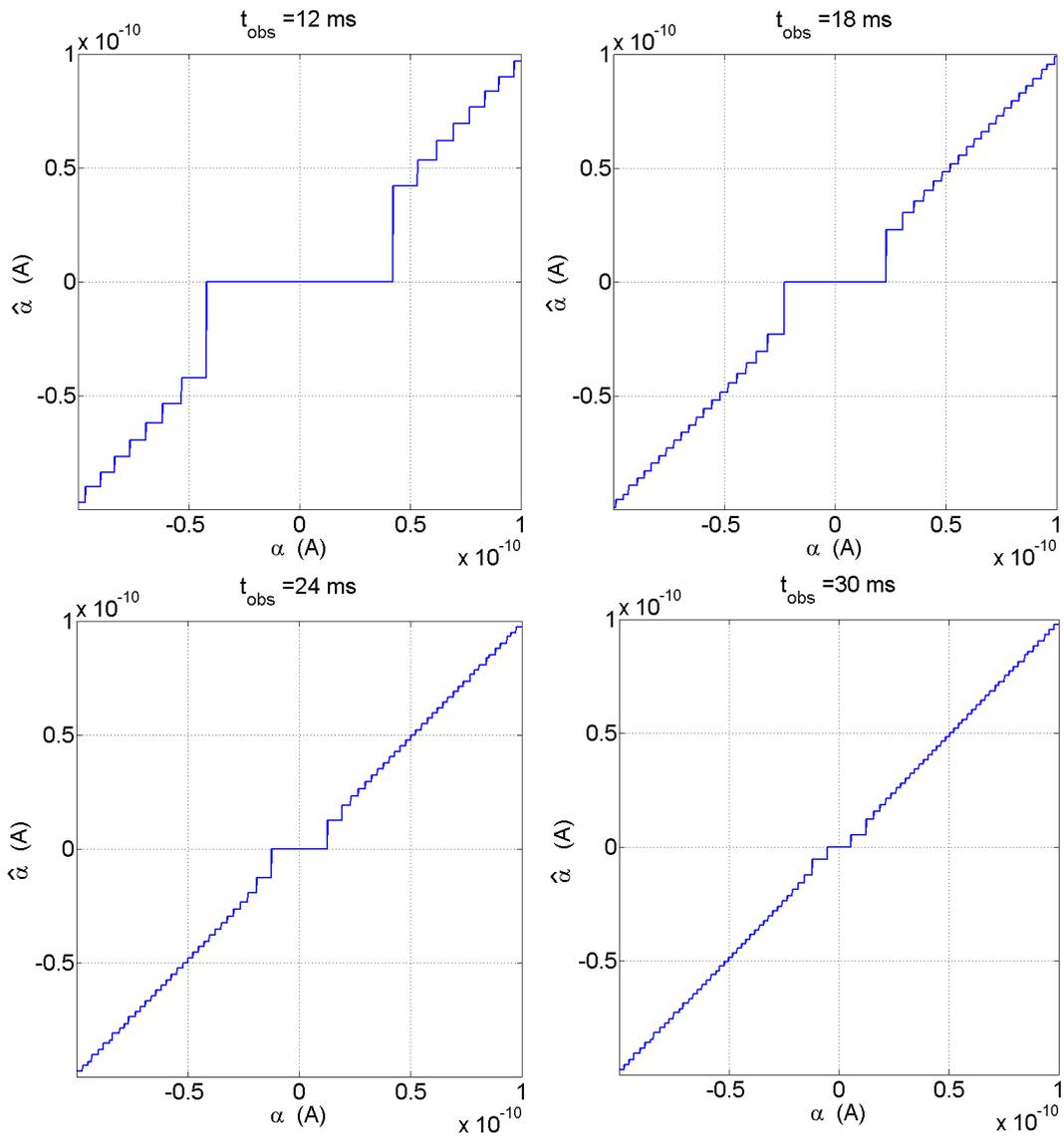


Figure 5.7: Time-evolution of the characteristic function of our retina-inspired A/D converter. The abscissa represents the magnitude of the step input current α . The ordinate represents the decoded value $\hat{\alpha}^{t_{\text{obs}}}$. The resulting function is shown at 12, 18, 24 and 30 ms.

5.3.2 A beyond the standards A/D converter

Interestingly, the behavior of the A/D converter that we specified is similar to several already existing devices. Figure 5.8 raises the analogy between our retina-inspired A/D converter and the standard A/D converters. Indeed, a standard A/D converter encompasses also three stages, namely (i) a gain controller, (ii) a non-linear rectifier and (iii) an uniform quantizer.

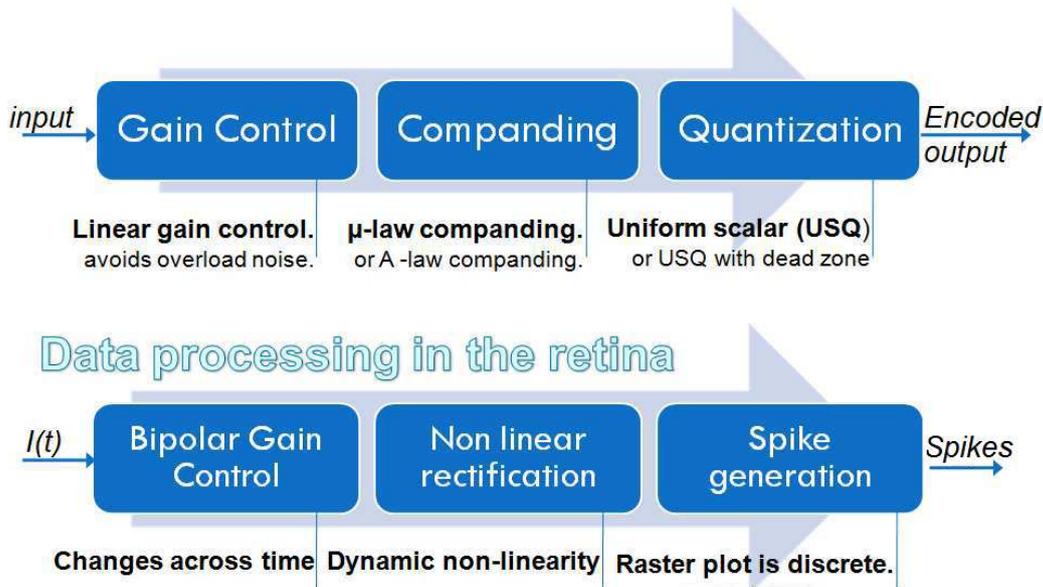


Figure 5.8: The analogy between the deep retina layers and the standard A/D converters.

Original time-dependent gain amplifier/compander

In the literature, the non-linear rectifiers are commonly referred to as companding circuits [Clark 1928]. Companding is a technique that is widely used in telecommunications to make the quantization steps unequal. This technique has an analogous functional role to what the IPL stage does in our case. It is also interesting to denote that, for audio recordings, a compander is generally preceded by a variable-gain amplifier which is locally linear. The functional role of this amplifier is also analogous to the one of the bipolar cells in our case. In the following, we show the effects of each layer on the input current $I^{opl}(t)$, this at different observation timings t_{obs} .

First, we start from the study of $V^b(t)$ which curves are shown in Figure 5.2. We then perform a transversal cut to get the maps $V_{t_{obs}}^b(\alpha)$ that

associate a bipolar voltage to a stimulus magnitude α at a given observation time t_{obs} . Figure 5.9(a) schematizes the transversal cut performed. The resulting maps are shown in Figure 5.9(b). These maps describe the time evolution of the bipolar cells behavior as a function of the input magnitude. These results highlights the fact that the contrast gain control in the retina is linear for short latencies (cf. $t \leq 20$ ms), but then it saturates leading to a flat map (cf. $t > 20$ ms). We retrieve as in consumer electronic devices a variable-gain linear amplifying system, that also takes into account a saturation bound.

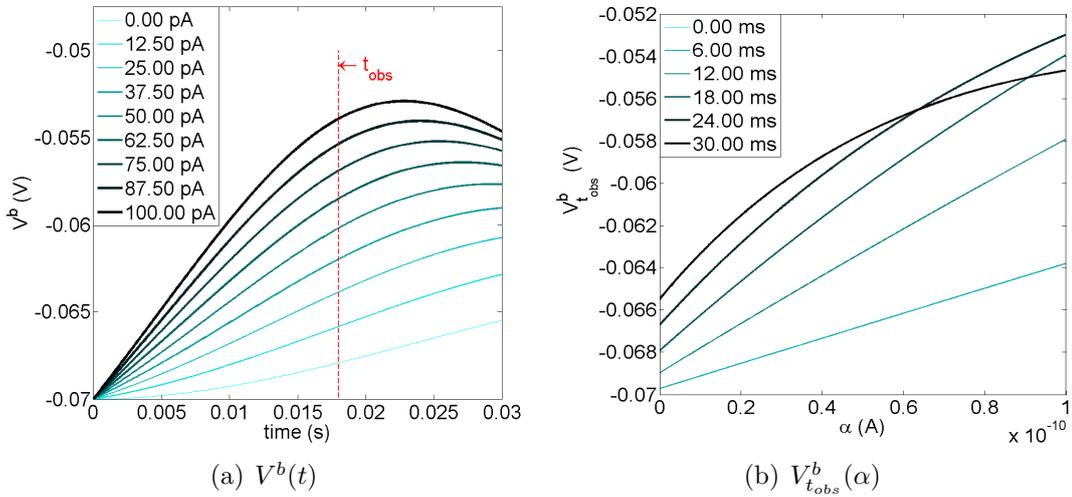


Figure 5.9: Behavior evolution of the gain control in the bipolar cells. 5.9(a): The abscissa axis represents the time. The ordinate axis represents the time-varying bipolar voltage $V^b(t)$. 5.9(b): Transversal cuts in 5.9(a). The abscissa axis represents the magnitude of the step input α . The ordinate axis represents the intensity-varying bipolar voltage $V_{t_{obs}}^b(\alpha)$.

Second, we cascade the non-linear rectification stage to the gain control one. We start from the study of the time behavior of the INL/IPL stages and infer their behavior as a function of the input magnitude α . The process that we employ is schematized in Figure 5.10(a) and is analogous to what we did for the INL stage study. We show in Figure 5.10(b) the resulting $I_{t_{obs}}^g(\alpha)$ maps that associate a ganglionic current to a stimulus magnitude α at a given observation time t_{obs} . We notice that the IPL non-linear rectification (i) compresses the low magnitudes dynamic range, and (ii) expands the high magnitudes one. This implies that the high magnitude inputs are well rendered while low magnitude ones are coarsely approximated.

The INL/IPL stages behavior in our case is in contradiction with the

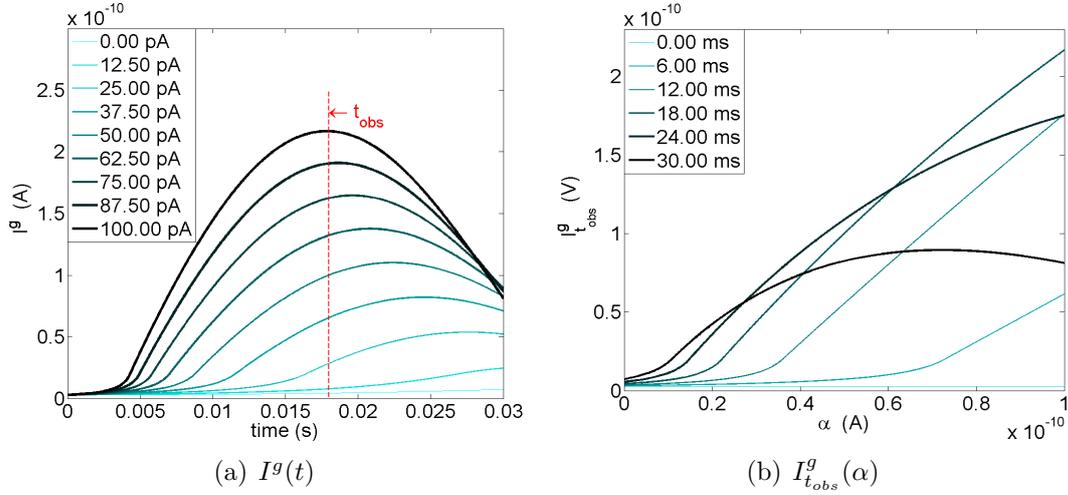


Figure 5.10: Behavior evolution of the gain control cascaded to the non-linear rectification in the IPL layer. 5.10(a): The abscissa axis represents the time. The ordinate axis represents the time-varying ganglionic current $I^g(t)$. 5.10(b): Transversal cuts in 5.10(a). The abscissa axis represents the magnitude of the step input α . The ordinate axis represents the intensity-varying ganglionic current $I_{t_{obs}}^g(\alpha)$.

traditional companding algorithms behavior as the A -law or the μ -law companders. Indeed, these algorithms work in a logarithmic fashion and emphasize low magnitudes and compress high magnitudes range. Figure 5.11 shows a comparison between the INL/IPL stages and the standard gain amplification/companding algorithms. Considering the above results, two main features make the INL/IPL stages go beyond the standards when compared to classical gain amplifiers/companders:

1. The INL/IPL stages emphasize high energy signals while the companders based on the μ -law emphasize high probability ones which have often the lowest energy.
2. The INL/IPL stages implement a time-dependent gain controller coupled to a time-dependent non-linear rectifier while the companders based on the μ -law have a static behavior.

Comparison to the Lloyd-Max quantizer

The originality of our approach arises also from its comparison to the standard of the quantization application: the Lloyd-Max quantizer [Garey 1982,

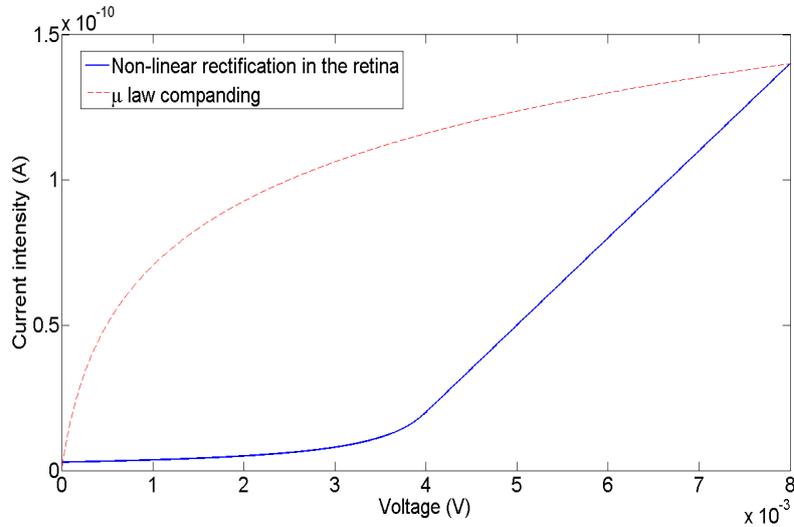


Figure 5.11: A comparison between the INL/IPL stages and the standard gain amplification/comparing algorithms. In solid line: the non-linear rectification used in the retina. In dotted line: the companding rescaled μ -law map.

Antonini 2002]. Figure 5.12 compares our retina-inspired A/D converter to the Lloyd-Max quantizer. The origin of these differences is that the retina-inspired A/D conversion is time-dependent and scalable while the Lloyd-Max quantizer is distribution-dependent and non-scalable by essence. The scalability of the retina-inspired is tuned by the only observation timing t_{obs} . Indeed, the choice of a given t_{obs} value implicitly imposes the choice of a quantizer and a decoder. The accuracy of our system increases proportionally to t_{obs} and this makes it scalable. On the contrary, if we would make the Lloyd-Max quantizer scalable, this would amount to tune the number n^q of quantization steps considered. Obviously for any value of n^q , the quantization process would require the off-line computation of the corresponding codebook. The Lloyd-Max quantizer thus defined is the optimal quantizer in terms of the rate/distortion behavior for a given distribution of the values to encode. The philosophy underlying the conception of the the Lloyd-Max quantizer makes it distribution-dependent and distribution-optimal. Thus, two major properties differentiate our retina-inspired A/D converter from the optimal quantizer:

1. The time dependency of our retina-inspired A/D conversion scheme allows it to be scalable. Indeed, the choice of the observation timing t_{obs} sets the level of accuracy of the quantizer/decoder.

2. Our retina-inspired A/D conversion offers scalability with no training process and no dictionary to transmit to the decoder.

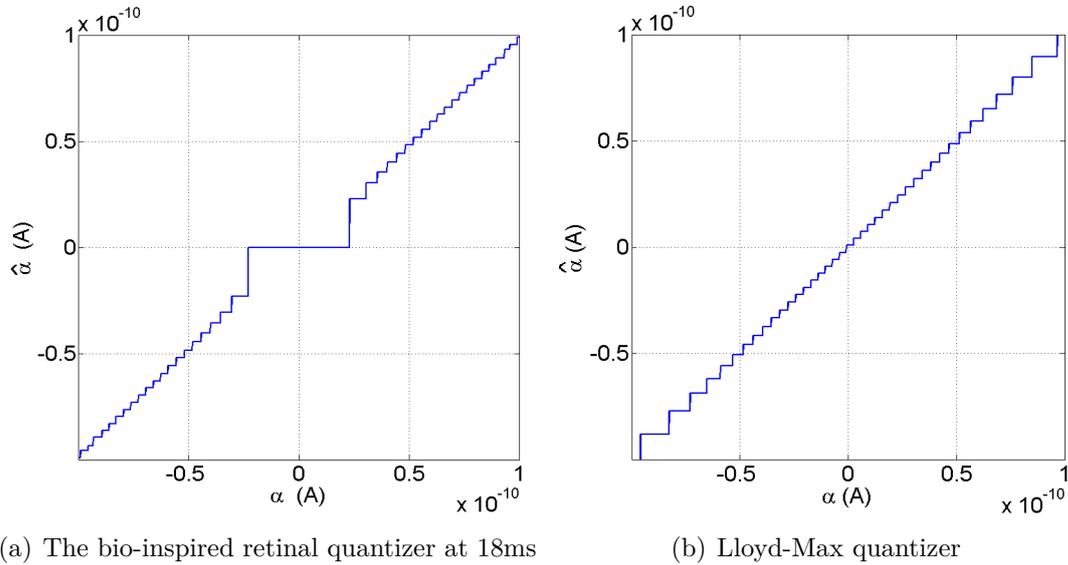


Figure 5.12: A comparison between our retina-inspired A/D converter and the standard Lloyd-Max quantizer. 5.12(a)The characteristic function of our retina-inspired A/D converter. 5.12(b)The characteristic function of a Lloyd-Max quantizer for a Gaussian source.

We specified an original quantizer/decoder mimicking the retina behavior. Though we restrained our study to the temporal behavior of the deep retina layers, we reproduced many mechanisms involved in the actual biological system. Interestingly, our quantizer/decoder behavior evolves across time (i) from coarse to fine and (ii) from uniform to non-uniform. As shown in this section, the time dynamics allow scalability. Besides, the philosophy underlying is in contradiction with the traditional Lloyd-Max quantizers because (i) it is distribution-independent (ii) it renders high magnitudes precisely while it renders low magnitudes coarsely.

5.4 The dithering hypothesis

In the preceding sections, we showed the behavior of the deep retina seen as an A/D converter. Despite being time-dependent, this retinal quantizer evolves

according to a deterministic law. Though, one major issue encountered by neuroscientists is the non-determinism of the retinal neural code. Indeed, given a single visual stimulus, spikes timings in the retina output are not exactly reproducible across trials. Yet, no clear evidence is established about the phenomena at the origin of this trial-to-trial variability. Several hypotheses were discussed in the literature and yielded two different points of view. The first hypothesis is that the precise timings of individual spikes convey a large amount of information [Perkell 1968b, Panzeri 2001]. This hypothesis suggests that the stimulus coding process in the retina is deterministic and reports detailed information about the stimulus with a high temporal fidelity. In this case, each single spike timing makes sense. The second hypothesis is that only a few statistical quantities measured over the spike-based output convey the relevant information about the stimulus to the visual cortex [Brown 2004]. For instance, since [Adrian 1926] it was widely assumed that the variable spike patterns corresponding to a single stimulus are random instantiations of a desired firing rate. In this case, the precise timing of each single spike may not be meaningful and thus spikes may carry some amount of noise. The spike based-output should then be averaged to reveal meaningful signals [de Ruyter van Steveninck 1997].

The role of spikes timings variability in the neural code of the retina is still an open issue and no clear evidence establishes whether this variability conveys precise information or random noise [Shadlen 1998]. Here, we make the proposal that the non-determinism in the retinal processing prior to the ganglionic layer yields a dither noise [Masmoudi 2010a, Masmoudi 2010c]. This noise, while corrupting the input of the ganglionic layer by a completely random signal, brings interesting features to the spike-based output of the retina. For this to be possible, the distribution of the noise that we introduce obeys specific constraints defined in [Wannamaker 2000]. Obviously, the dither noise hypothesis is one possible assumption among several others and we do not claim its biological exactness. Still, our present effort aims at bridging the differences between the different points of view reported above by exploring the hypothesis of a “retinal useful noise”.

Commonly, the dithering consists in adding a noise to a given input before passing it through an uniform scalar quantizer. In our case, the quantizer user

is the LIF neuron. We can prove experimentally that this neuron, under the assumption of a rate code, is an uniform scalar quantizer. Figure 5.13 shows the behavior of the LIF neuron when used in our A/D converter with no prior gain control and non-linear rectification. The results demonstrate that the LIF neuron could be considered as an uniform scalar quantizer.

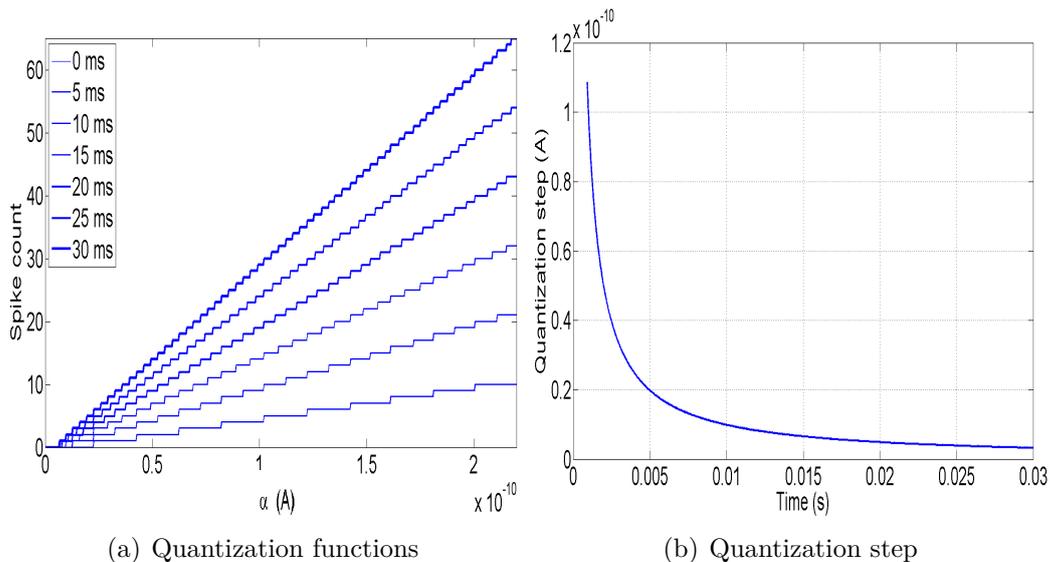


Figure 5.13: The LIF neuron seen as an uniform scalar quantizer. Figure 5.13(a): The LIF neuron quantization maps at different observation timings t_{obs} . Figure 5.13(b): Time evolution of the mean quantization step of the LIF neuron.

Two implementations of the dithering process are possible. The first one is the subtractive dithering (SD) and the second is the non-subtractive dithering (NSD). Theoretically, SD can add interesting features to the quantizing system without degrading its signal-to-noise ratio. Unfortunately, SD requires the subtraction of the exact noise signal from the quantizer output before decoding. So that, the subtractive dithering supposes that the exact values of the noise signal are transmitted to the decoder. As a consequence, the hypothesis that we study is that the retina is a non-subtractive dithered system (NSD).

Up to our knowledge, little have been done to explicit the probability distribution of such a noise. In the literature, it is generally and empirically assumed that the retinal noise η is Gaussian [Wohrer 2008]. Thus, we refer to the results established in [Wannamaker 2000], and recall the following funda-

mental theorem of dither noise distribution for the case of an uniform scalar quantizer:

Theorem 1. *The choice of zero-mean dither probability distribution function (pdf) which renders the first and second moments of the total error independent of the input, such that the first moment is zero and the second is minimized, is unique and is a triangular pdf of 2 LSB peak-to-peak amplitude.*

Thus, we suppose that (i) η^l magnitude (cf. Equation (5.6)) has a triangular probability distribution function across the possible locations in an image, and (ii) that the dynamic range of η^l is twice wider than the quantization step of the considered ganglion cell. Having these two conditions we verify the theorem. Under the restriction of these hypotheses correctness, we identify the retinal noise η to a dither signal. As we do not subtract the dither signal in the decoding process, our coder is an NSD. Although not intuitive, adding such a random dither signal to the input stimulus allow the quantizer to have interesting features. Mainly, the quantization error values $\varepsilon = (\alpha - \hat{\alpha}^{t_{obs}})$ and the input stimulus magnitude α are de-correlated. This feature is clearly demonstrated when computing the cross correlation between ε and α as shown in the Figures 5.14(a) and 5.14(b) for the test image Lena. Besides, quantization error is whitened so that error is uniformly distributed over the stimulus spectrum. Figures 5.14(c) 5.14(d) show a comparison between the spectra of the ganglion cell quantizer with and without NSD.

The whitening and de-correlation features yield a greater reconstruction error in terms of mean squared error [Wannamaker 2000]. Though, the visual quality of the reconstruction \tilde{I} is better when using a dithered system. Figure 5.15 shows the great impact of an NSD ganglion cell quantizer when compared to a non-dithered one, for the same observation time t_{obs} . Note that, the number of quantization steps in a dithered and a non-dithered coder is roughly the same, but the distribution of the quantization values varies greatly. This gives us an idea on what could be transmitted by the non-deterministic retinal coder and its impact on the stimuli reconstruction.

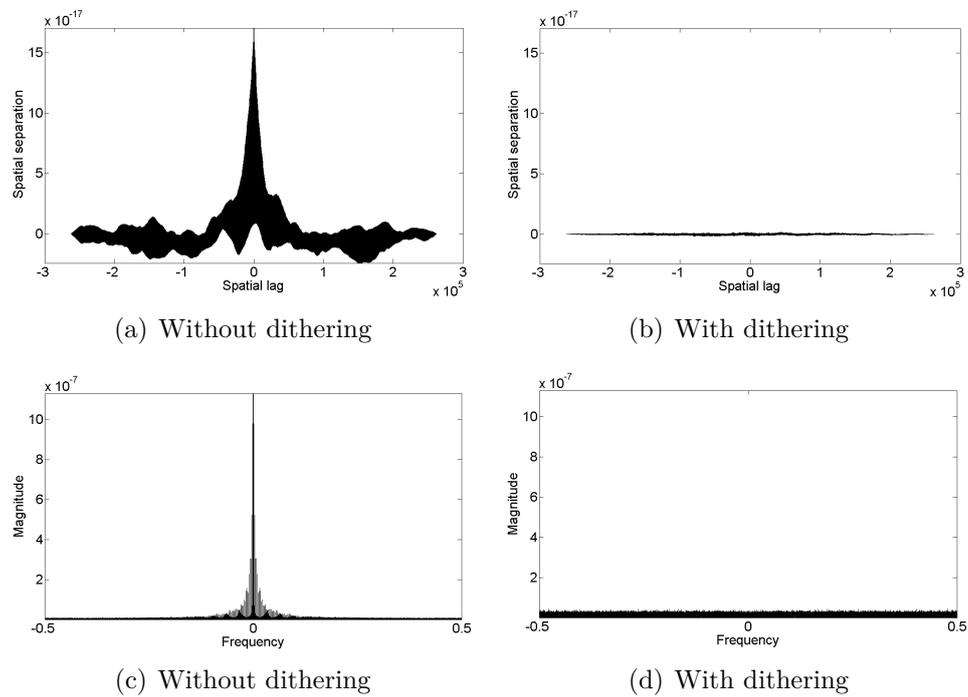


Figure 5.14: 5.14(a) 5.14(b): Cross correlation of the quantization error and the input stimuli. The abscissa represents the spatial lag and the ordinate the cross correlation magnitude. 5.14(c) 5.14(d): Noise whitening using a dithered quantizing ganglion cell: A comparison of reconstruction error spectra between non-dithered and dithered quantizing ganglion cell. The test image is Lena. The observation time is $t_{obs} = 55ms$.



Figure 5.15: Comparison of the reconstruction visual quality between non-dithered and dithered quantizing ganglion cell. The test image is Lena. The observation time is $t_{obs} = 55ms$.

5.5 Discussion

We presented a bio-inspired quantizer/decoder mapping the retina behavior. The model of the retina that we adopted, though restrained to its temporal aspect, reproduces many mechanisms involved in the actual biological system. Our quantizer behavior evolves dynamically, and thus, it permits scalability as it goes from coarse to fine across time. Interestingly, the quantizer evolves also from uniform to non-uniform, but in contradiction with traditional Lloyd-Max quantizers, as it renders high magnitudes precisely while it maps low magnitudes coarsely. Besides, we emitted a biologically plausible hypothesis that supposes the retinal noise distribution to have specific characteristics, yielding the definition of a non-subtractive dithered system. We do not claim that the retinal noise is a dither signal, but still such an hypothesis is seducing by the noise whitening and de-correlation features it allows. Our future work aims at adding several mechanisms of the retinal processing that are not taken into account in the current quantizer/decoder. The spatial filtering and lateral inhibitions are two examples of features that could be integrated in an upcoming effort. A long term goal would be to infer, starting from a sufficiently realistic model, a decoding algorithm that could decipher actual neural recordings.

STREAMING AN IMAGE THROUGH THE RETINA: A NOVEL DITHERED AND SCALABLE IMAGE CODER

Contents

| | | |
|------------|---|------------|
| 6.1 | The coding pathway | 140 |
| 6.1.1 | The image transform: The outer layers of the retina | 140 |
| 6.1.2 | The A/D converter | 143 |
| 6.2 | The decoding pathway | 146 |
| 6.3 | Results: The noiseless case | 148 |
| 6.4 | Introducing the noise in the coder | 151 |
| 6.4.1 | Study context | 151 |
| 6.4.2 | Introducing the multiscale NSD | 152 |
| 6.5 | Results: The dithered case | 156 |
| 6.6 | Conclusion | 160 |

Keywords

Static image compression, bio-inspired signal coding, retina, spiking neuron, A/D conversion, dithering, scalability.

Overview

We propose the design of an original scalable image coder/decoder that is inspired from the mammalians retina. Our coder accounts for the time-dependent and also non-deterministic behavior of the actual retina. This chapter is based mainly on the results that we published in [Masmoudi 2012c] and [Masmoudi 2011].

Contribution

The present work brings two main contributions. In a first step, we design a deterministic image coder mimicking most of the retinal processing stages. Then in a second step, we introduce a retinal noise in the coding process, that we model here as a multiscale dither signal. Concerning the first contribution, we specify an original model of the retinal transform through the introduction of time delays. Then, we cascade the specified transform to the A/D converter defined in the previous chapter to get a complete, original and scalable coder/decoder. Concerning the second contribution, the dithering process introduced is original since it takes into accounts the multiscale nature of the prior transform. Besides, it enables (i) reconstruction error whitening, (ii) reconstruction error decorrelation from the input, and (iii) a faster recognition of the fine details of the image during the decoding process.

Organization

This chapter is organized into two part:

1. The first part consists of the Sections 6.1 to 6.3 and presents the design of our bio-inspired scalable image coder/decoder with a deterministic behavior.
2. The second part consists of the Sections 6.4 to 6.5. In Section 6.4, we show how we integrated the dithering process in our coder/decoder. Then, in Section 6.5, we detail the perceptual impact of it.

We propose the design of an original scalable image coder/decoder that is inspired from the mammalian retina. Our coder accounts for the time-dependent and also non-deterministic behavior of the actual retina. The present work brings two main contributions. The first one is the design of a deterministic image coder/decoder mimicking most of the retinal processing stages. The second contribution is the introduction of a retinal noise in the coding process. Here, we model this noise by a multiscale dither signal in order to gain interesting perceptual features. Regarding our first contribution, our main source of inspiration will be the retina models introduced in [Van Rullen 2001b] and [Wohrer 2009a]. The coder that we propose has two stages. The first stage is an image transform which is performed by the outer layers of the retina. Here we specify an original model for this transform. We reproduce the outer layers behavior by filtering the image with a bank of difference of Gaussians with time-delays. The second stage is a time-dependent analog-to-digital conversion which is performed by the inner layers of the retina as specified in Chapter 5. The main novelty of this coder is to show that the time-dependent behavior of the retina cells could ensure, in an implicit way, scalability and bit allocation. Regarding our second contribution, we reconsider the inner layers of the retina. We emit a possible interpretation for the non-determinism observed by neurophysiologists in their output. For this sake, we model the retinal noise that occurs in these layers by a dither signal. The dithering process introduced here is an extension to what we presented in Chapter 5 in the sense that it accounts for the multiscale nature of the transform. The dithering process that we propose adds several interesting features to our image coder. The dither noise whitens the reconstruction error and decorrelates it from the input stimuli. Furthermore, integrating the dither noise in our coder allows a faster recognition of the fine details of the image during the decoding process. The present chapter goal is twofold. First, we aim at mimicking as closely as possible the retina for the design of a novel image coder while keeping encouraging performances. Second, we bring a new insight concerning the non-deterministic behavior of the retina.

This chapter is organized into two parts. The first part consists of the Sections 6.1 to 6.3 and presents the design of our bio-inspired scalable image coder/decoder with a deterministic behavior. This part is organized as follows.

In Section 6.1, we detail the specification of a novel retina-inspired image coder. In Section 6.2, we present the decoding pathway. In Section 6.3, we show the main results that demonstrate the properties of our model. The second part consists of the Sections 6.4 to 6.5. In Section 6.4, we show how we integrated the dithering process in our coder/decoder. Then, in Section 6.5, we detail the perceptual impact of it. Finally, in Section 6.6, we summarize our main conclusions.

6.1 The coding pathway

The coding pathway is schematized in Figure 6.1. It follows the same architecture as the *Virtual Retina* model. However, since we have to define also a decoding pathway, we need to think about the invertibility of each processing stage. For this reason some adaptations are required and described in this section.

6.1.1 The image transform: The outer layers of the retina

In *Virtual Retina*, the outer layers were modelled by a non-separable spatio-temporal filtering. This processing produces responses corresponding to spatial or temporal variations of the signal because it models time-dependent interactions between two low-pass filters: this is termed center-surround differences. This stage has the property that it responds first to low spatial frequencies and later to higher frequencies. This *time-dependent frequency integration* was shown for *Virtual Retina* [Wohrer 2009b] and it was confirmed experimentally (see, e.g., [Sterling 1992]). This property is interesting since a large amount of the total signal energy is contained in the low frequencies subbands, whereas high frequencies bring further details. This idea already motivated bit allocation algorithms to concentrate the resources for a good recovery on lower frequencies.

However, it appears that inverting this non-separable spatio-temporal filtering is a complex problem [Wohrer 2009b, Zhang 2005]. To overcome this difficulty, we propose to model differently this stage while keeping its essential features. To do so, we decomposed this process into two steps. The first

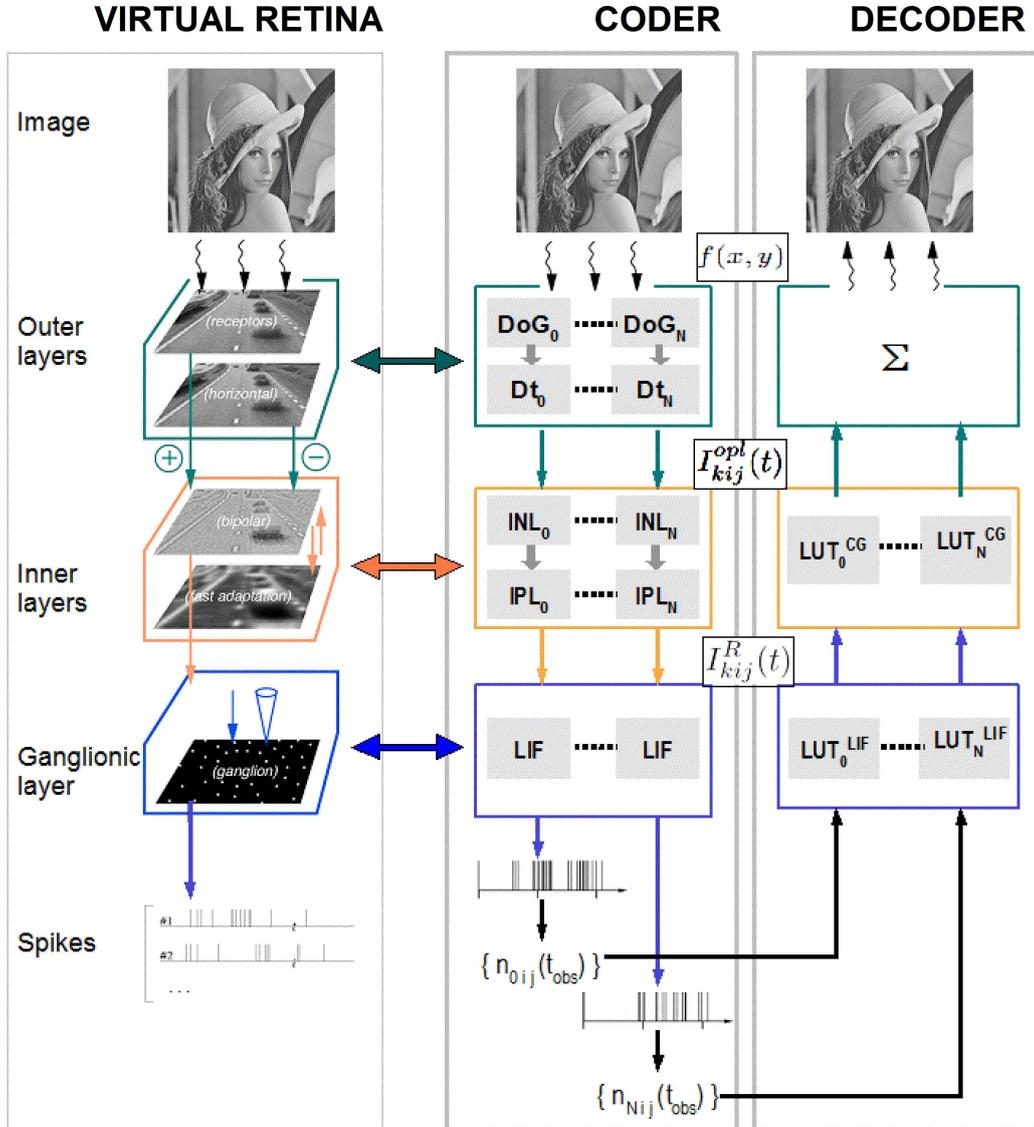


Figure 6.1: (a) Schematic view of the *Virtual Retina* model proposed by [Wohrer 2009a]. (b) and (c): Overview of our bio-inspired codec. Given an image, the static DoG-based multi-scale transform generates the subbands $\{F_k\}$. DoG filters are sorted from the lowest frequency-band filter DoG_0 to the highest one DoG_{N-1} . Each subband F_k is delayed using a time-delay circuit Dt_k , with $t_k < t_{k+1}$. The time-delayed multi-scale output is then made available to the subsequent coder stages. The final output of the coder is a set of spike series, and the coding feature adopted will be the spike count $n_{kij}(t_{obs})$ recorded for each neuron indexed by (kij) at a given time t_{obs} .

one considers only center-surround differences in the spatial domain (through differences of Gaussians) which is justified by the fact that our coder here gets static images as input. The second step reproduces the time-dependent frequency integration by the introduction of time-delays.

Center-surround differences in the spatial domain: The DoG model

Neurophysiologic experiments have shown that, as for classical image coders, the retina encodes the stimulus representation in a transform domain. In a first step, we will consider the model of the retinal transform defined in [Van Rullen 2001b] that we detailed in Chapter 2. In this model, the outer layers cells are implemented by a difference of two Gaussian filters. We recall that according to this model, the DoG cells can be arranged in a dyadic grid to sweep all the stimulus spectrum as schematized in Figure 6.2(a). Each layer k in the grid, is tiled with DoG_k cells having a scale k and generating a transform subband F_k , where $\sigma_{s_{k+1}} = \frac{1}{2}\sigma_{s_k}$ and $\sigma_{c_{k+1}} = \frac{1}{2}\sigma_{c_k}$. So, in order to measure the degree of activation c_{kij} of a given DoG_k cell at the location (i, j) with a scale k , we compute the convolution of the original image f by the DoG_k filter:

$$c_{kij} = \sum_{x,y=-\infty}^{\infty} DoG_k(i-x, j-y) f(x, y). \quad (6.1)$$

This transform generates a set of $(\frac{4}{3}N^2 - 1)$ coefficients for an N^2 -sized image. An example of such a bio-inspired multi-scale decomposition is shown in Figure 6.2(b). Note here that we added to this bank of filters a Gaussian low-pass scaling function that represents the state of the OPL filters at the time origin (see Chapter 3 for more details).

Integrating time dynamics through time-delay circuits

Of course, the model described in (6.1) has no dynamical properties. In the actual retina, the surround G_{σ_s} in Equation (2.4) appears progressively across time driving the filter passband from low frequencies to higher ones. Our goal is to reproduce this phenomenon that we called time-dependent frequency integration. To do so, we added in the coding pathway of each subband F_k a time-delay circuit D_{t_k} . The value of t_k is specific to F_k and is an increasing

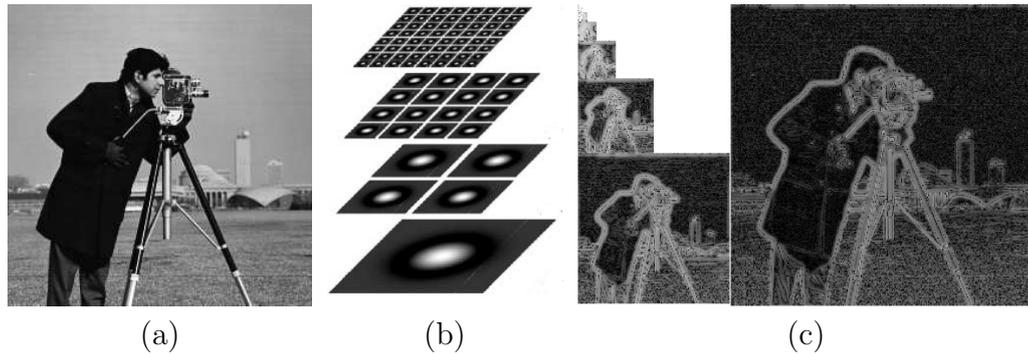


Figure 6.2: (a) Input image cameraman. (b) Example of a dyadic grid of DoG's used for the image analysis (from [Van Rullen 2001b]). (c) Example on image (a) of DoG coefficients generated by the retina model (the subbands are shown in the logarithmic scale)

function of k . The t_k delay causes the subband F_k to be transmitted to the subsequent stages of the coder starting from the time t_k . The time-delayed activation coefficient $I_{kij}^{opl}(t)$ computed at the location (i, j) for the scale k at time t is now defined as follows:

$$I_{kij}^{opl}(t) = c_{kij} \mathbb{1}_{\{t \geq t_k\}}(t), \quad (6.2)$$

where $\mathbb{1}_{\{t \geq t_k\}}$ is the indicator function such that, $\mathbb{1}_{\{t \geq t_k\}}(t) = 0$ if $t < t_k$ and 1 otherwise. While in our previous work [Masmoudi 2011] t_k is increasing linearly as a function of k , we changed the law governing t_k to an exponential one with a time constant denoted by τ^{opl} [Masmoudi 2012c]. This change is intended to bring more biological plausibility to our new coder as the time behavior of the outer layers cells is exponential [Field 1994, Wohrer 2009a]. Indeed, in the actual retina, the passband of the DoG cells runs through the low frequencies at a fast pace, then decelerates in an exponential fashion. So, the time-dependent frequency integration is not a linear phenomenon. The evolution of time delays t_k with respect to the scale k , in the present work, is detailed in Figure 6.3.

6.1.2 The A/D converter: The inner and ganglionic layers of the retina

The retinal A/D converter is defined based on the processing occurring in the inner and ganglionic layers, namely a contrast gain control, a non-linear recti-

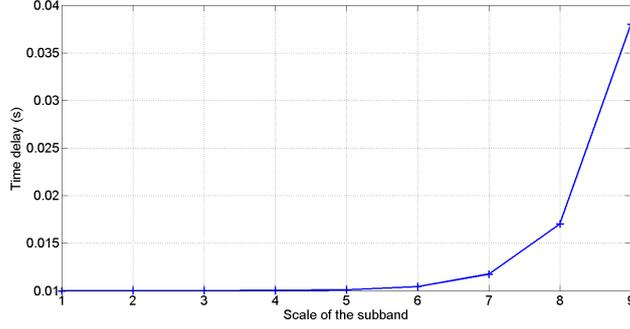


Figure 6.3: Time delays D_{t_k} introduced in the coding process. The time-dependent frequency integration is reproduced by delaying the coding process start of the subband F_k by t_k . The series t_k is represented as a function of the scale k . The progression law is exponential with a time constant $\tau^{opl} = 65 \text{ ms}$.

fication and a discretization based on LIF neurons (see Chapter 5). A different treatment will be performed for each delayed subband, and this produces a natural bit allocation mechanism. Indeed, as each subband F_k is presented at a different time t_k , it will be subject to a transform according to the state of our dynamic A/D converter at t_k .

6.1.2.1 Contrast gain control

The retina adjusts its operational range to match the input stimuli magnitude range. We described this stage in detail in Chapter 5. Given the scalar magnitude c_{kij} of the input step current $I_{kij}^{opl}(t)$, the contrast gain control is a non-linear operation on the potential of the bipolar cells. This potential varies according to both the time and the magnitude value c_{kij} ; and will be denoted by $V_{kij}^b(t, c_{kij})$. This phenomenon is modelled as described in Equation (5.2). We recall the model equation, with a slight adaptation made to the notations, as follows:

$$\begin{cases} c^b \frac{dV_{kij}^b(t, c_{kij})}{dt} + g^b(t)V_{kij}^b(t, c_{kij}) = I_{kij}^{opl}(t), & \text{for } t \geq 0, \\ g^b(t) = E_{\tau^b} \ast Q(V_{kij}^b(t, c_{kij})), \end{cases} \quad (6.3)$$

where $Q(V_{kij}^b) = g_0 + \lambda^b (V_{kij}^b(t))^2$ and $E_{\tau^b} = \frac{1}{\tau^b} \exp\left(-\frac{t}{\tau^b}\right)$, for $t \geq 0$. Figure 6.4(a) recalls the time behavior of $V_{kij}^b(t, c_{kij})$ for different magnitude values c_{kij} of $I_{kij}^{opl}(t)$.

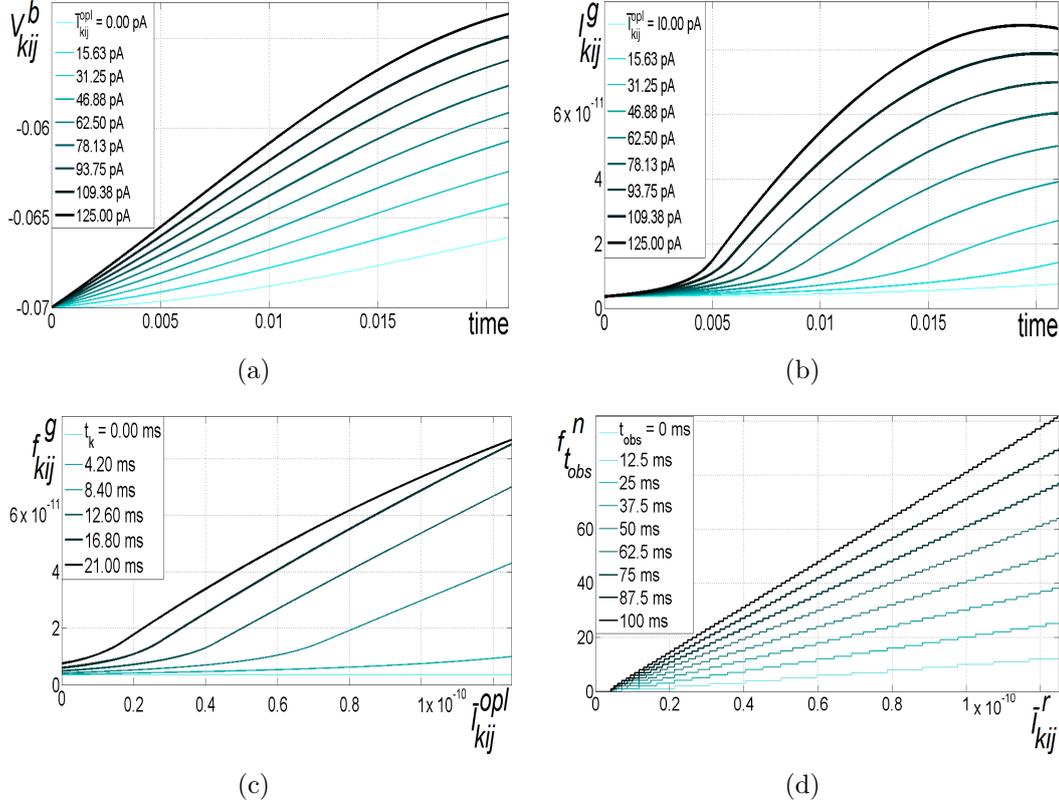


Figure 6.4: **6.4(a)**: $V_{kij}^b(t)$ as a function of time for different values of c ; **6.4(b)**: I_{kij}^g as a function of time for different values of c ; **6.4(c)**: The functions f_{kij}^g that map c_{kij} into I_{kij}^g for different values of t_k ; **6.4(d)**: The functions $f_{t_{obs}}^n$ that map \bar{I}_{kij}^r into n_{kij} for different values of t_{obs}

6.1.2.2 Non-linear rectification

In the next processing step, the potential $V_{kij}^b(t, c_{kij})$ is subject to a non-linear rectification yielding the so-called ganglionic current $I_{kij}^g(t, c_{kij})$. This phenomenon is modelled as described in Equation (5.3). For a constant scalar value c_{kij} , and with the notations adopted in this chapter, $I_{kij}^g(t, c_{kij})$ is defined as follows:

$$I_{kij}^g(t, c_{kij}) = N \left(T_{w^g, \tau^g}(t) * V_{kij}^b(t, c_{kij}) \right), \quad \text{for } t \geq 0,$$

where w^g and τ^g are constant scalar parameters, T_{w^g, τ^g} is a linear transient filter, and N is defined as in Equation (5.4). Figure 6.4(b) recalls the time behavior of $I_{kij}^g(t, c_{kij})$ for different values of c_{kij} .

As the currents c_{kij} are delayed with times $\{t_k\}$, our goal is to catch the

instantaneous behavior of the inner layers at these times $\{t_k\}$. This amounts to infer the transforms $I_{t_k}^g(c_{kij})$ that maps a given scalar magnitude c_{kij} into a rectified current \bar{I}_{kij}^r as the modelled inner layers would generate it at t_k . To do so, we start from the time-varying curves of $I_{kij}^g(t, c_{kij})$ in Figure 6.4(b) and we do a transversal cut at each time t_k . We show in Figure 6.4(c) the resulting maps $f_{t_k}^g$ such that $I_{kij}^g(t_k, c_{kij}) = f_{t_k}^g(c_{kij})$.

As for $I_{kij}^{opl}(t)$ (see Equation (6.2)), we introduce the time dimension using the indicator function $\mathbb{1}_{\{t \geq t_k\}}(t)$. The final output of this stage is the set of step functions $I_{kij}^r(t)$ defined by:

$$I_{kij}^r(t) = \bar{I}_{kij}^r \mathbb{1}_{\{t \geq t_k\}}(t), \quad \text{with } \bar{I}_{kij}^r = f_{t_k}^g(c_{kij}). \quad (6.4)$$

6.2 The decoding pathway

The decoding pathway is schematized in Figure 6.1(c). It consists in inverting, step by step, each coding stage described in Section 6.1. At a given time t_{obs} , the coding data is the set of $(\frac{4}{3}N^2 - 1)$ spike counts $n_{kij}(t_{obs})$, this section describes how we can recover an estimation $\tilde{f}_{t_{obs}}$ of the N^2 -sized input image $f(x, y)$. Naturally, the recovered image $\tilde{f}_{t_{obs}}(x, y)$ depends on the time t_{obs} which ensures time-scalability: the quality of the reconstruction improves as t_{obs} increases. The ganglionic and inner layers are inverted using look-up tables constructed off-line and the image is finally recovered by a direct reverse transform of the outer layers processing.

Recovering the input of the ganglionic layer:

First, given a spike count $n_{kij}(t_{obs})$, we recover $\tilde{I}_{kij}^r(t_{obs})$, the estimation of $I_{kij}^r(t_{obs})$. To do so, we compute off-line the look-up table $n_{t_{obs}}(\bar{I}_{kij}^r)$ that maps the set of current magnitude values \bar{I}_{kij}^r into spike counts at a given observation time t_{obs} (see Figure 6.4(d)). The reverse mapping is done by a simple interpolation in the reverse-look up table denoted $LUT_{t_{obs}}^{LIF}$. Here we draw the reader's attention to the fact that, as the input of the ganglionic layer is delayed, each coefficient of the subband F_k is decoded according to the reverse map $LUT_{t_{obs}-t_k}^{LIF}$. Obviously, the recovered coefficients do not match exactly the original ones due to the quantization performed in the LIF's.

Recovering the input of the inner layers:

Second, given a rectified current value $\tilde{I}_{kij}^r(t_{obs})$, we recover $\tilde{I}_{kij}^{opl}(t_{obs})$, the estimation of $I_{kij}^{opl}(t_{obs})$. In the same way as for the preceding stage, we infer the reverse “inner layers mapping” through the pre-computed look up table $LUT_{t_{obs}}^{CG}$. The current intensities $\tilde{I}_{kij}^{opl}(t_{obs})$, corresponding to the retinal transform coefficients, are passed to the subsequent retinal transform decoder.

Recovering the input stimulus:

Finally, given the set of $(\frac{4}{3}N^2 - 1)$ coefficients $\{\tilde{I}_{kij}^{opl}(t_{obs})\}$, we recover $\tilde{f}_{t_{obs}}(x, y)$, the estimation of the original image stimulus $f(x, y)$. Though the dot product of every pair of *DoG* filters is approximately equal to 0, the set of filters considered is not strictly orthonormal. We proved in [Masmoudi 2012a] that there exists a dual set of vectors enabling an exact reconstruction (see Chapter 3 for more details). Hence, the reconstruction estimate \tilde{f} of the original input f can be obtained as follows:

$$\tilde{f}_{t_{obs}}(x, y) = \sum_{\{kij\}} \tilde{I}_{kij}^{opl}(t_{obs}) \widetilde{DoG}_k(i - x, j - y), \quad (6.5)$$

where $\{kij\}$ is the set of possible scales and locations in the considered dyadic grid and \widetilde{DoG}_k are the duals of the DoG_k filters obtained as detailed in [Masmoudi 2012a]. Equation (6.5) defines a progressive reconstruction depending on t_{obs} . This provides our code with an important feature: the scalability. Despite the fact that the input of our coder is a static image, we will be referring to this feature as time-scalability. Indeed, in our case different levels of rate and quality levels are achievable thanks to the observation time t_{obs} .

6.3 Results: Case of the bio-inspired and noiseless scalable image coder

We show examples of image reconstruction using our bio-inspired coder at different times¹. Then, we study these results in terms of quality and bit-cost. Quality is assessed by classical image quality criteria (PSNR and mean SSIM [Wang 2004c]). The cost is measured by the Shannon entropy $H(t_{obs})$ upon the population of $\{n_{kij}(t_{obs})\}$. The entropy computed in bits per pixel (bpp), for an N^2 -sized image, is defined by:

$$H(t_{obs}) = \frac{1}{N^2} \sum_{k=0}^{K-1} 2^{2k} H(\{n_{s_k ij}(t_{obs}), (i, j) \in \llbracket 0, 2^k - 1 \rrbracket^2\}), \quad (6.6)$$

where K is the number of analyzing subbands. Figure 6.5 shows two examples of progressive reconstruction obtained with our new coder. Bit-rate/Quality are computed for each image in terms of the triplet (bit-rate in bpp / PSNR quality in dB / mean SSIM quality). Progressive reconstruction of cameraman in the left column yields: From top to bottom (0.006 bpp / 16.02 dB / 0.48), (0.077 bpp / 18.34 dB / 0.55), (0.23 bpp / 21.20 dB / 0.65), and (1.39 bpp / 26.30 dB / 0.84). Progressive reconstruction of baboon in the right column yields: From top to bottom (0.037 bpp / 16.98 dB / 0.18), (0.32 bpp / 19.07 dB / 0.35), (0.63 bpp / 20.33 dB / 0.49), and (2.24 bpp / 27.37 dB / 0.92).

The new concept of *time scalability* is an interesting feature as it introduces time dynamics in the design of the coder. Figure 6.6 illustrates this concept. This is a consequence of the mimicking of the actual retina. We also notice that, as expected, low frequencies are transmitted first to get a first approximation of the image, then details are added progressively to draw its contours. The bit-cost of the coded image is slightly high. This can be explained by the fact that Shannon entropy is not the most relevant metric in our case as no context is taken into consideration, especially the temporal context. Indeed, one can easily predict the number of spikes at a given time t knowing $n_{kij}(t - dt)$. Note also that no compression techniques, such that

¹In all experiments, the model parameters are set to biologically realistic values: $g_0^b = 8 \cdot 10^{-10} S$, $\tau^b = 12 \cdot 10^{-3} s$, $\lambda^b = 9 \cdot 10^{-7}$, $c^b = 1.5 \cdot 10^{-10} F$, $v_0^g = 4 \cdot 10^{-3} V$, $i_0^g = 15 \cdot 10^{-12} A$, $w^g = 8 \cdot 10^{-1}$, $\tau^g = 16 \cdot 10^{-3} s$; $\lambda^g = 12 \cdot 10^{-9} S$, $\delta = 2 \cdot 10^{-3} V$, $g^L = 2 \cdot 10^{-9} S$, $V_R^0 = 0 V$, $t_0 = 10 \cdot 10^{-3} s$, $t_{K-1} = 38 \cdot 10^{-3} s$, $\tau^{opl} = 65 \cdot 10^{-3} s$.

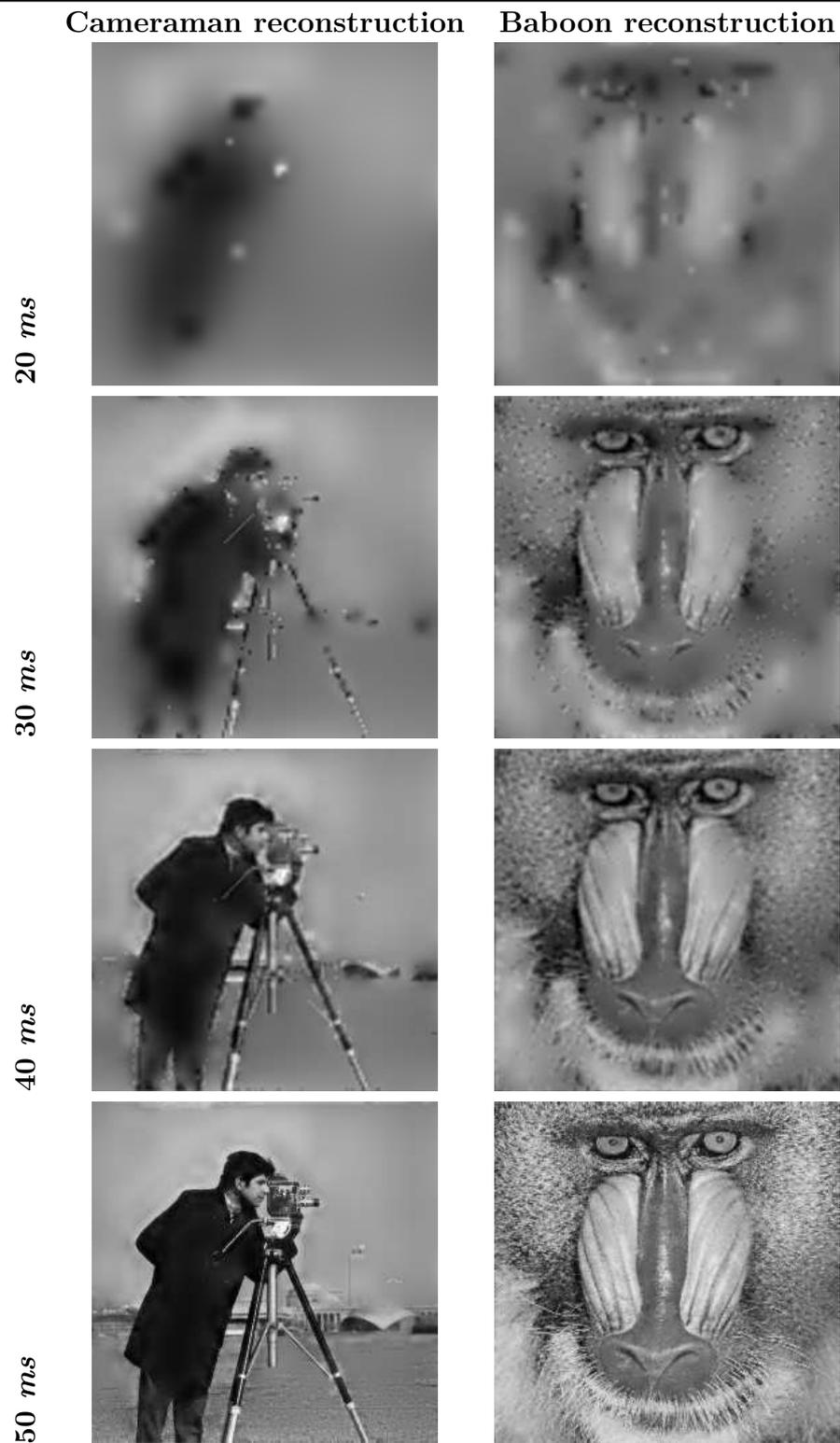


Figure 6.5: Progressive image reconstruction of cameraman and baboon using our new bio-inspired coder. The coded/decoded image is shown at: 20 ms, 30 ms, 40 ms, and 50 ms.

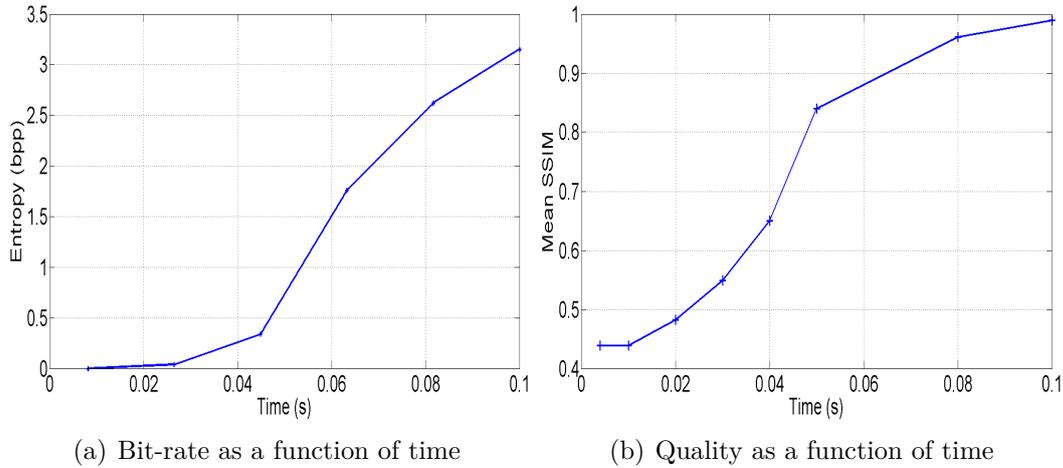


Figure 6.6: Illustration of the concept of time scalability. the test image is cameraman. 6.6(a) shows the bit-rate variation of the encoded image as a function of the observation time t_{obs} . The bit-rate is measured by means of the entropy in bits per pixel (bpp). 6.6(b) shows the reconstruction quality variation as a function of the observation time t_{obs} . The quality is measured by means of the mean structural similarity index (mean SSIM). The only parameter that is tuned by the user is t_{obs} . Both quality and cost increase in accordance with t_{obs} . We talk about time-scalability.

bit-plane coding, are yet employed. Our work aims mainly at setting the basis of new bio-inspired coding designs.

For the reasons cited above, the performance of our coding scheme in terms of bit-cost have still to be improved to be competitive with the well established JPEG and JPEG 2000 standards. Thus we show no comparison in this chapter. Though primary results are encouraging, noting that optimizing the bit-allocation mechanism and exploiting coding techniques as bit-plane coding [Taubman 2000] would improve considerably the bit-cost. Besides, the image as reconstructed with our bio-inspired coder shows no ringing and no block effect as in JPEG. Finally our codec enables scalability in an original fashion through the introduction of time dynamics within the coding mechanism.

Note also that differentiation in the processing of subbands, introduced through time-delays in the retinal transform, ensures an implicit bit-allocation mechanism. In particular the non-linearity in the inner layers stage amplifies singularities and contours, and these provide crucial information for the analysis of the image.

6.4 Introducing the noise in the coder: The non-subtractive dither hypothesis

In the preceding Sections 6.1 and 6.2, we presented the design of an image coder based on a bio-plausible model of the retina. We especially emphasized the deep retina layers analogy with A/D converters. Despite the fact that our coder takes into account several features of the actual retina as its time-dependent behavior, still it follows a deterministic law. Though, the actual neural code of the retina is clearly non-deterministic [de Ruyter van Steveninck 1997, Shadlen 1998]. Thus, in this section, we tackle the issue of the coding non-determinism in the retina. As in Chapter 5, we make the proposal that the processing stages prior to the ganglionic layer yield a special type of noise: the dither noise. We then experience the perceptual impact of such a noise in our coder and give an original and plausible interpretation of its role in the stimuli coding process.

6.4.1 Study context

We introduce in this section a multiscale dithering process that will be integrated in our bio-inspired image coder. Indeed, the coder that we designed has a multiscale architecture. So that the dither noise to be introduced must take into consideration the different scales of the retina model cells used for the image analysis. We will assume that the processing stages of the retina that precede the ganglionic layer introduce a noise. As this noise is prior to the quantization done in the ganglionic layer, it is referred to as a dither noise. Furthermore, this dither noise takes into consideration the multiscale architecture of the retina model. So that, we will be talking about a multiscale dithering. The present work extends our previous efforts in [Masmoudi 2010a, Masmoudi 2010c] to the multiscale case.

Few techniques referred to as multiscale dithering have been described in the literature. For example, in [Wang 2004a] the authors considered a hierarchical wavelet transform. The sibling subbands, id est lying in the same level, are decorrelated by applying a series of rotations. The transform applied on the subbands is loosely referred to as dithering because

it introduces a change on the wavelet coefficients prior to quantization. The resulting image is meant to reduce entropy while keeping the same perceptual quality. Another example is given in [Katsavounidis 1997]. The authors used an image hierarchical quadtree representation and employ an error diffusion algorithm to get a binary halftone image. The distribution of binary pixels over the image space gives the impression of a multi-gray level image while using only two quantization levels. Although interesting, these state-of-the-art algorithms have one major drawback regarding the goals of our present work. Indeed, the techniques described rely on a totally deterministic algorithm. No random behavior is introduced during the coding process. Whereas in our case, we need to consider a coding process that may lead to different codes across trials for a single image. Besides the two algorithms are iterative and time consuming and this contradicts the speed of processing in the retina.

6.4.2 Introducing the multiscale NSD

In order to define the dither noise that corrupts the current I_{kij}^r (cf. Equation (6.4)) at the input of the ganglionic layer, we reconsider the ganglion cell as a noisy leaky integrate and fire neuron (nLIF), that behaves according to Equation (5.6) that we recall in the following:

$$c^l \frac{dV_{kij}(t)}{dt} + g^l V_{kij}(t) = I_{kij}^r(t) + \eta_{kij}^l, \quad \text{for } t \in [t_{kij}^{(l)}, t_{kij}^{(l+1)}],$$

The choice of the noise η_{kij}^l distribution model to apply must obey two constraints: the biological plausibility and the mathematical constraints that provide our coder with interesting perceptual properties.

First, let us consider the biological plausibility constraint. Our aim is to mimic as closely as possible the actual retina behavior while modelling the multiscale dithering η_{kij}^l . So that, one must consider the nature of the dependency (if any) between the scale and the noise strength according to neurophysiologists observations. In this context, the authors in [Kier 1995] stated that: “*The main difference between small and large cells is that the larger ones have lower peak sensitivity*”. This means that the large retina

cells have a low reactivity to stimulus variations and thus are poorly affected by noise. On the contrary, small cells are extremely sensitive to stimulus variations and thus could be highly affected by noise. Our aim is to reproduce this phenomenon of noise strength variability as a function of retina cells scale. So that, we will corrupt the currents (I_{kij}^r) at the input of the retina ganglionic layer with noise coefficients η_{kij}^l , such that the dynamic range of this noise distribution depends on the cells scale k . The larger the subband F_k cells are, the lower the noise dynamic range is. Thus, we will have to generate K noise subbands, with an increasing dynamic range, to corrupt K subbands of rectified currents I_{kij}^r .

Second, let us consider the mathematical constraint. Indeed we must consider the statistical properties that have to be verified by the added noise to provide our coder with interesting perceptual features. To this end, we refer to theorem 1 (Chapter 5). Thus, we suppose that (i) η_{kij}^l has a triangular probability distribution function with no loss of biological plausibility, and (ii) that the dynamic range of η_{kij}^l is twice wider than the quantization step of the considered ganglion cell. Having these two conditions we verify the theorem. In this way, we identify the retinal noise η_{kij}^l to a dither signal. As we do not subtract the dither signal in the decoding process, our coder is said to be an NSD [Wannamaker 2000, Wannamaker 2004].

According to the discussion above we will consider that the noise η_{kij}^l dynamic range (i) is an increasing function of the scale k of the considered *DoG* retina cell, and (ii) is twice the width of the quantization step of the subsequent ganglion cell. Here we remind the reader that the ganglion cells are modelled, in our coder, by LIF neurons that are dynamic quantizers. Indeed the ganglionic layer evolves from a coarse to a fine quantizer. The quantization step of a LIF neuron will be denoted Q^{lif} . Obviously, Q^{lif} is a decreasing function of the observation time t_{obs} as shown in Figure 6.7. Furthermore, according to our original retina transform (cf. Section 6.1.1), the coding process of each subband F_k is delayed in time by t_k . So that, the ganglion cells will have different levels of progression at a given time t_{obs} depending on the subband scale k . We set the dithering parameters for an optimal observation time t_{obs}^* . So that, each subband will be corrupted by a

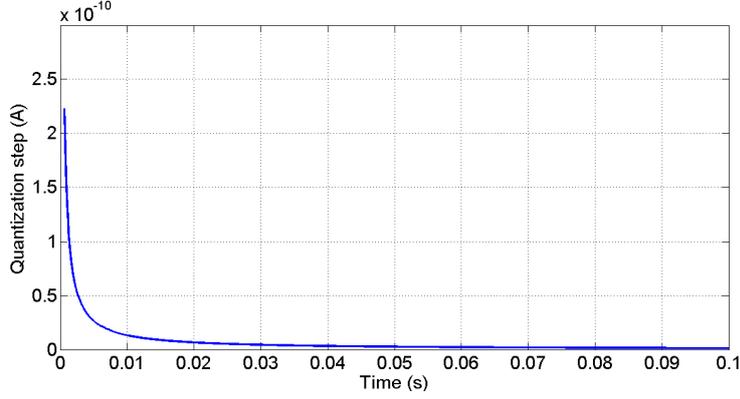


Figure 6.7: Estimation of the LIF neuron quantization step Q^{lif} as a function of the observation time t_{obs} . The abscissa shows the observation times t_{obs} between 0 ms and 100 ms . The ordinate axis shows the mean quantization step Q^{lif} estimated at a given t_{obs} in Amperes.

noise subband having a triangular pdf which dynamic range Δ_k depends on the scale k such that:

$$\begin{cases} Q^{lif}(t_{obs}^* - t_k) = Q_k^* \\ \Delta_k = 2 Q_k^* \end{cases} \quad (6.7)$$

An example of a multiscale dither noise thus defined is given in Figure 6.8. The test image is cameraman and the optimal observation time chosen is $t_{obs}^* = 52\text{ ms}$. The rectified currents I_{kij}^r in each subband of scale k are subject to a dither noise η_{kij}^l that has a triangular distribution with a dynamic range Δ_k . We can notice that large cells in the low frequency subbands are poorly corrupted with noise while tight cells in the high frequency subbands are highly corrupted with noise. This is due to the fact that $\Delta_{k+1} > \Delta_k, \forall 0 \leq k < K - 2$. Interestingly, we remark that the time delays introduced in our model of the retinal transform allow us to implicitly satisfy the constraint of noise dynamic range Δ_k being an increasing function of the cells scale k .

Adding such a dither noise to the input of the ganglionic layer I_{kij}^r induces interesting features. As specified in the theorem 1 (Chapter 5), one important feature is the decorrelation between the reconstruction error at the output of the de-quantizer and the original signal at the input of the corresponding quantizer. The results of the theorem were demonstrated for uniform scalar quantizers. Whereas in our coder the ganglionic layer is not strictly a scalar

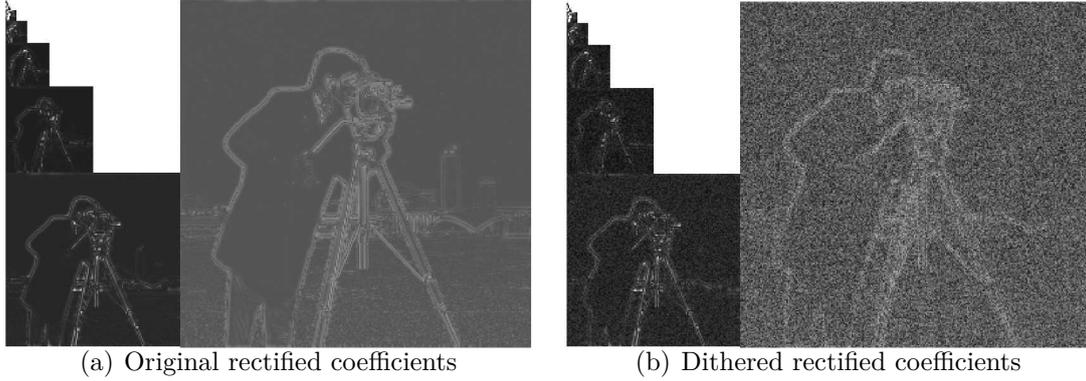


Figure 6.8: Example of dither noise introduced at the input of the ganglionic layer. The test image is cameraman. 6.8(a) shows the noiseless rectified coefficients (I_{kij}^r). 6.8(b) shows the rectified coefficients ($I_{kij}^r + \eta_{kij}^l$) with the dither noise η_{kij}^l added. The noise parameters are set for the optimal observation time is $t_{obs}^* = 52ms$. The dither noise has a triangular distribution with a dynamic range Δ_k that depends on the subband F_k considered. The larger the subband cells are, the lower the noise dynamic range is. The high frequencies are more corrupted with noise than the low frequencies.

quantizer but rather an approximation of it and, furthermore, the bio-inspired A/D converter that we designed is not uniform due to the preceding gain control and non-linear rectification stages. So that, we must verify the relevance of our approach. As the dithering process occurs in the DoG transform domain, we measure the error/input correlation in the transform domain. The error that we will denote by ε_{kij} is defined, in this case, as the difference between the output of the OPL layer I_{kij}^{opl} and the estimation of it after decoding \tilde{I}_{kij}^{opl} , such that:

$$\varepsilon_{kij} = I_{kij}^{opl} - \tilde{I}_{kij}^{opl} \quad (6.8)$$

We can experimentally verify that, in fact, ε_{kij} and the input stimuli I_{kij}^{opl} are decorrelated. This feature is clearly demonstrated when computing the cross correlation between ε_{kij} and I_{kij}^{opl} as shown in Figures 6.9(a) and 6.9(b) for the test image cameraman and the highest frequency subband F_{K-1} . Comparable observations are made on the other subbands. Figure 6.9(a) shows the cross-correlation between ε_{kij} and I_{kij}^{opl} measured for the noiseless case. The correlation is high especially when the spatial lag is small. 6.9(b) shows the same cross-correlation measures for the dithered case. We observe a very high decrease in the correlation even for the small spatial lags cases. Then, we can conclude that the signals ε_{kij} and I_{kij}^{opl} are clearly decorrelated.

Another perceptually important feature that is induced by the dithering process is the error whitening. We verify also this feature in our case. As shown in Figures 6.9(c), the spectrum of the error obtained when using our coder with no addition of noise is non-uniform. This denotes strong geometric correlations in the error image which yields annoying artefacts. On the contrary, we notice in Figure 6.9(d) that the error spectrum is equally dispatched in the Fourier domain if we add a dither noise. Thus our new dithered scalable image coder gained interesting features through the integration of a dithering process.

The whitening and de-correlation features yield a greater reconstruction error in terms of mean squared error [Wannamaker 2000]. Though, the error whitening and decorrelation features acquired in the transform domain are perceptually important. Indeed, a strong correlation between the coding error and the original signal implies annoying artefacts. Besides the error whitening is important because all frequencies are affected by the same noise. The perceptual impact of dithering on the final image reconstruction $\tilde{f}_{t_{obs}}$ is shown in the next section.

6.5 Results: Case of the bio-inspired and dithered scalable image coder

We show in this section the perceptual impact of the dithering on the reconstructed images using our decoder. Our experiments demonstrate the ability of the dither noise to accelerate the recognition of the image details and singularities during the decoding process.

A first example is given in Figure 6.10. The left column shows the evolution of the reconstruction $\tilde{f}_{t_{obs}}$ with increasing times t_{obs} , in the case of noiseless coding. The right column shows the evolution of the reconstruction $\tilde{f}_{t_{obs}}$ with increasing times t_{obs} , in the case of addition of a dither noise to the input of the ganglionic layer. The central column shows a filtered version of cameraman. Cameraman is sharpened to enhance the image details. The comparison between the noiseless case reconstruction (on the left) and the dithered reconstruction (on the right) demonstrates perceptual importance

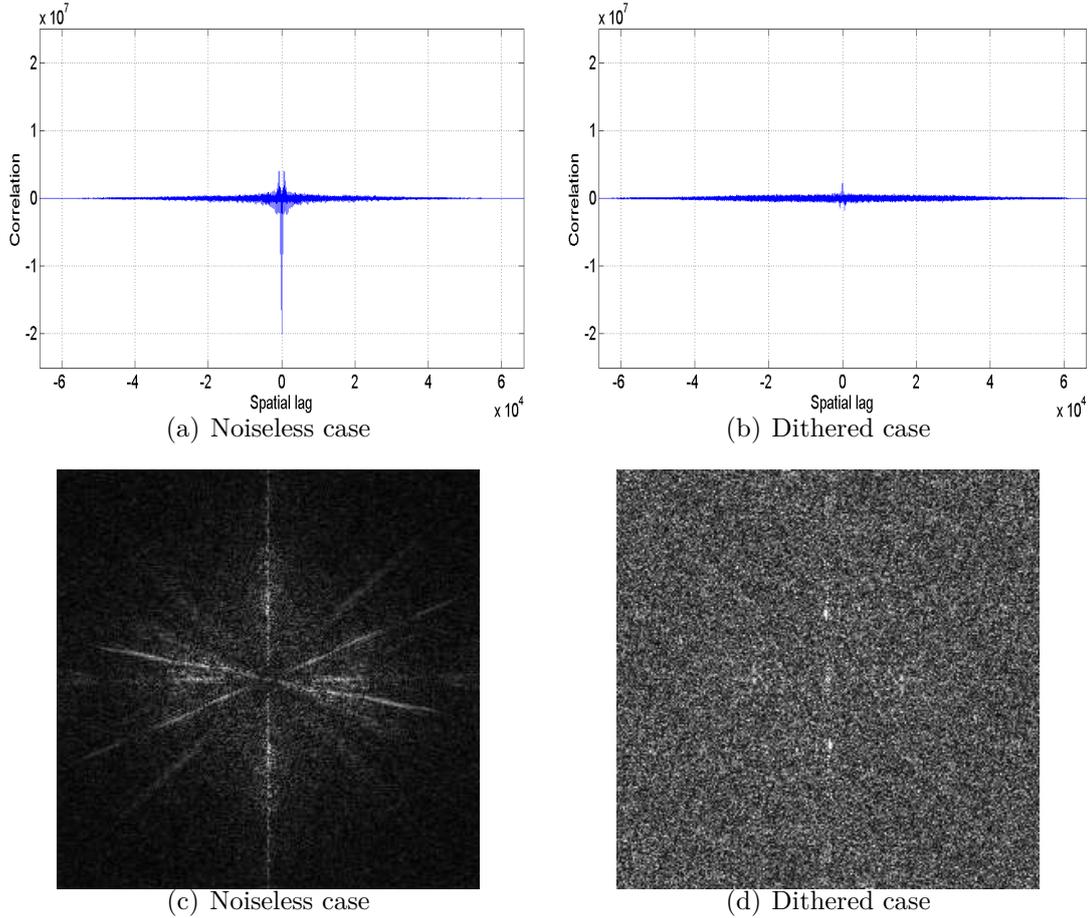


Figure 6.9: Error whitening and decorrelation in the DoG transform domain induced by the dither noise addition. The results are shown for the highest frequency subband B_{K-1} , but comparable observations are made on the other subbands. The dither noise is introduced at the input of the ganglionic layer. The dither noise parameters are set for the optimal observation time is $t_{obs}^* = 52ms$. 6.9(a) (respectively 6.9(b)) shows the cross-correlation between ε_{kij} and I_{kij}^{opl} measured for the noiseless (resp. dithered) case. We observe a very high decrease in the correlation induced by the noise. The error is decorrelated from the input. 6.9(c) (respectively 6.9(d)) shows the amplitude spectrum of ε_{kij} computed for the noiseless (resp. dithered) case. We observe a wide spreading of the error spectrum in Fourier domain induced by the noise. The error is whitened.

of noise in the image coding process in the retina. With the addition of noise, details of cameraman are well rendered “before date”. For example, the hand of cameraman and the tower in the background appear since $t_{obs} = 44ms$ for the dithered case while still invisible in the noiseless case at the same observation time. We can also notice that the horizontal stripes in the background, the grass details, the pant folds, and the hand are well rendered since $t_{obs} = 48ms$. On the contrary these details are still invisible or highly blurry in the noiseless case at the same observation time. Finally, at the optimal observation $t_{obs} = t_{obs}^* = 52ms$ all the fine details of the image, including the coat and the background details, are clearly distinguished in the dithered case while still blurry or invisible in the noiseless case.

A second example is given in Figure 6.11 for the baboon test image. This image is rich of details and singularities and thus particularly challenging. Though, our dithered coder still renders the image details “before date in this case” (with another adequate parametrization for the dither noise). As for the preceding example, the left image shows the reconstruction $\tilde{f}_{t_{obs}}$ in the case of noiseless coding. The right image shows the reconstruction $\tilde{\tilde{f}}_{t_{obs}}$ in the case of addition of a dither noise to the input of the ganglionic layer. The central image is a sharpened version of baboon. The observation time shown in this figure is also the optimal observation $t_{obs} = t_{obs}^* = 44ms$. The comparison between the noiseless case reconstruction (on the left) and the dithered reconstruction (on the right) confirms the observations made in the first example. The dither noise helps the recognition of fine details “before date”. While in the noiseless case face and bear details of baboon are still blurry, these details are well rendered in the dithered reconstruction case.

On one hand, the integration of a dither noise in the coding process yield a greater reconstruction error in terms of mean squared error [Wannamaker 2000]. Besides, as the dither noise is a disordered signal, it also increases the entropy of the image code. On the other hand, the error whitening and de-correlation features acquired by our system are perceptually important. This is a crucial point because our current results may prove that the retina conveys a code that is optimized for the tasks to be performed by the visual cortex as categorization. While the rate/distortion trade-off remains an important goal for a coding scheme it may not be the central performance



Figure 6.10: Perceptual impact of the multiscale dithering on the reconstruction of cameraman. The observation times t_{obs} are shown on the left. From top to bottom, t_{obs} take successively the values of: 40 ms , 44 ms , 48 ms , and 52 ms . The observation time shown in this figure is also the optimal observation $t_{obs} = t_{obs}^* = 44\text{ ms}$.

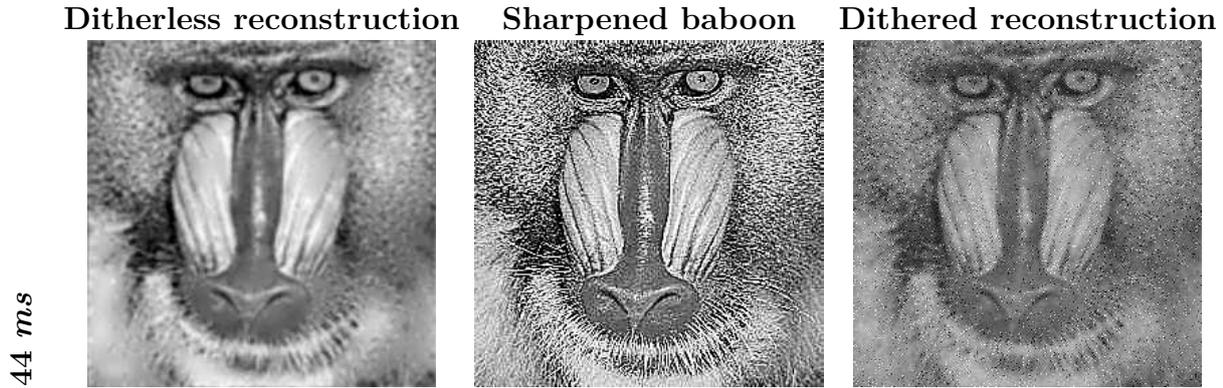


Figure 6.11: Perceptual impact of the multiscale dithering on the reconstruction of baboon. The observation time shown in this figure is also the optimal observation $t_{obs} = t_{obs}^* = 44 ms$.

criterion for the retina.

6.6 Conclusion

The work that we presented brings two main contributions. As a first step, we proposed a bio-inspired codec for static images with a deterministic behavior. The image coder is based on two stages. The first stage is the image transform as performed by the outer layers of the retina. In order to integrate time dynamics, we added to this transform time delays that are subband specific so that, each subband is processed differently. The second stage is a succession of two dynamic processing steps mimicking the deep retina layers behavior. These latter perform an A/D conversion and generate a spike-based, invertible, retinal code for the input image in an original fashion.

In a second step, we investigated the issue of non-determinism in the retina neural code. We proposed to model the retinal noise by a multiscale dither signal with specific statistical properties. The dithering process that we proposed whitens the reconstruction error and decorrelates it from the input stimuli. Besides, from a perceptual point of view, our coder allows an earlier recognition of the image details and singularities during the decoding process.

In conclusion, our coding scheme offers interesting features such as (i) time-scalability, as the choice of the observation time of our codec enables

different reconstruction qualities, and (ii) bit-allocation, as each subband of the image transform is separately mapped according to the corresponding state of the inner layers. In addition, when integrating a dithering process our coder gained interesting perceptual features. These features, if the dithering hypothesis is confirmed, help the visual cortex recognize the fine details of the image. This latter point is interesting because it may prove that the retina conveys a code that is optimized for the tasks to be performed by the visual cortex. Interestingly, our dithering hypothesis found an echo recently in the computational neurosciences community [Vidne 2012]. We are convinced that further neurophysiologic investigations may also confirm the relevance of dithering in the retinal processing.

In terms of rate/distortion, the results accomplished by our coding scheme are encouraging. Though the rate/distortion performance is not the primary goal of this work, our coder could still be improved to be competitive with the well established JPEG and JPEG 2000 standards. Optimizing techniques as bit-plane coding are to be investigated.

This work is at the crossroads of diverse hot topics in the fields of neurosciences, brain-machine interfaces, and signal processing and tries to bridge the gap between these different domains towards the conception of new biologically inspired coders.

DISCUSSION AND PERSPECTIVES

The work that we presented in this manuscript had the aim of conceiving novel image coders inspired from the retina. We organized our effort into two parts. In the first part of the manuscript, we assumed that the rank order is the coding feature within the retinal neural code. We started from the classical the retina model specified in [Van Rullen 2001b]. We enhanced this model and combined it with data compression techniques to get a first bio-inspired image coder/decoder. This coder/decoder has several interesting features, among them the scalability. In addition to [Van Rullen 2001b], the second part of my manuscript relied also on a retina simulation software called *Virtual Retina* [Wohrer 2009a]. This simulator specification helped us enhance our first coder/decoder by taking into account several supplemental retinal processing stages. The bio-inspired coder/decoder that we devised in this part is an enhancement of the first one in the sense that we lead our effort towards more biological plausibility. Our coder/decoder is time scalable, where the time scalability designates the ability of tuning the rate and quality through the choice of the reconstruction time. We kept the design of our coder/decoder as close as possible to the biological reality, while keeping an interesting rate/distortion trade-off.

Regarding the first part of this work, we proposed an original and exact decoding procedure for the retina model considered. Our contribution encompassed a theoretical and a technical aspect. Regarding the theoretical aspect, (i) we proved that the classical bio-inspired transform used to model the retina is non-invertible as it is, and (ii) we gave an original mathematical proof that this transform if augmented with an adequate scaling function is a

frame. We then defined the corresponding dual frame that is necessary for the exact image reconstruction. Regarding the technical aspect, we overcame the problem of memory overhead encountered while computing this dual frame using an original recursive out-of-core blockwise algorithm.

Then, we have proposed a new bio-inspired coder/decoder for static images. First, the image is converted into a ROC code, then a stack run coder is applied, followed by a first order arithmetic compressor. The performances of this coding scheme were tested against well established JPEG standards, and we obtain encouraging results for low bandwidth transmissions, especially when dealing with noisy data. This compression scheme also offers interesting features such as scalability and reasonable complexity. Though, limitations have been observed in terms of rate/quality, when compared to JPEG2000 for noiseless data transmissions.

Regarding the second part of this work, we first presented a bio-inspired A/D converter mapping the retina behavior. Our A/D converter is dynamic and implicitly enables scalability. Interestingly, our quantizer evolves from uniform to non-uniform, but in contradiction with traditional Lloyd-Max quantizers, as it renders high magnitudes precisely while it maps low magnitudes coarsely.

We proposed a bio-inspired coder/decoder for static images with a deterministic behavior. Our image coder/decoder is based on two stages. The first stage is the image transform as performed by the outer layers of the retina. In order to integrate time dynamics, we added to this transform time delays that are subband specific so that, each subband is processed differently. The second stage is our A/D converter previously designed.

We investigated the issue of non-determinism in the retina neural code. We emitted a biologically plausible hypothesis that supposes the retinal noise is a dither signal. We do not claim that the exactness of our hypothesis, but still it is seducing as it enables the noise whitening and de-correlation features. In order to adapt this hypothesis to our coder/decoder, we modelled the retinal noise by a multiscale dither signal. The dithering process that we proposed (i) whitens the reconstruction error and (ii) decorrelates it from the analysis coefficients of the input stimuli. Besides, from a perceptual point of view, our coder allows an earlier recognition of the image details and

singularities during the decoding process.

As conclusion, our novel coder/decoder scheme offers interesting features such as (i) time-scalability, and (ii) bit-allocation. In addition, when integrating a dithering process our coder gained interesting perceptual features. These features, if the dithering hypothesis is confirmed, help the visual cortex recognize the fine details of the image.

Regarding the future work and perspectives, several aspects are to be extended and enhanced. One possible extension could be the addition of several mechanisms of the retinal processing that are not taken into account in the current coder/decoder. The lateral inhibitory mechanism is an example of features that could be integrated in an upcoming effort. A long term goal would be to infer, starting from a sufficiently realistic model, a decoding algorithm that could decipher actual neural recordings.

One perspective concerns the coding features used in our different schemes, namely rank order or spike counts. Several coding models relying on alternative coding features. For example, it appears that bursts or synchronies are features that could encode for the stimulus. The goal is then to use these models in order to reproduce some spiking pattern as observed in real cell recordings, and establish with more accuracy how spikes are triggered by a stimulus then decoded by the nervous system [Rieke 1997, Lesica 2004].

In terms of rate/distortion, the results accomplished by our coding scheme are encouraging. Though the rate/distortion performance is not the primary goal of this work, our coder could still be improved to be competitive with the well established JPEG and JPEG 2000 standards. Optimizing techniques as bit-plane coding are to be investigated.

Regarding our dithering hypothesis, we think that the neurosciences community could conduct interesting investigations. Indeed, this hypothesis found an echo recently in the computational neurosciences community [Vidne 2012]. We are convinced that further neurophysiologic experimentations may confirm the relevance of dithering in the retinal processing.

One more perspective concerns the design of a novel framework to capture and represent the retinal code at a rate that we expect to be below the Nyquist rate. This framework would rely on the theory of compressive sensing that

employs combined and non-adaptive linear projections to encode the signal. In a compressive sensing system, the unknown is a highly sparse signal, and given the adequate restrictions could be recovered exactly though coded below the Nyquist rate [Candès 2008]. Both compressive sensing systems and rank order coders suppose that the signals transmitted are highly sparse. So that we think that an analogy is to be made between the issues underlying the retinal coding and the tools developed in the compressive sensing literature. This work would bring a new insight on the coding and decoding processes within the retina and might answer to the question: Is the retina is a compressive sensing system?

One last perspective concerns our retina inspired A/D converter. Indeed, our A/D converter mimics the retina neurons behavior. Interestingly, recent efforts try to raise analogies between some neuron models and Σ - Δ modulators [Rastogi 2011]. These Σ - Δ modulators have several applications in particular in brain machine interfaces [Moxon 2000]. Yet, Σ - Δ modulators are already applied in neural prosthetics. Thus we think that our A/D converter could be extended and highly improved to be incorporated in the design of retinal prosthetics. Therefore this issue is of particular importance.

This work is at the crossroads of diverse hot topics in the fields of neurosciences, brain-machine interfaces, and signal processing and tries to bridge the gap between these different domains towards the conception of new biologically inspired coders. Thus we hope that this manuscript would be helpful for researchers from the different cited domains.

AN OUT-OF-CORE ALGORITHM FOR HIGH-DIMENSION MATRIX INVERSION

We give in the following a pseudo-code description of the out-of-core matrix inversion evoked in Chapter 3. This algorithm is not specific to the matrices used in this work and could be used for any other invertible high-dimensionality matrix. The inversion of a high-dimension matrix can be achieved bloc by bloc. Each bloc is stored separately on disk. The inversion is then performed using a recursive algorithm that relies on the blockwise matrix inversion formula that follows:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1} B Q^{-1} C A^{-1} & -A^{-1} B Q^{-1} \\ -Q^{-1} C A^{-1} & Q^{-1} \end{pmatrix},$$

where Q is the Schur complement of A . Inverting a matrix amounts to the inversion of two matrices that are 4 times smaller. The inversion consists then in subdividing the problem by a factor 4 at each recursion level until we reach a single bloc problem. This algorithm requires a set of out-of-core blockwise matrix routines for multiplication, subtraction and addition and also other utility functions. All of these functions calls are marked in dark green in the code that follows and their names are self-explanatory. It is also to be noted that the algorithm that we implemented is multi-threaded to accelerate the computation but this aspect is discarded in the pseudo-code below:

```

void computeBlockWiseInversion (baseNameMatrixFile,...
xLimits, yLimits, baseNameInverseMatrixFile, recursionLevel)
2:
  { //Initialize boundary variables}
4:
  xBegin ← xLimits[1];
6: xEnd ← xLimits[2];
  yBegin ← yLimits[1];
8: yEnd ← yLimits[2];
  { //Treat the single bloc case: The recursion end-point}
10: if (getNumberOfColumns(baseNameMatrixFile)== 1 &&...
  getNumberOfLines(baseNameMatrixFile)== 1) then
  bloc ← computeGaussJordanElimination(baseNameMatrixFile)
12: store(baseNameInverseMatrixFile, bloc)
  { //Treat the multi bloc case: The recursion must go on}
14: else
  numberOfLinesInSubBloc ← round((xEnd-xBegin+1) / 2)
16: numberOfColumnsInSubBloc ← round((yEnd-yBegin+1) / 2)
  xLimitsOfA ← {xBegin, xBegin+numberOfLinesInSubBloc-1}
18: yLimitsOfA ← {yBegin, yBegin+numberOfColumnsInSubBloc-1}
  xLimitsOfB ← {xBegin, xBegin+numberOfLinesInSubBloc-1}
20: yLimitsOfB ← {yBegin+numberOfColumnsInSubBloc, yEnd-1}
  xLimitsOfC ← {xBegin+numberOfLinesInSubBloc, xEnd}
22: yLimitsOfC ← {yBegin, yBegin+numberOfColumnsInSubBloc-1}
  xLimitsOfD ← {xBegin+numberOfLinesInSubBloc, xEnd}
24: yLimitsOfD ← {yBegin+numberOfColumnsInSubBloc, yEnd}
  computeBlockWiseInversion (baseNameMatrixFile, xLimitsOfA,
  yLimitsOfA, baseNameInverseOfA, recursionLevel+1) { //A-1}
26: auxiliaryVariableName ← concatenateStrings(baseNameSchurA,
  "Aux1")
  multiplyBlockwiseMatrices(baseNameInverseOfA, 1 nbrLignesSous-
  Bloc, 1 nbrColonnesSousBloc, baseNameMatrixFile, xLimitsOfB,
  yLimitsOfB, auxiliaryVariableName) { //D - CA-1B}

```

```

28: auxiliaryVariableName2 ← concatenateStrings(baseNameSchurA,
    "Aux2")
    multiplyBlockwiseMatrices ( baseNameMatrixFile, indicesCII, in-
    dicesCJJ,...auxiliaryVariableName1, 1 nbrLignesSousBloc, 1 nbrColon-
    nesSousBloc,...auxiliaryVariableName2);
30: delete(concatenateStrings(auxiliaryVariableName1, "*")
    subtractBlockWiseMatrices( baseNameMatrixFile, indicesDII, indices-
    DJJ, auxiliaryVariableName2, 1 nbrLignesSousBloc, 1 nbrColonnes-
    SousBloc, baseNameSchurA);
32: delete(concatenateStrings(auxiliaryVariableName2, "*")
    baseNameSchurAinv ← concatenateStrings (baseNameInverseMatrix-
    File, "SchurAinverseRecursionLevel", itoa(recursionLevel+1))
34: computeBlockWiseInversion( baseNameSchurA, 1 nbrLignesSous-
    Bloc, 1 nbrColonnesSousBloc, baseNameSchurAinv, recursionLevel+1)
    {//System command}
36: command = 'del ', baseNameSchurA, '*';
    system(command);
38: for ii=indicesDII(1):indicesDII(2) do
    for jj=indicesDJJ(1):indicesDJJ(2) do
40:     nomBlocInverse = baseNameInverseMatrixFile, num2str(ii),
        num2str(jj), '.mat';
        nomBlocSchurAinvIIJJ = baseNameSchurAinv, num2str(ii-
        indicesDII(1)+1), num2str(jj-indicesDJJ(1)+1), '.mat';
42:     load(nomBlocSchurAinvIIJJ)
        save(nomBlocInverse, 'blocIIJJ')
44: auxiliaryVariableName1 = baseNameInverseMatrixFile, 'AuxB1';
    multiplyBlockwiseMatrices( nomBaseFichiersMatrice, indicesBII, in-
    dicesBJJ,...baseNameSchurAinv, 1 nbrLignesSousBloc, 1 nbrColonnes-
    SousBloc,...auxiliaryVariableName1);
46: auxiliaryVariableName2 = baseNameInverseMatrixFile, 'AuxB2';
    multiplyBlockwiseMatrices( baseNameInverseOfA, 1 nbrLignesSous-
    Bloc, 1 nbrColonnesSousBloc,...auxiliaryVariableName1, 1 nbrLignes-
    SousBloc, 1 nbrColonnesSousBloc,...auxiliaryVariableName2);
48: command = 'del ', auxiliaryVariableName1, '*';

```

```

    system(command);
50:  for ii=indicesBII(1):indicesBII(2) do
        for jj=indicesBJJ(1):indicesBJJ(2) do
52:      nomBlocInverse = baseNameInverseMatrixFile, num2str(ii),
          num2str(jj), '.mat';
          nomBlocAuxiliaire2IIJJ = auxiliaryVariableName2, num2str(ii-
            indicesBII(1)+1), num2str(jj-indicesBJJ(1)+1), '.mat';
54:      load(nomBlocAuxiliaire2IIJJ)
          blocIIJJ = -blocIIJJ;
56:      save(nomBlocInverse, 'blocIIJJ')
          auxiliaryVariableName1 = baseNameInverseMatrixFile, 'AuxA1';
58:      multiplyBlockwiseMatrices( nomBaseFichiersMatrice, indicesCII, in-
        dicesCJJ,...baseNameInverseOfA, 1 nbrLignesSousBloc, 1 nbrColonnes-
        SousBloc,...auxiliaryVariableName1);
          auxiliaryVariableName3 = baseNameInverseMatrixFile, 'AuxA4';
60:      multiplyBlockwiseMatrices( auxiliaryVariableName2, 1 nbrLignesSous-
        Bloc, 1 nbrColonnesSousBloc,...auxiliaryVariableName1, 1 nbrLignes-
        SousBloc, 1 nbrColonnesSousBloc,...auxiliaryVariableName3);
          command = 'del ', auxiliaryVariableName2, '*';
62:      system(command);
          addBlockwiseMatrices( auxiliaryVariableName3, 1 nbrLignesSousBloc,
            1 nbrColonnesSousBloc,...baseNameInverseOfA, 1 nbrLignesSousBloc,
            1 nbrColonnesSousBloc,...baseNameInverseMatrixFile);
64:      command = 'del ', auxiliaryVariableName3, '*';
          system(command);
66:      auxiliaryVariableName2 ← baseNameInverseMatrixFile, 'AuxC1';
          multiplyBlockwiseMatrices( baseNameSchurAinv, 1 nbrLignesSous-
            Bloc, 1 nbrColonnesSousBloc,...auxiliaryVariableName1, 1 nbrLignes-
            SousBloc, 1 nbrColonnesSousBloc,...auxiliaryVariableName2);
68:      for ii=indicesCII(1):indicesCII(2) do
          for jj=indicesCJJ(1):indicesCJJ(2) do
70:          nomBlocInverse = baseNameInverseMatrixFile, num2str(ii),
            num2str(jj), '.mat';

```

```
nomBlocAuxiliaire2IIJJ = auxiliaryVariableName2, num2str(ii-
indicesCII(1)+1), num2str(jj-indicesCJJ(1)+1), '.mat';
72: load(nomBlocAuxiliaire2IIJJ)
    blocIIJJ = -blocIIJJ;
74: save(nomBlocInverse, 'blocIIJJ')
    delete(concatenateStrings(auxiliaryVariableName1, "*")
76: delete(concatenateStrings(auxiliaryVariableName2, "*")
    delete(concatenateStrings(baseNameSchurAinv, "*")
78: delete(concatenateStrings(baseNameInverseOfA, "*"))
```


Bibliography

- [Adelson 1987] E.H. Adelson, E.P. Simoncelli, R. Hingorani and Massachusetts Institute of Technology. Media Laboratory. Orthogonal pyramid transforms for image coding. Media Laboratory, Massachusetts Institute of Technology, 1987. (Cited on page 7.)
- [Adrian 1926] E.D. Adrian. *The impulses produced by sensory nerve endings*. Journal of Physiology, vol. 61, no. 1, pages 49–72, 1926. (Cited on pages 25, 117, 121 and 131.)
- [Ahmed 1974] N. Ahmed, T. Natarajan and K.R. Rao. *Discrete cosine transform*. Computers, IEEE Transactions on, vol. 100, no. 1, pages 90–93, 1974. (Cited on page 6.)
- [Allison 2000] J.D. Allison, P. Melzer, Y. Ding, AB Bonds and V.A. Casagrande. *Differential contributions of magnocellular and parvocellular pathways to the contrast response of neurons in bush baby primary visual cortex(V 1)*. Visual neuroscience, vol. 17, no. 1, pages 71–76, 2000. (Cited on page 20.)
- [Antonini 1992] M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies. *Image coding using wavelet transform*. IEEE Transactions on Image Processing, 1992. (Cited on pages 3 and 6.)
- [Antonini 2002] M. Antonini and V. Ricordel. *Traité ic2, chapitre Quantification*, pages 45–72. Hermès, Paris, 2002. (Cited on page 129.)
- [Antonini 2006] M. Antonini, J. Bensa, M. Oger and S. Ragot. *Stack-Run Audio Coding*. In Audio Engineering Society Convention 120, 5 2006. (Cited on page 90.)
- [Baccus 2002] S.A. Baccus and M. Meister. *Fast and slow contrast adaptation in retinal circuitry*. Neuron, vol. 36, no. 5, pages 909–919, 2002. (Cited on page 112.)

- [Beaudoin 2007] D.L. Beaudoin, B.G. Borghuis and J.B. Demb. *Cellular basis for contrast gain control over the receptive field center of mammalian retinal ganglion cells*. Journal of Neuroscience, Soc Neuroscience, vol. 27, no. 10, page 2636, 2007. (Cited on page 110.)
- [Bell 1990] T.C. Bell, I. H. Witten and J.G. Cleary. Text compression. Prentice Hall, Englewood Cliffs, N.J. :, 1990. (Cited on page 90.)
- [Bhattacharya 2007] Basabdatta Sen Bhattacharya and Steve Furber. *Maximising Information Recovery from Rank-Order Codes*. In Proceedings of SPIE, volume 6570, Orlando, FL, U.S.A., 2007. (Cited on pages 61 and 73.)
- [Bhattacharya 2010] Basabdatta Sen Bhattacharya and Steve Furber. *Biologically Inspired Means for Rank-Order Encoding Images: A Quantitative Analysis*. IEEE Trans. Neural Netw., vol. 21, no. 7, pages 1087–1099, 2010. (Cited on pages 73, 74 and 76.)
- [Bialek 1991] W. Bialek, F. Rieke, R. de Ruyter van Steveninck and D. Warland. *Reading a neural code*. Science, vol. 252, no. 5014, pages 1854–1857, June 1991. (Cited on page 4.)
- [Binkofski 1998] F. Binkofski, C. Dohle, S. Posse, KM Stephan, H. Hefter, RJ Seitz and HJ Freund. *Human anterior intraparietal area subserves prehension*. Neurology, vol. 50, no. 5, pages 1253–1259, 1998. (Cited on page 19.)
- [Bodden 2007] E. Bodden, M. Clasen and J. Kneis. *Arithmetic Coding revealed : A guided tour from theory to praxis*. Rapport technique, McGill University School of Computer Science Sable Research Group, 2007. (Cited on page 93.)
- [Brown 2004] E.N. Brown, R.E. Kass and P.P. Mitra. *Multiple neural spike train data analysis: state-of-the-art and future challenges*. Nature Neuroscience, vol. 7, no. 5, pages 456–461, 2004. (Cited on page 131.)
- [Bullier 2001] J. Bullier. *Integrated model of visual processing*. Brain Research Reviews, vol. 36, no. 2-3, pages 96–107, 2001. (Cited on page 21.)

- [Burt 1983] P. Burt and E. Adelson. *The Laplacian pyramid as a compact image code*. IEEE Trans. Commun., vol. 31, no. 4, pages 532–540, 1983. (Cited on pages 6 and 57.)
- [Cabrera Fernández 2005] D. Cabrera Fernández, H.M. Salinas and C.A. Puliafito. *Automated detection of retinal layer structures on optical coherence tomography images*. Optics Express, vol. 13, no. 25, pages 10200–10216, 2005. (Cited on page 13.)
- [Camuñas-Mesa 2010] L. Camuñas-Mesa, JA Pérez-Carrasco, C. Zamarreño-Ramos, T. Serrano-Gotarredona and B. Linares-Barranco. *On scalable spiking convnet hardware for cortex-like visual sensory processing systems*. In IEEE International Symposium on Circuits and Systems (ISCAS 2010), pages 249–252. IEEE, 2010. (Cited on page 115.)
- [Candès 2008] E.J. Candès and M.B. Wakin. *An introduction to compressive sampling*. IEEE Signal Processing Magazine, vol. 25, no. 2, pages 21–30, 2008. (Cited on page 166.)
- [Cao 2008] D. Cao, J. Pokorny, V.C. Smith and A.J. Zele. *Rod contributions to color perception: linear with rod contrast*. Vision research, vol. 48, no. 26, pages 2586–2592, 2008. (Cited on page 14.)
- [Cardoso 2006] C.G.S. Cardoso, M.C. Cunha, A. Gomide, D.J. Schiozer and J. Stolfi. *Finite Elements on Dyadic Grids with Applications*. Mathematics and Computers in Simulation, vol. 73, no. 1–4, pages 87–104, November 2006. Applied and Computational Mathematics - Selected Papers of the Fifth PanAmerican Workshop - June 21-25, 2004, Tegucigalpa, Honduras. (Cited on page 44.)
- [Chen 2008] G. Chen, H.D. Lu and A.W. Roe. *A map for horizontal disparity in monkey V2*. Neuron, vol. 58, no. 3, pages 442–450, 2008. (Cited on page 19.)
- [Christopoulos 2000] C. Christopoulos, A. Skodras and T. Ebrahimi. *The JPEG2000 still image coding system: An overview*. IEEE Transactions on Consumer Electronics, vol. 16, no. 4, pages 1103–1127, 2000. (Cited on pages 3 and 96.)

- [Clark 1928] A.B. Clark and et al. *Electrical picture-transmitting system*. US Patent assigned to AT& T, 1928. (Cited on page 126.)
- [Cover 1991] T.M. Cover and J.A. Thomas. *Elements of information theory*. Wiley-Interscience, August 1991. (Cited on pages 8 and 40.)
- [Crouzet 2011] S.M. Crouzet, M. Cauchoi *et al.* *When does the visual system need to look back?* *The Journal of Neuroscience*, vol. 31, no. 24, pages 8706–8707, 2011. (Cited on page 22.)
- [Curcio 1990] C.A. Curcio, K.R. Sloan, R.E. Kalina and A.E. Hendrickson. *Human photoreceptor topography*. *The Journal of comparative neurology*, vol. 292, no. 4, pages 497–523, 1990. (Cited on page 14.)
- [Dacey 2000] D. Dacey, O.S. Packer, L. Diller, D. Brainard, B. Peterson and B. Lee. *Center surround receptive field structure of cone bipolar cells in primate retina*. *Vision Research*, vol. 40, no. 14, pages 1801–1811, 2000. (Cited on page 16.)
- [Daubechies 1990] I. Daubechies. *The wavelet transform, time-frequency localization and signal analysis*. *Information Theory, IEEE Transactions on*, vol. 36, no. 5, pages 961–1005, 1990. (Cited on page 6.)
- [de Ruyter van Steveninck 1997] R.R. de Ruyter van Steveninck, G.D. Lewen, S.P. Strong, R. Koberle and W. Bialek. *Reproducibility and variability in neural spike trains*. *Science*, vol. 275, no. 5307, page 1805, 1997. (Cited on pages 131 and 151.)
- [Delorme 2003] A. Delorme. *Early cortical orientation selectivity: how fast inhibition decodes the order of spike latencies*. *Journal of computational Neuroscience*, vol. 15, no. 3, pages 357–365, 2003. (Cited on pages 41 and 48.)
- [Do 2003] M.N. Do and M. Vetterli. *Framing pyramids*. *IEEE Trans. Signal Process.*, vol. 51, no. 9, pages 2329–2342, 2003. (Cited on pages 66, 67 and 75.)

- [Dowling 1966] JE Dowling and BB Boycott. *Organization of the primate retina: electron microscopy*. Proceedings of the Royal Society of London. Series B, Biological Sciences, pages 80–111, 1966. (Cited on page 12.)
- [Duffin 1952] R. J. Duffin and A. C. Schaeffer. *A Class of Nonharmonic Fourier Series*. Trans. Amer. Math. Soc., vol. 72, no. 2, pages pp. 341–366, 1952. (Cited on page 66.)
- [Escobar 2009] M.J. Escobar, G.S. Masson, T. Vieville and P. Kornprobst. *Action recognition using a bio-inspired feedforward spiking network*. International journal of computer vision, vol. 82, no. 3, pages 284–301, 2009. (Cited on page 25.)
- [Feichtinger 2011] H.G. Feichtinger, J.C. Príncipe, J.L. Romero, A. Singh Alvarado and G.A. Velasco. *Approximate reconstruction of bandlimited functions for the integrate and fire sampler*. Advances in computational mathematics, pages 1–12, 2011. (Cited on page 114.)
- [Felleman 1991] D.J. Felleman and D.C. Van Essen. *Distributed hierarchical processing in the primate cerebral cortex*. Cerebral cortex, vol. 1, no. 1, page 1, 1991. (Cited on page 21.)
- [Field 1994] D.J. Field. *What is the goal of sensory coding?* Neural Comput., vol. 6, no. 4, pages 559–601, 1994. (Cited on pages 44 and 143.)
- [Freeman 2001] W.J. Freeman and J.M. Barrie. *Chaotic oscillations and the genesis of meaning in cerebral cortex*. Nonlinear dynamics in the life and social sciences, vol. 320, page 45, 2001. (Cited on page 27.)
- [Fu 2005] Y. Fu, H.W. Liao, M.T.H. Do and K.W. Yau. *Non-image-forming ocular photoreception in vertebrates*. Current opinion in neurobiology, vol. 15, no. 4, pages 415–422, 2005. (Cited on page 12.)
- [Garey 1982] MR Garey, D. Johnson and H. Witsenhausen. *The complexity of the generalized Lloyd-max problem (corresp.)*. IEEE Transactions on Information Theory, vol. 28, no. 2, pages 255–256, 1982. (Cited on page 129.)

- [Gaudart 1993] L. Gaudart, J. Crebassa and JP Petrakian. *Wavelet transform in human visual channels*. Applied optics, vol. 32, no. 22, pages 4119–4127, 1993. (Cited on page 109.)
- [Gautrais 1998] J. Gautrais and S. Thorpe. *Rate coding vs temporal order coding : a theoretical approach*. Biosystems, vol. 48, no. 1, pages 57–65, 1998. (Cited on page 28.)
- [Gersho 1992] A. Gersho and R.M. Gray. Vector quantization and signal compression, volume 159. Springer, 1992. (Cited on pages 9 and 10.)
- [Gerstner 2000] W. Gerstner. *Population dynamics of spiking neurons: fast transients, asynchronous states, and locking*. Neural computation, vol. 12, no. 1, pages 43–89, 2000. (Cited on page 28.)
- [Gerstner 2002] W. Gerstner and W. Kistler. Spiking neuron models : Single neurons, populations, plasticity. Cambridge University Press, 2002. (Cited on pages 23, 28 and 30.)
- [Girod 1993] B. Girod. *What’s wrong with mean-squared error?* pages 207–220, 1993. (Cited on pages 92 and 96.)
- [Gollisch 2008] T. Gollisch and M. Meister. *Rapid Neural Coding in the Retina with Relative Spike Latencies*. Science, vol. 319, no. 5866, pages 1108–1111, 2008. (Cited on pages 29 and 40.)
- [Gollisch 2010] T. Gollisch and M. Meister. *Eye smarter than scientists believed: Neural computations in circuits of the retina*. Neuron, vol. 65, no. 2, pages 150–164, 2010. (Cited on pages 14, 15 and 103.)
- [Golub 1996] G.H. Golub and C.F. Van Loan. Matrix computations. Johns Hopkins Univ Pr, 1996. (Cited on page 57.)
- [Goodale 1992] M.A. Goodale and A.D. Milner. *Separate visual pathways for perception and action*. Trends in neurosciences, vol. 15, no. 1, pages 20–25, 1992. (Cited on pages 3, 17, 18 and 19.)
- [Goodale 2004] M.A. Goodale and D.A. Westwood. *An evolving view of duplex vision: separate but interacting cortical pathways for perception and*

- action*. Current opinion in Neurobiology, vol. 14, no. 2, pages 203–211, 2004. (Cited on pages 18 and 19.)
- [Grill-Spector 2001] K. Grill-Spector, Z. Kourtzi and N. Kanwisher. *The lateral occipital complex and its role in object recognition*. Vision research, vol. 41, no. 10-11, pages 1409–1422, 2001. (Cited on page 19.)
- [Gross 1973] C.G. Gross. *Inferotemporal cortex and vision*. Progress in physiological psychology, vol. 5, pages 77–123, 1973. (Cited on page 19.)
- [Guo 2007] X. Guo, X. Qi and J.G. Harris. *A time-to-first-spike CMOS image sensor*. Sensors Journal, IEEE, vol. 7, no. 8, pages 1165–1175, 2007. (Cited on page 29.)
- [Hammond 1974] P. Hammond. *Cat retinal ganglion cells: size and shape of receptive field centres*. The Journal of Physiology, vol. 242, no. 1, pages 99–118, October 1974. (Cited on page 46.)
- [Hoang 2009] T.M.N. Hoang, S. Ragot, M. Oger and M. Antonini. *A new bit-plane coder for scalable transform audio coding*. In IEEE International Conference on Acoustics, Speech and Signal Processing 2009 (ICASSP 2009), pages 4137–4140. IEEE, 2009. (Cited on page 92.)
- [Hupe 1998] JM Hupe, AC James, BR Payne, SG Lomber, P. Girard and J. Bullier. *Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons*. Nature, vol. 394, no. 6695, pages 784–787, 1998. (Cited on page 21.)
- [Johansson 2004] R.S. Johansson and I. Birznieks. *First spikes in ensembles of human tactile afferents code complex spatial fingertip events*. Nature Neuroscience, vol. 7, no. 2, pages 170–177, 2004. (Cited on pages 4 and 29.)
- [Katsavounidis 1997] I. Katsavounidis and C.C. Jay Kuo. *A multiscale error diffusion technique for digital halftoning*. IEEE Transactions on Image Processing, vol. 6, no. 3, pages 483–490, 1997. (Cited on page 152.)

- [Keat 2001] J. Keat, P. Reinagel, R.C. Reid and M. Meister. *Predicting every spike: a model for the responses of visual neurons*. *Neuron*, vol. 30, no. 3, pages 803–817, 2001. (Cited on page 114.)
- [Kier 1995] C.K. Kier, G. Buchsbaum and P. Sterling. *How retinal microcircuits scale for ganglion cells of different size*. *The Journal of Neuroscience*, vol. 15, no. 11, pages 7673–7683, 1995. (Cited on page 152.)
- [Kim 2001] K.J. Kim and F. Rieke. *Temporal contrast adaptation in the input and output signals of salamander retinal ganglion cells*. *The Journal of Neuroscience*, vol. 21, no. 1, pages 287–299, 2001. (Cited on page 112.)
- [Koivisto 2011] M. Koivisto, H. Railo, A. Revonsuo, S. Vanni and N. Salminen-Vaparanta. *Recurrent processing in V1/V2 contributes to categorization of natural scenes*. *The Journal of Neuroscience*, vol. 31, no. 7, pages 2488–2492, 2011. (Cited on page 22.)
- [Kostal 2007] L. Kostal, P. Lansky and J.P. Rospars. *Neuronal coding and spiking randomness*. *European Journal of Neuroscience*, vol. 26, no. 10, pages 2693–2701, 2007. (Cited on page 114.)
- [Kovacevic 2008] J. Kovacevic and A. Chebira. An introduction to frames. *Foundations and Trends in Signal Processing*, Now Pub, 2008. (Cited on pages 66, 70 and 73.)
- [Langdon 1984] G.G. Langdon. *An introduction to arithmetic coding*. *IBM Journal of Research and Development*, vol. 28, no. 2, pages 135–149, 1984. (Cited on page 93.)
- [Lazar 2004] A.A. Lazar. *Time encoding with an integrate-and-fire neuron with a refractory period*. *Neurocomputing*, vol. 58, pages 53–58, 2004. (Cited on page 115.)
- [Lazar 2008] A.A. Lazar and E.A. Pnevmatikakis. *A video time encoding machine*. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 717–720. IEEE, 2008. (Cited on page 115.)

- [Lee 1998] B.B. Lee, J. Kremers and T. Yeh. *Receptive fields of primate retinal ganglion cells studied with a novel technique*. *Visual Neuroscience*, vol. 15, no. 1, pages 161–175, January–February 1998. (Cited on page 46.)
- [Lehky 1996] S.R. Lehky and J.H.R. Maunsell. *No binocular rivalry in the LGN of alert macaque monkeys*. *Vision Research*, vol. 36, no. 9, pages 1225–1234, 1996. (Cited on page 17.)
- [Lesica 2004] N.A. Lesica and G.B. Stanley. *Encoding of Natural Scene Movies by Tonic and Burst Spikes in the Lateral Geniculate Nucleus*. *Journal of Neuroscience*, vol. 24, no. 47, pages 10731–10740, November 2004. (Cited on page 165.)
- [Li 2002] Z. Li. *A saliency map in primary visual cortex*. *Trends in Cognitive Sciences*, vol. 6, no. 1, pages 9–16, 2002. (Cited on page 19.)
- [Livingstone 1988] M. Livingstone and D. Hubel. *Segregation of form, color, movement, and depth: anatomy, physiology, and perception*. *Science*, vol. 240, no. 4853, page 740, 1988. (Cited on page 16.)
- [Mainen 1995] Z.F. Mainen and T.J. Sejnowski. *Reliability of spike timing in neocortical neurons*. *Science*, vol. 268, no. 5216, pages 1503–1506, 1995. (Cited on page 4.)
- [Mallat 1989] S.G. Mallat. *Multifrequency channel decompositions of images and wavelet models*. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 37, no. 12, pages 2091–2110, 1989. (Cited on page 109.)
- [Marpe 2001] D. Marpe, G. Blattermann, G. Heising and T. Wiegand. *Video compression using context-based adaptive arithmetic coding*. In *IEEE International Conference on Image Processing 2001. (ICIP 2001)*, volume 3, pages 558–561. IEEE, 2001. (Cited on page 95.)
- [Masmoudi 2010a] K. Masmoudi, M. Antonini and P. Kornprobst. *Another look at the retina as an image dithered scalar quantizer*. In *11th International Workshop on Image Analysis for Multimedia Interactive*

- Services (WIAMIS), pages 1–4. IEEE, 2010. (Cited on pages x, 106, 131 and 151.)
- [Masmoudi 2010b] K. Masmoudi, M. Antonini and P. Kornprobst. *Another look at the retina as an image scalar quantizer*. In IEEE International Symposium on Circuits and Systems (ISCAS 2010), pages 3076–3079. IEEE, 2010. (Cited on pages x and 106.)
- [Masmoudi 2010c] K. Masmoudi, M. Antonini and P. Kornprobst. *Encoding and decoding stimuli using a biologically realistic model: The non-determinism in spike timings seen as a dither signal*. page 75, 2010. (Cited on pages x, 106, 131 and 151.)
- [Masmoudi 2010d] K. Masmoudi, M. Antonini, P. Kornprobst and L. Perinet. *A novel bio-inspired static image compression scheme for noisy data transmission over low-bandwidth channels*. In IEEE International Conference on Acoustics, Speech and Signal Processing 2010 (ICASSP 2010), pages 3506–3509. IEEE, 2010. (Cited on pages ix, 56, 57, 59, 62 and 82.)
- [Masmoudi 2011] K. Masmoudi, M. Antonini and P. Kornprobst. *A bio-inspired image coder with temporal scalability*. In Advanced Concepts for Intelligent Vision Systems, pages 447–458. Springer Berlin/Heidelberg, 2011. (Cited on pages x, 138 and 143.)
- [Masmoudi 2012a] K. Masmoudi, M. Antonini and P. Kornprobst. *Frames for exact inversion of the rank order coder*. IEEE Transactions on Neural Networks and Learning Systems, vol. 23, pages 353–359, 2012. (Cited on pages ix, 54, 84 and 147.)
- [Masmoudi 2012b] K. Masmoudi, M. Antonini and P. Kornprobst. *A perfectly invertible rank order coder*. In SciTePress, editeur, Proceedings of the International Conference on Bio-inspired Systems and Signal Processing, pages 153–162, 2012. (Cited on pages ix, 54 and 84.)
- [Masmoudi 2012c] K. Masmoudi, M. Antonini and P. Kornprobst. *Streaming an image through the eye: The retina seen as a dithered scalable image*

- coder.* to appear in Signal Processing: Image Communication, 2012. (Cited on pages x, 138 and 143.)
- [Masquelier 2011] T. Masquelier. *Relative spike time coding and STDP-based orientation selectivity in the early visual system in natural continuous and saccadic vision: a computational model.* Journal of computational neuroscience, 2011. (Cited on page 109.)
- [Maunsell 1999] J.H.R. Maunsell, G.M. Ghose, J.A. Assad, C.J. McADAMS, C.E. Boudreau, B.D. Noerager *et al.* *Visual response latencies of magnocellular and parvocellular LGN neurons in macaque monkeys.* Visual Neuroscience, vol. 16, no. 1, pages 1–14, 1999. (Cited on page 16.)
- [Mazer 2003] J.A. Mazer and J.L. Gallant. *Goal-related activity in V4 during free viewing visual search: Evidence for a ventral stream visual salience map.* Neuron, vol. 40, no. 6, pages 1241–1250, 2003. (Cited on page 19.)
- [Meister 1999] M. Meister and M.J. Berry. *The neural code of the retina.* Neuron, pages 435–450, 1999. (Cited on page 103.)
- [Moxon 2000] K.A. Moxon, J. Morizio, J.K. Chapin, MA Nicolelis and P. Wolf. *Designing a brain-machine interface for neuroprosthetic control.* CRC Press, Boca Raton, 2000. (Cited on page 166.)
- [Nauge 2010] M. Nauge, MC Larabi and C. Fernandez. *Benchmark de métriques de qualité sur bases de données d'images compressées.* In CORESA 2010, 2010. (Cited on page 8.)
- [Nawrot 1999] M. Nawrot, A. Aertsen and S. Rotter. *Single-trial estimation of neuronal firing rates: from single-neuron spike trains to population activity.* Journal of neuroscience methods, vol. 94, no. 1, pages 81–92, 1999. (Cited on page 27.)
- [Nealey 1994] TA Nealey and JH Maunsell. *Magnocellular and parvocellular contributions to the responses of neurons in macaque striate cortex.* The Journal of neuroscience, vol. 14, no. 4, page 2069, 1994. (Cited on page 16.)

- [Nelson 1996] M. Nelson and J.L. Gailly. The data compression book, volume 95. M&T Books, 1996. (Cited on page 10.)
- [Ockleford 1977] EM Ockleford, AD Milner, W. Dewar and IA Sneddon. *Form perception in stump-tail macaques following posterior parietal and lateral frontal lesions*. *Cortex; a journal devoted to the study of the nervous system and behavior*, vol. 13, no. 4, page 361, 1977. (Cited on page 19.)
- [Olshausen 1996] B.A. Olshausen and D.J. Field. *Emergence of simple-cell receptive-field properties by learning a sparse code for natural images*. *Nature*, vol. 381, no. 6583, pages 607–609, June 1996. (Cited on page 88.)
- [Oram 1992] M.W. Oram and D.I. Perrett. *Time Course of Neural Responses Discriminating Different Views of the Face and Head*. *Journal of Neurophysiology*, vol. 68, no. 1, pages 70–84, July 1992. (Cited on page 39.)
- [Oswald 2007] A.M.M. Oswald, B. Doiron and L. Maler. *Interval coding. I. Burst interspike intervals as indicators of stimulus intensity*. *Journal of neurophysiology*, vol. 97, no. 4, page 2731, 2007. (Cited on page 29.)
- [Panzeri 2001] S. Panzeri, R.S. Petersen, S.R. Schultz, M. Lebedev and M.E. Diamond. *The role of spike timing in the coding of stimulus location in rat somatosensory cortex*. *Neuron*, vol. 29, no. 3, pages 769–777, 2001. (Cited on pages 29 and 131.)
- [Pascual-Leone 2001] A. Pascual-Leone and V. Walsh. *Fast backprojections from the motion to the primary visual area necessary for visual awareness*. *Science*, vol. 292, no. 5516, page 510, 2001. (Cited on page 21.)
- [Perkell 1968a] D. Perkell and T. Bullock. *Neural Coding*. *Neurosciences Research Program Bulletin*, vol. 6, pages 221–343, 1968. (Cited on page 25.)
- [Perkell 1968b] D. Perkell and T. Bullock. *Neural Coding*. *Neurosciences Research Program Bulletin*, vol. 6, pages 221–343, 1968. (Cited on page 131.)

- [Perrett 1982] DI Perrett, ET Rolls and W. Caan. *Visual neurones responsive to faces in the monkey temporal cortex*. Experimental brain research, vol. 47, no. 3, pages 329–342, 1982. (Cited on page 39.)
- [Perrinet 2003] L. Perrinet. *Comment déchiffrer le code impulsionnel de la Vision? Étude du flux parallèle, asynchrone et épars dans le traitement visuel ultra-rapide*. PhD thesis, Université Paul SABATIER, February 2003. (Cited on pages 48, 49 and 87.)
- [Perrinet 2004] L. Perrinet, M. Samuelides and S. Thorpe. *Coding static natural images using spiking event times: do neurons Cooperate?* IEEE Trans. Neural Netw., vol. 15, no. 5, pages 1164–1175, 2004. (Cited on pages 48, 56, 57, 74 and 76.)
- [Phillips 1998] DP Phillips. *Factors shaping the response latencies of neurons in the cat's auditory cortex*. Behavioural brain research, vol. 93, no. 1-2, pages 33–41, 1998. (Cited on page 29.)
- [Purpura 1988] K. Purpura, E. Kaplan and RM Shapley. *Background light and the contrast gain of primate P and M retinal ganglion cells*. Proceedings of the National Academy of Sciences, vol. 85, no. 12, page 4534, 1988. (Cited on page 16.)
- [Rabbani 1991] M. Rabbani and P.W. Jones. *Digital image compression techniques*. SPIE-International Society for Optical Engineering, 1991. (Cited on page 3.)
- [Rakshit 1995] S. Rakshit and C.H. Anderson. *Error correction with frames: the Burt-Laplacian pyramid*. IEEE Trans. Inf. Theory, vol. 41, no. 6, pages 2091–2093, 1995. (Cited on pages 66 and 75.)
- [Rastogi 2011] M. Rastogi, A. Singh Alvarado, J.G. Harris and J.C. Principe. *Integrate and Fire Circuit as an ADC replacement*. In IEEE International Symposium on Circuits and Systems (ISCAS 2011), pages 2421–2424. IEEE, 2011. (Cited on pages 114 and 166.)
- [Rieke 1997] F. Rieke, D. Warland, R. de Ruyter van Steveninck and W. Bialek. *Spikes: Exploring the neural code*. The MIT Press, Cambridge, MA, USA, 1997. (Cited on pages 3, 23, 27, 99 and 165.)

- [Rieke 2001] F. Rieke. *Temporal contrast adaptation in salamander bipolar cells*. *Journal of Neuroscience*, vol. 21, no. 23, pages 9445–9454, December 2001. (Cited on page 112.)
- [Rodieck 1965] R.W. Rodieck. *Quantitative analysis of the cat retinal ganglion cells response to visual stimuli*. *Vision Research*, vol. 5, no. 11, pages 583–601, 1965. (Cited on page 44.)
- [Sayood 2000] K. Sayood. *Introduction to data compression*, second edition (morgan kaufmann series in multimedia information and systems). Morgan Kaufmann, 2000. (Cited on pages 90, 92 and 93.)
- [Schiller 1990] P.H. Schiller and N.K. Logothetis. *The color-opponent and broad-band channels of the primate visual system*. *Trends in Neurosciences*, vol. 13, no. 10, pages 392–398, 1990. (Cited on page 17.)
- [Schmolesky 1998] M.T. Schmolesky, Y. Wang, D.P. Hanes, K.G. Thompson, S. Leutgeb, J.D. Schall and A.G. Leventhal. *Signal timing across the macaque visual system*. *Journal of Neurophysiology*, vol. 79, no. 6, page 3272, 1998. (Cited on pages 18 and 19.)
- [Serre 2007] T. Serre, A. Oliva and T. Poggio. *A feedforward architecture accounts for rapid categorization*. *Proceedings of the National Academy of Sciences*, vol. 104, no. 15, page 6424, 2007. (Cited on page 22.)
- [Shadlen 1998] M.N. Shadlen and W.T. Newsome. *Noise, neural codes and cortical organization*. *Findings and Current Opinion Cognitive Neuroscience*, vol. 4, pages 569–79, 1998. (Cited on pages 131 and 151.)
- [Shapley 1978] RM Shapley and JD Victor. *The effect of contrast on the transfer properties of cat retinal ganglion cells*. *The Journal of Physiology*, vol. 285, no. 1, pages 275–298, 1978. (Cited on page 112.)
- [Shih 2011] J.Y. Shih, C.A. Atencio and C.E. Schreiner. *Improved stimulus representation by short interspike intervals in primary auditory cortex*. *Journal of Neurophysiology*, vol. 105, no. 4, page 1908, 2011. (Cited on page 29.)

- [Skodras 2001] A. Skodras, C. Christopoulos and T. Ebrahimi. *JPEG2000: The upcoming still image compression standard*. Pattern Recognition Letters, vol. 22, no. 12, pages 1337–1345, 2001. (Cited on pages 3 and 96.)
- [Stein 2005] R.B. Stein, E.R. Gossen and K.E. Jones. *Neuronal variability: noise or part of the signal?* Nature Reviews Neuroscience, vol. 6, no. 5, pages 389–397, 2005. (Cited on pages 26, 28 and 114.)
- [Sterling 1992] P. Sterling, E. Cohen, R.G. Smith and Y. Tsukamoto. *Retinal circuits for daylight: why ballplayers don't wear shades*. Analysis and Modeling of Neural Systems, pages 143–162, 1992. (Cited on page 140.)
- [Tallon-Baudry 2011] C. Tallon-Baudry, F. Meyniel and S. Bourgeois-Gironde. *Fast and Automatic Activation of an Abstract Representation of Money in the Human Ventral Visual Pathway*. PloS one, vol. 6, no. 11, page e28229, 2011. (Cited on pages 22 and 40.)
- [Tanaka 2009] H. Tanaka and I. Ohzawa. *Surround suppression of V1 neurons mediates orientation-based representation of high-order visual features*. Journal of neurophysiology, vol. 101, no. 3, page 1444, 2009. (Cited on page 19.)
- [Taubman 2000] D. Taubman. *High performance scalable image compression with EBCOT*. IEEE transactions on Image Processing, vol. 9, no. 7, pages 1158–1170, 2000. (Cited on page 150.)
- [Thorpe 1989] S.J. Thorpe and M. Imbert. *Biological constraints on connectionist modelling*. Connectionism in perspective, pages 63–92, 1989. (Cited on page 38.)
- [Thorpe 1990] S. Thorpe. *Spike arrival times: A highly efficient coding scheme for neural networks*. In R. Eckmiller, G. Hartmann and G. Hauske, editors, Parallel Processing in Neural Systems and Computers, pages 91–94. Elsevier, 1990. (Cited on pages 4, 35, 37, 38, 39, 41 and 42.)

- [Thorpe 1996] S. Thorpe, D. Fize and C. Marlot. *Speed of processing in the human visual system*. Nature, vol. 381, pages 520–522, 1996. (Cited on pages 22, 29, 40 and 41.)
- [Thorpe 1998] S.J. Thorpe and J. Gautrais. *Rank order coding*. Computational Neuroscience: Trends in Research, vol. 13, pages 113–119, 1998. (Cited on page 84.)
- [Tsai 1996] M.J. Tsai, J.D. Villasenor and F. Chen. *Stack-run image coding*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 6, no. 5, pages 519–521, October 1996. (Cited on pages 90, 91 and 92.)
- [Ungerleider 2000] S.K.L.G. Ungerleider. *Mechanisms of visual attention in the human cortex*. Annual review of neuroscience, vol. 23, no. 1, pages 315–341, 2000. (Cited on page 21.)
- [Van Rullen 2001a] R. Van Rullen and S. Thorpe. *Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artificial objects*. Perception, vol. 30, no. 6, pages 655–668, 2001. (Cited on pages 40 and 117.)
- [Van Rullen 2001b] R. Van Rullen and S. Thorpe. *Rate Coding versus temporal order coding: What the Retinal ganglion cells tell the visual cortex*. Neural Comput., vol. 13, pages 1255–1283, 2001. (Cited on pages ii, vi, ix, x, 31, 35, 36, 37, 41, 43, 44, 46, 47, 48, 49, 50, 51, 52, 53, 55, 56, 57, 58, 60, 61, 66, 67, 75, 79, 84, 97, 103, 121, 139, 142, 143 and 163.)
- [Van Rullen 2002] R. Van Rullen and S. Thorpe. *Surfing a spike wave down the ventral stream*. Vision Research, vol. 42, no. 23, pages 2593–2615, October 2002. (Cited on pages 30, 31, 39, 41, 46 and 48.)
- [VanRullen 2005] R. VanRullen, R. Guyonneau and S.J. Thorpe. *Spike times make sense*. Trends in neurosciences, vol. 28, no. 1, pages 1–4, 2005. (Cited on pages 4, 29 and 30.)
- [Vetterli 1995] M. Vetterli and J. Kovačević. *Wavelets and subband coding*. Prentice Hall PTR Upper Saddle River, NJ, 1995. (Cited on page 6.)

- [Vidne 2012] M. Vidne, Y. Ahmadian, J. Shlens, J.W. Pillow, J. Kulkarni, A.M. Litke, EJ Chichilnisky, E. Simoncelli and L. Paninski. *Modeling the impact of common noise inputs on the network activity of retinal ganglion cells*. to appear in Journal of Computational Neuroscience, 2012. (Cited on pages 106, 161 and 165.)
- [Wallace 1992] G.K. Wallace. *The JPEG still picture compression standard*. Consumer Electronics, IEEE Transactions on, vol. 38, no. 1, pages xviii–xxxiv, 1992. (Cited on page 3.)
- [Walsh 1996] V. Walsh and S.R. Butler. *The effects of visual cortex lesions on the perception of rotated shapes*. Behavioural brain research, vol. 76, no. 1-2, pages 127–142, 1996. (Cited on page 19.)
- [Wang 2004a] Chung-Neng Wang, Chi-Min Liu and Tihao Chiang. *Perceptual dithering for octave subband image coding*. Journal of Visual Communication and Image Representation, vol. 15, no. 2, pages 163 – 184, 2004. (Cited on page 151.)
- [Wang 2004b] Z. Wang, A.C. Bovik, H.R. Sheikh and E.P. Simoncelli. *Image quality assessment: from error visibility to structural similarity*. IEEE Trans. Image Process., vol. 13, no. 4, pages 600–612, 2004. (Cited on pages 74 and 96.)
- [Wang 2004c] Z. Wang, A.C. Bovik, H.R. Sheikh and E.P. Simoncelli. *Image quality assessment: From error visibility to structural similarity*. IEEE Transactions on Image Processing, vol. 13, no. 4, pages 600–612, 2004. (Cited on page 148.)
- [Wannamaker 2000] R.A. Wannamaker, S.P. Lipshitz, J. Vanderkooy and J.N. Wright. *A theory of nonsubtractive dither*. IEEE Transactions on Signal Processing, vol. 48, no. 2, pages 499–516, 2000. (Cited on pages 131, 132, 133, 153, 156 and 158.)
- [Wannamaker 2004] R.A. Wannamaker. *Subtractive and nonsubtractive dithering: A mathematical comparison*. J. Audio Eng. Soc, vol. 52, no. 12, pages 1211–1227, 2004. (Cited on page 153.)

- [Watson 1987] A.B. Watson. *The cortex transform: rapid computation of simulated neural images*. Computer vision, graphics, and image processing, vol. 39, no. 3, pages 311–327, 1987. (Cited on page 6.)
- [Witten 1987] I.H. Witten, R.M. Neal and J.G. Cleary. *Arithmetic coding for data compression*. Communications of the ACM, vol. 30, no. 6, pages 520–540, 1987. (Cited on page 95.)
- [Wohrer 2008] A. Wohrer. *Model and large-scale simulator of a biological retina, with contrast gain control*. PhD thesis, Graduate School of Information and Communication Sciences, University of Nice-Sophia Antipolis, 2008. (Cited on pages 103, 104 and 132.)
- [Wohrer 2009a] A. Wohrer and P. Kornprobst. *Virtual Retina : A biological retina model and simulator, with contrast gain control*. Journal of Computational Neuroscience, vol. 26, no. 2, pages 219–249, 2009. (Cited on pages ii, vii, x, 99, 103, 106, 108, 110, 111, 112, 139, 141, 143 and 163.)
- [Wohrer 2009b] Adrien Wohrer, Pierre Kornprobst and Marc Antonini. *Retinal filtering and image reconstruction*. Research Report RR-6960, INRIA, 2009. (Cited on page 140.)
- [Zhang 2005] Y. Zhang, A. Ghodrati and D.H. Brooks. *An analytical comparison of three spatio-temporal regularization methods for dynamic linear inverse problems in a common statistical framework*. Inverse Problems, vol. 21, page 357, 2005. (Cited on page 140.)
- [Zohar 2011] O. Zohar, T.M. Shackleton, I. Nelken, A.R. Palmer and M. Shamir. *First Spike Latency Code for Interaural Phase Difference Discrimination in the Guinea Pig Inferior Colliculus*. The Journal of Neuroscience, vol. 31, no. 25, page 9192, 2011. (Cited on page 29.)