



ETSIT
UPM



ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE TELECOMUNICACIÓN
UNIVERSIDAD POLITÉCNICA DE MADRID

INSTITUTO DE ÓPTICA
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Tesis Doctoral

New contributions in overcomplete image
representations inspired from the functional
architecture of the primary visual cortex

Nuevas contribuciones en representaciones sobrecompletas de
imágenes inspiradas por la arquitectura funcional de la corteza
visual primaria

SYLVAIN GAEL FREDERIC FISCHER

Directores de Tesis:

GABRIEL CRISTÓBAL PÉREZ

ANDRÉS SANTOS Y LLÉO

2007

Tribunal nombrado por el Mgfc. y Excmo. Sr. Rector de la Universidad Politécnica de Madrid, el día

PRESIDENTE:

VOCAL:

VOCAL:

VOCAL:

SECRETARIO:

Realizado el acto de defensa de la Tesis en Madrid, el día
acuerda otorgarle la calificación de:

EL PRESIDENTE

EL SECRETARIO

LOS VOCALES

Abstract

The present thesis aims at investigating parallelisms between the functional architecture of primary visual areas and image processing methods. A first objective is to refine existing models of biological vision on the base of information theory statements and a second is to develop original solutions for image processing inspired from natural vision. The available data on visual systems contains physiological and psychophysical studies, Gestalt psychology and statistics on natural images. The thesis is mostly centered in overcomplete representations (i.e. representations increasing the dimensionality of the data) for multiple reasons. First because they allow to overcome existing drawbacks of critically sampled transforms, second because biological vision models appear overcomplete and third because building efficient overcomplete representations raises challenging and actual mathematical problems, in particular the problem of sparse approximation.

The thesis proposes first a self-invertible log-Gabor wavelet transformation inspired from the receptive field and multiresolution arrangement of the simple cells in the primary visual cortex (V1). This transform shows promising abilities for noise elimination. Second, interactions observed between V1 cells consisting in lateral inhibition and in facilitation between aligned cells are shown efficient for extracting edges of natural images. As a third point, the redundancy introduced by the overcompleteness is reduced by a dedicated sparse approximation algorithm which builds a sparse representation of the images based on their edge content. For an additional decorrelation of the image information and for improving the image compression performances, edges arranged along continuous contours are coded in a predictive manner through chains of coefficients. This offers then an efficient representation of contours. Fourth, a study on contour completion using the tensor voting framework based on Gestalt psychology is presented. There, the use of iterations and of the curvature information allow to improve the robustness and the perceptual quality of the existing method.

Resumen

La presente tesis doctoral tiene como objetivo indagar en algunos paralelismos entre la arquitectura funcional de las áreas visuales primarias y el tratamiento de imágenes. Un primer objetivo consiste en mejorar los modelos existentes de visión biológica basándose en la teoría de la información. Un segundo es el desarrollo de nuevos algoritmos de tratamiento de imágenes inspirados de la visión natural. Los datos disponibles sobre el sistema visual abarcan estudios fisiológicos y psicofísicos, psicología Gestalt y estadísticas de las imágenes naturales. La tesis se centra principalmente en las representaciones sobrecompletas (i.e. representaciones que incrementan la dimensionalidad de los datos) por las siguientes razones. Primero porque permiten sobrepasar importantes desventajas de las transformaciones ortogonales; segundo porque los modelos de visión biológica necesitan a menudo ser sobrecompletos y tercero porque construir representaciones sobrecompletas eficientes involucra problemas matemáticos relevantes y novedosos, en particular el problema de las aproximaciones dispersas.

La tesis propone primero una transformación en ondículas log-Gabor auto-inversible inspirada del campo receptivo y la organización en multiresolución de las células simples del cortex visual primario (V1). Esta transformación ofrece resultados prometedores para la eliminación del ruido. En segundo lugar, las interacciones observadas entre las células de V1 que consisten en la inhibición lateral y en la facilitación entre células alineadas se han mostrado eficientes para extraer los bordes de las imágenes naturales. En tercer lugar, la redundancia introducida por la transformación sobrecompleta se reduce gracias a un algoritmo dedicado de aproximación dispersa el cual construye una representación dispersa de las imágenes sobre la base de sus bordes. Para una decorrelación adicional y para conseguir más altas tasas de compresión, los bordes alineados a lo largo de contornos continuos están codificado de manera predictiva por cadenas de coeficientes, lo que ofrece una representación eficiente de los contornos. Finalmente se presenta un estudio sobre el cierre de contornos utilizando la metodología de tensor voting. Proponemos el uso de iteraciones y de la información de curvatura para mejorar la robustez y la calidad perceptual de los métodos existentes.

Resumé

La présente thèse doctorale propose d'étudier en parallèle l'architecture fonctionnelle de aires visuelles primaires et les méthodes de traitement d'image. Un premier objectif consiste à améliorer les modèles de vision biologique sur la base de la théorie de l'information, et un second est le développement de solutions originales inspirées de la vision naturelle pour le traitement des images. Les données disponibles sur le système visuel recouvrent des études physiologiques et psychophysiques, la psychologie Gestalt et les statistiques des images naturelles. La thèse est centrée principalement sur les représentations surcomplètes (c'est-à-dire des représentations qui accroissent la dimensionnalité des données) pour de multiples raisons. Premièrement parce que ces représentations permettent de résoudre les difficultés inhérentes aux transformations à échantillonnage critique. Deuxièmement parce que les modèles de vision biologique ont généralement besoin d'être surcomplètes et troisièmement parce que construire des représentations surcomplètes efficaces pose des problèmes mathématiques actuels et importants, en particulier le problème de l'approximation parcimonieuse.

La thèse propose premièrement une transformation en ondelettes log-Gabor auto-inversible inspirée des champs récepteurs et de l'organisation en multirésolution des cellules simples du cortex visuel primaire (V1). Cette transformation possède des capacités prometteuses pour l'élimination du bruit. La deuxième partie de la thèse consiste à modéliser les non-linéarités et les interactions entre les cellules de V1, en particulier les inhibitions latérales et les facilitations entre les cellules alignées. Cette modélisation se révèle efficace pour l'extraction des bords des images naturelles. Troisièmement, la redondance introduite par la transformation surcomplète est réduite grâce à un algorithme dédié qui construit une approximation parcimonieuse des images sur la base de leur contenu en bords. Pour une réduction supplémentaire de la redondance et pour améliorer les performances de compression d'image, les bords organisés en contours continus sont codés de manière prédictible par des chaînes de coefficients. Cette codification offre ainsi une représentation efficace des contours. En quatrième point nous présentons une étude sur la fermeture des contours utilisant la méthodologie de vote par tenseurs elle-même basée sur la psychologie Gestalt. Nous proposons l'implémentation itérative et l'usage de l'information de courbature pour améliorer la robustesse et la qualité perceptuelle.

A Laura

Agradezco a Gabriel Cristóbal, mi director de tesis, el haberme aportado el apoyo necesario para llevar a cabo este proyecto; y a Andrés Santos, el haberme interesado en el tema del sistema visual y la bioingeniería. Me alegro de haber colaborado durante toda la tesis con mi compañero Rafa.

Agradecimientos a los compañeros del Instituto, cuya solidaridad me permitió seguir adelante con la tesis, a pesar de un cierto autoritarismo en algunas instituciones, en el CSIC en particular que permite expulsar doctorandos sin ofrecerles la posibilidad de defenderse y sin ni siquiera informarles de los procesos en curso. Agradecimientos en particular a Raul, Oscar, Sergio, Lourdes por su apoyo. Agradecimientos también a los investigadores que respetan a los doctorandos, en particular a Gabriel, Susana y a Priscilla.

Agradecimientos al Ministerio de Educación y Ciencia (MEC) por la beca FPU ofrecida para la presente tesis. Sin embargo, tengo que lamentar la denegación de dos estancias por motivos todavía sin esclarecer. Afortunadamente, la vida ha sido agradable durante estos años, gracias a mis compañeros, en particular Laura, Sergio, Ana, Rafa, Oscar, Ainara, Filip, Manuel, Lourdes, Patricia, Raul, Amelia, Isi, Zhisong Xiao, Olivares, Mane, Noemi, Jeremie, Vincenzo, Hector, Victor, Casper, George, Sergio Ortiz, Elena, Carlos, Salvador, Armando, Paquita; a los de afuera, Jose, Justo, Begoña, Celeste, Tatiana, Junior, Suzanne, Eike, Paul, Tobias, Isabel, Maroto, Guillermo, Cano, Fleur, Nuria, Dani, Urue, Anne-So, y los que faltan en esta lista. Gracias a Josefina, Manuel, Albert, Ramón y Ludivina por su hospitalidad y su simpatía.

Remerciements à mes parents et à Irène, Raphaël, Alexis, Nicolas, Annabelle, Virginie, encouragements à Maxime, Alys et Antoine.

Remerciements aux équipes de l'INCM à Marseille et leurs collaborateurs, qui forment un centre scientifique multidisciplinaire et de bonne ambiance, où des recherches en traitement d'images peuvent s'enrichir des connaissances de biologie, de psychologie, de modélisation et de mathématiques. Remerciements en particulier aux chercheurs Laurents, Frédo, Guillaume Masson, Eric Castet, Daniele, Jean-Luc Velay, Ivo. Merci aussi pour leur présence et le temps passé ensemble à JB, Alex, Seb, Fred, Elisa, Sylvains, Carole, Camille, Viet, Sophie, Alexandrine, Vincent Hok, Stéph, Andreia, Andrea, Ana, Di Pech, Björk, ainsi que Matthieu et le couloir de Cuques; aux anciens camarades retrouvés à Marseille, Vincent, Béatrice, JC, Nader et Blandine. J'aurais aimé rester plus longtemps à Marseille, aussi bien pour collaborer et discuter plus étroitement avec Laurent et l'équipe de chercheurs que pour y vivre avec les camarades qui y demeurent.

Ich danke Prof. Heiko Neumann und Dr. Pierre Bayerl für den Aufenthalt in Ulm und für die einföhrung zu contour ausfüllen and modellierung.

Dekuji Janu Flusserovi a Filipu Šroubkovi, že mi umožnili týdenní stáž v Ústavu teorie informace a automatizace v Praze.

I thank also the researchers who dedicate some time for helping me to improve the present thesis, Laurent Perrinet, Laura Rebollo-Neira, Sandrine Antoine, Nader Yeganefer, Javier Portilla, Baltasar Beferull-Lozano and Peter Schelkens.

Contents

Abstract	iii
Resumen	iv
Resumé	v
Table of contents	x
Resumen extendido	1
0.1 Antecedentes y estado del arte	1
0.2 Ondículas log-Gabor auto-invertibles	6
0.3 Extracción de bordes por modelos energéticos	7
0.4 Representaciones dispersas basadas en extracción de bordes	9
0.5 Cierre de contornos por tensor voting	10
0.6 Principales aportaciones de la presente tesis	11
0.7 Conclusiones	12
1 Introduction	14
1.1 Wavelet representations	14
1.2 Overcomplete representations	15
1.3 Sparse representations	16
1.4 Visual system physiology	18
1.5 Biological modeling	20
1.6 Statistics of natural images	21
1.7 Viewpoints in cognitive sciences	22
1.8 Novelty of the approach	24
1.9 Organization of the thesis and description of the new contributions	25
2 Self-invertible log-Gabor wavelets	27
2.1 Introduction	28
2.2 Method	30
2.2.1 Bandpass log-Gabor filters	30
2.2.2 High-pass oriented filters	34

2.2.3	Low-pass filter	38
2.2.4	Direct log-Gabor wavelet transform	39
2.2.5	Reconstruction transform	39
2.2.6	Matrix notation	40
2.2.7	Sparse downsampling	41
2.2.8	Border strategy	43
2.2.9	Methods for reconstruction improvement and exact reconstruction	43
2.3	Application to image denoising	45
2.3.1	Number of orientations	47
2.3.2	Comparison between different multiresolution schemes	48
2.4	Conclusions	56
3	Edge extraction through energy models	58
3.1	Introduction to edge extraction	58
3.2	Edge extraction using log-Gabor wavelets	60
3.2.1	Hard thresholding for noise elimination	61
3.2.2	Oriented inhibition	62
3.2.3	Facilitation across scales	66
3.2.4	Facilitation in space and across orientations	66
3.3	Results on edge extraction	68
3.4	Conclusion	71
4	Sparse approximation based on edge extraction	72
4.1	Introduction	73
4.2	Methods	75
4.2.1	Non-uniqueness under overcomplete dictionaries	76
4.2.2	The problem of amplitude assignation	77
4.2.3	Iterative sparse approximation algorithm	79
4.2.4	Non-iterative sparse approximation	85
4.2.5	Chain coding	87
4.3	Applications	90
4.3.1	Image compression through iterative sparse approximation . . .	91
4.3.2	Image compression through non-iterative sparse approximation and chain coding	92
4.3.3	Non-iterative sparse approximation applied to denoising	98
4.4	Conclusions	100
5	Contour completion using tensor voting	106
5.1	Introduction	107
5.2	A brief introduction to tensor voting	109
5.3	Iterated tensor voting	111

5.3.1	Example	111
5.3.2	Iterative tensor voting (IT) method	111
5.3.3	Statistics on the influence of iterations	113
5.3.4	Results	114
5.3.5	Further remarks on the use of iterations	115
5.4	Curvature improvement	117
5.4.1	Method	117
5.4.2	Statistical study	120
5.4.3	Hand-written text example	122
5.5	Robustness to noise	123
5.5.1	Evaluation methods	123
5.5.2	Measurement based on average gaps	124
5.5.3	Results on noise robustness	125
5.6	Conclusion	126
6	Conclusion	128
6.1	Further work	130
	Bibliography	131
A	Philosophical viewpoints in cognitive science	147
B	Developing Big Brother	149

Resumen extendido

”El capitalismo de nuestros días exhibe, en su centro universal de poder, una identidad evidente de los monopolios privados y el aparato estatal. Las corporaciones multinacionales utilizan directamente al Estado para acumular, multiplicar y concentrar capitales, profundizar la revolución tecnológica, militarizar la economía y, mediante diversos mecanismos, asegurar el éxito de la norteamericanización del mundo capitalista.”

Eduardo Galeano in *Las Venas Abiertas de America Latina*.

0.1 Antecedentes y estado del arte

Desde hace veinte años las *ondículas* (*wavelets* en inglés) han revolucionado las técnicas de tratamiento de la señal y las imágenes. Fueron introducidas por primera vez por Morlet y colaboradores [132], y los aspectos teóricos y matemáticos fueron desarrollados más adelante por Meyer [130], Mallat [117] y Daubechies [33]. Las ondículas son una alternativa entre las representaciones puramente espaciales (o temporales) y las representaciones frecuenciales de tipo Fourier ya que ofrecen una caracterización de las señales tanto espacial como frecuencial. Se aplican sobre la señal mediante traslaciones y dilataciones de una *ondícula madre* la cual está localizada espacialmente y es específica de una banda de frecuencia. Proporcionan entonces una representación espacio-frecuencial de la señal que se ha mostrado eficiente para el análisis y la caracterización de señales. La estructura *diádica*, es decir la limitación del conjunto de dilataciones posibles de la ondícula madre a potencias de dos, y un submuestreo crítico, permite la construcción de familias de ondículas (bi-) ortogonales [117]. Las ondículas ortogonales implementan los mismos filtros en la descomposición y en la reconstrucción de la señal. La ortogonalidad impone condiciones excesivamente restrictivas sobre la forma de los filtros. En consecuencia se prefieren a menudo las ondículas *biortogonales* que implementan filtros diferentes para la transformación directa e inversa. En el tratamiento de señales digitales, dichas ondículas biortogonales constituyen una referencia obligada. Sin embargo, ambas ondículas ortogonales y biortogonales sufren de

importantes desventajas. Ninguna de las ondículas ortogonales o biortogonales son *invariantes a la traslación*, lo que significa que un leve desplazamiento de la señal en el dominio espacial puede inducir una variación grande (sin proporción) de los coeficientes de la transformación. El problema está relacionado con la presencia de *dispersión de frecuencias* o *aliasing*. Las ondículas están organizadas en bandas de frecuencias que no están totalmente disociadas y que se entremezclan debido al submuestreo crítico. El aliasing se compensa en la transformación inversa y no impide la reconstrucción exacta de la señal. Sin embargo en aplicaciones donde se eliminan coeficientes (compresión, restauración, etc), el aliasing puede dar lugar a la aparición de artefactos. Además, el hecho de que los filtros de las ondículas biortogonales sean diferentes para la descomposición y la reconstrucción puede también vulnerar la robustez de la reconstrucción cuando se eliminan coeficientes, ya que pueden aparecer eventos que no corresponden estrictamente a la señal analizada. Por otra parte, en tratamiento de imágenes se utilizan ondículas (bi-) ortogonales en dos dimensiones separables en X e Y, es decir, compuestas por el producto de dos ondículas de una dimensión, una vertical y una horizontal. Las ondículas así generadas se comportan peor en las direcciones oblicuas.

En los últimos diez años diversas transformaciones alternativas han sido propuestas para solucionar los defectos anteriormente descritos a costa de sacrificar las propiedades de (bi-) ortogonalidad. Las *ondículas no diezmadas* están constituidas por las mismas funciones separables en X e Y que las ondículas (bi-) ortogonales, pero no están submuestreadas, por lo que no introducen aliasing y son invariantes a la traslación. Se han mostrado más eficientes que las ondículas biortogonales, en particular en aplicaciones de eliminación de ruido [28, 111, 20]. Las *steerable pyramids* [182, 159] presentan la propiedad de invariancia a la traslación gracias a un submuestreo menos crítico que permite evitar la introducción de aliasing. Además las *steerable pyramids* ofrecen un mejor comportamiento en direcciones oblicuas gracias a la implementación de filtros de diversas orientaciones. Ondículas en cuadratura de fase sensibles tanto a señales simétricas como antisimétricas han sido también propuestas con éxito para, por ejemplo, el análisis de texturas o la compresión [96, 74, 159]. Otras transformaciones tales como *curvelets* [189, 46] o *contourlets* [41] han sido propuestas principalmente para mejorar la sensibilidad en orientación de las ondículas y adaptarse mejor a la forma de los bordes de contraste de las imágenes naturales.

Por su parte las *ondículas de Gabor* [74, 134, 173] utilizan como ondícula madre una función inicialmente propuesta por Gabor [68] en 1946 y utilizada posteriormente por Morlet en 1982 en las primeras transformaciones en ondículas [132]. Un filtro de Gabor consiste en una Gaussiana modulada por exponencial compleja. La forma es Gaussiana tanto en el dominio espacial como frecuencial, por lo que el filtro está estrechamente localizado tanto en el espacio como en frecuencia. De hecho, se puede demostrar que dicho filtro alcanza el óptimo teórico de localización conjunta en espacio y en frecuencia.

Por lo tanto, ha sido considerado como un filtro óptimo en el tratamiento de la señal [68]. Sin embargo este filtro presenta también desventajas:

1. No es posible construir una base ortogonal completa con dichos filtros.
2. Los filtros no tienen una media nula, lo que introduce una componente continua en los coeficientes correspondientes a bandas intermedias de frecuencia. Este último problema puede, sin embargo, solucionarse mediante la utilización de filtros de log-Gabor, los cuales son funciones Gabor en escala frecuencial logarítmica [53].
3. Es particularmente difícil cubrir las altas y las bajas frecuencias con filtros de Gabor, obtener un recubrimiento uniforme del dominio frecuencial y reconstruir exactamente la señal (las ondículas de Gabor no son invertibles).

Por estas razones, las ondículas de Gabor han sido utilizadas principalmente en aplicaciones que no necesitan reconstrucción de la imagen, tales como el análisis de texturas [27, 158] o la extracción de bordes [175, 102], aunque sus buenas cualidades han permitido también su utilización en la restauración o la eliminación de ruido [101].

El uso de transformaciones no críticamente submuestreadas permite entonces remediar algunas de las desventajas de las ondículas (bi-) ortogonales. Dichas transformaciones son llamadas *sobrecompletas* (*overcomplete*) para indicar que inducen una representación de dimensión superior a la de la señal: contienen más coeficientes de transformación que la señal original contiene de muestras (o píxeles). Consecuentemente, las transformaciones introducen *redundancia*, siendo ineficientes para las aplicaciones como la compresión que necesitan representaciones *dispersas*, es decir representaciones que contienen una mínima cantidad de coeficientes no-nulos. Se puede, sin embargo, reducir la redundancia de las representaciones sobrecompletas eliminando una gran proporción de sus coeficientes por medio de algoritmos de *codificación dispersa* o *sparse coding* [147]. Considerando que los coeficientes de las transformaciones ya no son *linealmente independientes*, dichos algoritmos seleccionan una base reducida de coeficientes para representar con ella una aproximación de la señal. Los principales algoritmos son *Matching Pursuit* (MP) [118, 152] y *Basis Pursuit* (BP) [22]. MP es un algoritmo *glo-tón* (*greedy*) que elige una por una las ondículas que ofrecen la mejor correlación con la señal, substrayendo en cada iteración la ondícula elegida de la señal e iterando sobre la señal restante. BP resuelve el sistema lineal minimizando una función de penalización que consiste en la suma de las amplitudes de todos los coeficientes. Las desventajas de dichos algoritmos son sus altos costes computacionales, sus fallos en la búsqueda de la mejor aproximación de la señal, y la aparición de artefactos. Diferentes variantes de estos algoritmos han sido propuestas, y pueden encontrarse en el resumen propuesto por Pece [147].

Las ondículas sobrecompletas y dispersas son finalmente transformaciones que ya no poseen las propiedades de ortogonalidad ni de linealidad. La optimización de tales transformaciones resulta entonces una tarea particularmente difícil por los altos grados de libertad que ofrecen este nuevo tipo de representaciones y por la dificultad de manejar transformaciones no lineales. Asimismo es extremadamente difícil estudiar las propiedades matemáticas de transformaciones no-ortogonales y no-lineales.

En cambio un paralelismo directo puede establecerse con los modelos de visión biológica. Desde los años 1960 se propuso modelos de los sistemas sensoriales considerándolos como mecanismos optimizados en términos de reducción de redundancia de manera consistente con la *teoría de la información* [3, 7, 8, 54]. Específicamente, transformaciones espacio-frecuenciales similares a las transformaciones en ondículas se han utilizado para modelar la arquitectura funcional de las primeras etapas de los sistemas de visión biológica. En los primates, la información visual captada por la retina se procesa primero por las propias neuronas de la retina, enviándose a continuación al *núcleo lateral genículado* (LGN) antes de llegar al *córtex visual primario* (V1) [220]. Los trabajos de los premios Nobel Hubel y Wiesel a partir de los años 1960 [88, 87] han llevado a clasificar las neuronas del cortex visual primario en tres clases: *células simples*, *células complejas* y *células hipercomplejas*. El campo receptivo de las *células simples* ha sido descrito en primera aproximación como lineal y se modela habitualmente mediante funciones de Gabor [124, 36, 53]. Las células simples son consecuentemente sensibles a detalles visuales en una posición específica de la retina, así como una banda de frecuencia y una orientación específicas. La red de células simples se ha descrito como cubriendo todas las posiciones del campo visual, con una densidad superior en la *fovea*, al menos tres bandas de frecuencias en cada posición de la retina [40], y bandas de orientaciones de entre 20 y 40 grados [36, 87]. Dicha red tiene entonces importantes similitudes con las ondículas de Gabor. Por su parte las células complejas tienen la misma localización en posición, escala y orientación que las células simples pero son invariantes a la fase, y son sensibles a estímulos en movimiento. Las células hipercomplejas también llamadas *end-stopped* son o simples o complejas y son sensibles únicamente a estímulos limitados en extensión es decir que no salen de un lado u otro del campo receptivo de la célula [87]. Las células complejas y hipercomplejas son ambas altamente no lineales. También han sido descritas no-linealidades en las células simples. Destacan en particular interacciones entre células [94]. Dichas interacciones son de dos tipos. El primer tipo consiste en *inhibición* entre células vecinas¹. El segundo tipo de no-linealidad consiste en la *facilitación* de células también próximas aunque hasta una distancia más grande. La facilitación ocurre entre células que por su posición y orientación preferente se encuentran relativamente alineadas. Diversos modelos computacionales se han propuesto

¹Las células del área V1 tienen una organización *retinotópica* es decir que las células próximas en V1 tienen campos receptivos centrados en posiciones próximas de la retina. La inhibición de células vecinas corresponde entonces a una inhibición de señales procedente de posiciones cercanas.

para mostrar las características y habilidades de tales redes neuronales. Grossberg y colaboradores han desarrollado un modelo simulando células del LGN, células simples y complejas [131], mostrando la capacidad de tal red para realizar tareas de restauración de imagen y de eliminación de ruido. Heitger y colaboradores [83], Yen y Finkel [221] y Li [112] implementaron modelos similares mostrando la eficiencia de la arquitectura del córtex visual V1 para la extracción de bordes y contornos. También se ha especulado que V1 lleve a cabo tareas similares a los algoritmos de codificación dispersa con el fin de obtener una representación eficiente de la información visual [139]. Por ello se han propuestos modelos basados en redes neuronales implementando MP [151]. Otra herramienta para optimizar las transformaciones sobrecompletas consiste entonces en utilizar los conocimientos que se tienen del sistema visual primario. Es interesante inspirarse del conocimiento de V1, ya que las prestaciones y capacidades del sistema visual humano están todavía muy por encima de los sistemas de visión artificial, tanto por la capacidad de análisis de escenas complejas, el reconocimiento de formas y patrones, el tratamiento en tiempo real, etc. De hecho se puede considerar el sistema visual humano como optimizado para el tratamiento eficiente de la información visual por los millones de años de evolución.

El estudio de las estadísticas de las imágenes naturales es una tercera herramienta para ayudar en el diseño de las ondículas no-ortogonales. Importantes resultados han sido obtenidos por técnicas de Análisis de Componentes Independientes (ICA). Las componentes independientes de las imágenes naturales extraídas por tales métodos se asemejan a los modelos de los campos receptivos de las células simples y a las ondículas: en la mayoría de los casos dan lugar a funciones orientadas de frecuencia particular [141, 43]. Las funciones de Gabor serían entonces los elementos básicos de construcción de las imágenes naturales. Las estadísticas de las imágenes naturales permiten también encontrar reglas de agrupamiento de los bordes en contornos continuos [69] congruentes con los estudios psicofísicos de percepción de los contornos [55, 84, 99]. Existe además una relación directa entre la visión biológica y las estadísticas de las imágenes naturales: en el ámbito de las Redes Neuronales Artificiales se suele considerar que las redes biológicas aprenden características del mundo real. Así los campos receptivos de las células del córtex V1 pueden ser interpretados como procedentes de la extracción de componentes independientes aprendidas de las imágenes naturales.

Dentro de las posibles aplicaciones en tratamiento de imágenes, las ondículas se han mostrado particularmente eficientes para la *eliminación de ruido*. En efecto las ondículas caracterizan particularmente bien los cambios de contraste que corresponden a bordes o líneas de las imágenes naturales. Producen coeficientes de amplitud elevada en los bordes mientras que el ruido induce un nivel relativamente uniforme y bajo en todos los coeficientes. La utilización de métodos de umbralización permite entonces eliminar la mayoría del ruido conservando bien los bordes y líneas [47]. Posteriormente se propu-

so mejoras por la utilización de ondículas sobrecompletas invariantes a la traslación [28] así como mediante el uso de un número mayor de orientaciones [189, 160, 41]. Otra aplicación, la *extracción de bordes* ha sido relacionada con los modelos de visión biológica desde los primeros trabajos de Marr [125]. Posteriormente Canny [16] desarrolló una formalización y una justificación puramente matemáticas de los métodos de extracción de bordes. Propone un método de construcción de filtros 2D antisimétricos y orientados para la detección de los bordes a exclusión de las líneas (solo se consideran cambios antisimétricos de contraste). El tamaño de los filtros se adapta en función de la señal y de la cantidad de ruido. En este tipo de esquemas se utilizan habitualmente entre 6 y 8 orientaciones y los filtros tienen formas generalmente elongadas. Un método de histéresis permite la selección de grupos enteros de bordes evitando así que un contorno sea extraído sólo por partes. Más adelante se ha demostrado que el uso de filtros (complejos) de Gabor permite la extracción simultánea de los bordes (elementos antisimétricos) y las líneas (elementos simétricos). El proceso de extracción se asimila entonces estrechamente a los modelos funcionales del cortex visual V1 [133, 175, 221, 112, 131, 102] descritos anteriormente. Las respuestas de las células hipercomplejas o *end-stopped* se han mostrado también importantes para completar la extracción de bordes [83]. En *compresión* de imagen las ondículas biortogonales han dado lugar al estándar JPEG-2000 [184], el cual constituye la referencia en materia de compresión. Consiste en una transformación en ondículas biortogonales seguida por una quantization de los coeficientes y una codificación por planos de bits. Los bordes de contraste de las imágenes producen en el dominio ondicular un reducido número de coeficientes de gran amplitud. La cuantificación preserva dichos bordes mientras elimina los coeficientes pequeños correspondientes a detalles poco perceptibles de las imágenes. En consecuencia las imágenes se pueden codificar mediante un número reducido de coeficientes no-nulos del dominio ondicular. Aparte las ondículas han sido utilizadas con éxito en toda una serie de aplicaciones adicionales, donde las ondículas sobrecompletas a menudo muestran un mejor comportamiento que las (bi-) ortogonales, como p.e. el análisis de texturas [159] o la fusión de imágenes [215] o el reconocimiento de patrones [164].

0.2 Ondículas log-Gabor auto-invertibles

El primer objetivo de la tesis es el desarrollar una multiresolución inversible que a la vez posee excelentes propiedades matemáticas y sea al mismo tiempo próxima a los conocimientos disponibles sobre la fisiología de los campos receptivos de las células simples. La transformación propuesta implementa filtros de tipo Gabor con anchos de banda y distribución en orientaciones y escalas próximas de las células simples.

Tres problemas surgen desde el punto de vista técnico: primero las funciones de Gabor no son de media cero, por lo que no serían adecuadas en principio para la multiresolución (un componente continuo contamina todas las bandas de frecuencias). Este problema se resuelve utilizando filtros log-Gabor que son Gabores en coordenadas log-polar [53]. El uso de tales filtros mejora también el recubrimiento del espectro de Fourier.

Segundo, se considera usualmente que no se puede obtener un recubrimiento uniforme del espacio de Fourier mediante filtros de tipo Gabor. Mostramos que por el desplazamiento en orientación de las escalas pares (propuesto en [74]) y con el ajuste del ancho de banda se consigue un recubrimiento casi uniforme (con variaciones menores a 0.6%), lo que permitirá una reconstrucción exacta después de una muy leve modificación de los filtros.

El tercer problema consiste en construir filtros orientados de alta frecuencia que estén estrechamente localizados espacialmente y frecuencialmente. Los filtros siendo selectivos a frecuencias próximas de la frecuencia de Nyquist, no es posible construirlos directamente a través de la ecuación clásica de los log-Gabores.

Finalmente se consigue la construcción de una multiresolución con filtros de log-Gabor que posee la propiedad de auto-inversibilidad, es decir que la transformación inversa es de reconstrucción exacta y utiliza los mismos filtros que la transformación directa. Tal propiedad aporta una estabilidad adicional frente a la aparición de artefactos.

Las ondículas log Gabor así definidas son particularmente adecuadas para codificar los cambios de contraste local de las imágenes naturales. De hecho muestran una gran capacidad para la segregación y la eliminación del ruido en imágenes naturales por simple umbralización de los coeficientes de la transformación en ondículas. Comparado con otras multiresoluciones, como las ondículas bi-ortogonales [44], ondículas invariantes a traslación [28], y steerable pyramids [182, 160], las ondículas log-Gabor aquí presentadas muestran una mejora importante en la capacidad de eliminación de ruido por umbralización.

0.3 Extracción de bordes por modelos energéticos

El capítulo 3 de la tesis propone un método de extracción de bordes basado en modelos previos inspirados en el córtex visual V1. Se ha demostrado que la extracción simul-

tánea de los *bordes* (cambios antisimétricos de contraste o escalones) y de las *líneas* (cambios simétricos de contraste, o crestas) no se pueden conseguir por medio de filtros lineales [150]. Consecuentemente se han propuesto *modelos energéticos* (es decir cuadráticos) basados en la suma cuadrada de dos filtros de tipo Gabor en cuadratura de fase [1, 133, 175, 149, 75, 221, 83, 131]. Cada filtro detecta o los cambios simétricos o los antisimétricos, mientras la suma cuadrada garantiza una respuesta monomodal.

El modelo energético está directamente inspirado en la células complejas de V1 que muestran una respuesta similar a las células simples (es decir una respuesta a estímulos localizados y específicos de una orientación y una banda de frecuencia), pero que mantienen una invariancia de la fase gracias a la suma cuadrada de dos filtros en cuadratura de fase [17].

Llamaremos *canal* al resultado de la convolución de la imagen original mediante un filtro de Gabor de una orientación y escala determinada. Después de la suma cuadrada, la extracción de borde y líneas implementa una operación local de *supresión* de los coeficientes no-máximos. Es una operación clásica que se encuentra en muchos métodos de extracción de bordes (p.e. en Canny [16]). La supresión consiste en la eliminación de los coeficientes que no son máximos locales en la dirección normal a la orientación del canal (o del filtro correspondiente). Este proceso se llama también ocasionalmente *competición espacial* [75]. Además se eliminan también los coeficientes que no son máximos locales entre orientaciones (*competición en orientación*). Mediante dichos procesos de supresión, se eliminan la gran mayoría de los coeficientes y los coeficientes restantes suelen formar cadenas continuas localizadas en los contornos de los objetos de la imagen, aunque aparecen también coeficientes aislados debido al ruido o detalles menos perceptibles de la imagen. Finalmente para refinar la extracción, se suele aplicar un paso de *facilitación* (o *cooperación*) entre los coeficientes alineados o co-circulares: criterios de proximidad, de co-linealidad, de co-circularidad están aplicados para preservar los coeficientes pertenecientes a contornos y eliminar los coeficientes aislados debidos al ruido.

Comparado con las técnicas lineales, p.e. el detector de bordes de Canny [16], el esquema es además más robusto a las detecciones dobles de bordes, detecta cambios de contraste de cualquier fase (no solo múltiplos de 90°). Una originalidad de la implementación propuesta en la presente tesis es la utilización de un modelo basado en ondículas para la extracción de borde. La utilización de múltiples escalas implica que los bordes se extraen separadamente para cada escala. Además el uso de la codificación dispersa (ver capítulo 4) permitirá la reconstrucción de las imágenes a partir de aquellos bordes.

0.4 Representaciones dispersas basadas en extracción de bordes

Las transformaciones sobrecompletas lineales inducen una importante redundancia. En aplicaciones como la compresión donde se pretende reducir la cantidad de información, es necesario implementar algoritmos de *codificación dispersa* o *sparse coding* [147, 118, 141, 22]. En casos particulares, p.e. cuando se busca un sub-base ortogonal dentro de una base de decomposición sobrecompleta, existen algoritmos rápidos y óptimos (i.e. que encuentran la decomposición de reconstrucción exacta que contenga la mínima cantidad de información) [29]. Pero en el caso general y particularmente cuando no se puede construir una sub-base ortogonal (como es el caso aquí con los filtros de tipo Gabor), solo existen algoritmos iterativos que encuentran aproximaciones a la solución óptima. Proponemos en el capítulo 4 de la presente tesis un algoritmo dedicado a las ondículas log-Gabor, el cual es de coste computacional bajo y está basado en los modelos biológicos de V1, en particular en la extracción de bordes por modelos energéticos (ver capítulo 3).

Proponemos la hipótesis que los coeficientes situados en los bordes de las diferentes escalas son suficientes para reconstruir la mayor parte de la imagen. Dicho de otro modo, se considera que la amplitud de los coeficientes alejados del lugar preciso donde está situado el borde puede ser eliminada a condición de realzar para compensar los coeficientes de bordes. Este aumento de amplitud de los coeficientes seleccionados y la reducción de la amplitud de los no-seleccionados se puede realizar mediante la proyección de los coeficientes no-seleccionados sobre los seleccionados. Por este método no se impera la reconstrucción, es decir que se obtiene todavía una reconstrucción exacta mediante la transformación en ondículas log-Gabor inversa.

El algoritmo propuesto procede entonces en tres pasos: primero se calcula la transformación en ondículas log-Gabor. Segundo se elige una selección de los coeficientes de bordes por el método energético descrito en el capítulo 3. Y finalmente se proyectan los coeficientes no seleccionados (llamados el *residuo*) sobre los seleccionados.

Como resultado se obtiene por este método una buena reconstrucción de los contornos aunque algunas distorsiones pueden aparecer en los diferentes tipos de esquinas y uniones así como en las texturas. Otros algoritmos dedicados a la codificación dispersa de uniones y texturas serían útiles para mejorar las prestaciones del método. Se podrían basar en métodos existentes de extracción de esquinas y uniones [83], de representación estadística de las texturas [159] y de codificación dispersa separando los bordes de las texturas [45]. Otra manera de mejorar el resultado consiste en iterar el método propues-

to. Mostramos que unas tres iteraciones son suficientes para obtener una importante mejora en uniones y texturas.

La transformación log-Gabor dispersa así definida se revela adecuada para una diversidad de aplicaciones. Extrae los bordes y reconstruye a partir de los mismos. Utilizada en compresión de imágenes limita la aparición de los artefactos usualmente molestos cuando se emplean transformaciones (bi-) ortogonales. El método realiza además simultáneamente una tarea de eliminación de ruido: el ruido por sus estadísticas produce un nivel bajo de amplitud repartido en todo el dominio de transformación y además no induce estructuras coherentes en los coeficientes, como por ejemplo cadenas. Consecuentemente, interfiere poco en la selección de los coeficientes de bordes. Los contornos serán entonces reconstruidos correctamente mientras que el ruido desaparece. Comparado con los métodos más comunes de eliminación de ruido con ondículas, el presente método basado en ondículas log-Gabor dispersas preserva particularmente bien los bordes y las líneas de poco contraste mientras que en otros métodos pueden totalmente desaparecer.

El presente método de codificación dispersa aporta una ventaja adicional para las aplicaciones de compresión: los coeficientes de bordes están localizados en cadenas por lo que sus posiciones y amplitudes pueden ser codificadas con alto grado de predictibilidad a lo largo de dichas cadenas. Aquello permite reducir la cantidad de información por codificar sin pérdida de calidad: provee entonces una decorrelación adicional de la señal. Gracias a la representación dispersa y a la codificación por cadenas, el esquema log-Gabor wavelets podría revelarse competitivo con los estándares de compresión como JPEG y JPEG-2000.

0.5 Cierre de contornos por tensor voting

El capítulo 5 de la tesis está dedicado al agrupamiento de bordes en contornos continuos. En este capítulo los bordes están representados por tensores que codifican tanto la orientación del borde, la amplitud del borde y la confianza de tener una unión. Un mecanismo de *votación* permite propagar la orientación de un borde hacia sus vecinos y encontrar así la orientación local del contorno. En presencia de hueco entre segmentos de contorno, los contornos se completarán a lo largo de la dirección preferida.

Proponemos aquí dos mejoras al método de *tensor voting*. Primero estudiamos las ventajas de implementar una versión iterativa de tensor voting. Mostramos que las iteraciones permiten refinar las estimaciones de orientación local y la confianza de tener uniones, sin por ello implicar desventajas importantes. Es entonces conveniente utilizar

un cierto número de iteraciones para el tensor voting.

El segundo estudio propuesto incorpora un método de evaluación de la curvatura local y realiza las operaciones de voto teniendo en cuenta dicha curvatura. El uso de este parámetro de curvatura no implica un coste computacional elevado y mejora de modo significativo el cierre de contornos. Los contornos completados muestran curvaturas más suaves y perceptualmente más agradables. La cantidad de artefactos se ve reducida de un modo importante y el método es menos vulnerable al desajuste del parámetro de escala empleado en el método clásico de tensor voting.

0.6 Principales aportaciones de la presente tesis

Las principales aportaciones originales de la presente tesis pueden resumirse por cada uno de los capítulos de la siguiente manera.

El Capítulo 2 presenta una multiresolución en ondículas de Gabor original en varios aspectos: primero incluye nuevos filtros de alta frecuencia los cuales son orientados, estrechamente localizados y con valores complejos. Segundo, es auto-inversible, lo que induce una mejora importante puesto que ninguna multiresolución por filtros de Gabor de reconstrucción exacta había sido propuesta anteriormente en la literatura. La metodología de optimización de la transformación en ondículas resulta también novedosa. Reúne criterios múltiples que incluyen (1) datos de inspiración biológica referentes a los campos receptivos de la células del cortex visual V1, (2) estadísticas de las imágenes naturales, (3) preservación de la calidad perceptual y (4) criterios de optimalidad matemática. Como resultado dicha transformación muestra habilidades inéditas para la separación estadística entre el ruido espacialmente incoherente y los elementos de las imágenes naturales, lo cual es prometedor no solo para aplicaciones de eliminación de ruido sino también para tareas más generales de análisis de imagen.

La innovación del capítulo 3 reside sobretudo en la construcción de un método energético de extracción de bordes multiescala como síntesis de diferentes métodos, es decir agrupando diversos mecanismos anteriormente propuestos en la literatura. Otras importantes innovaciones residen (1) en la utilización de un esquema ondículas auto-inversible para la extracción de los bordes y (2) en la posibilidad de reconstruir a partir de los bordes. Desde el punto de vista de la calidad de los resultados, los bordes extraídos muestran una similitud importante con los bordes percibidos en las imágenes naturales. Comparado con el método clásico de Canny [16], el presente método es capaz de extraer tanto los bordes simétricos como antisimétricos mientras Canny solo extrae

bordes antisimétricos.

El capítulo 4 introduce una innovación consecuente en los algoritmos de codificación dispersa por la proposición de basar la extracción de la sub-base de codificación sobre métodos de extracción de bordes. El algoritmo se asemeja en algunos aspectos con el Matching Pursuit (MP), pero innova autorizando la selección simultánea de múltiples coeficientes en la misma iteración. Además, mientras que los algoritmos generales de codificación dispersa son iterativos y computacionalmente costosos, el presente algoritmo puede implementarse de manera no-iterativa con coste computacional bajo. El método propuesto permite la compresión de imágenes con una calidad aproximándose a los estándares de compresión (JPEG y JPEG-2000) y con la importante ventaja de limitar drásticamente la presencia de artefactos de alta frecuencia y de preservar la calidad perceptual. La codificación dispersa beneficia además de la robustez al ruido de ambas etapas de transformación en ondículas de Gabor y de extracción de bordes. Es entonces capaz de operar la compresión bajo niveles de ruido muy importante, mientras otros algoritmos de codificación dispersa como MP muestran respuestas inestables en presencia de ruido. Finalmente la codificación dispersa puede utilizarse en tareas de denoising donde muestra una habilidad particular para preservar los bordes.

El capítulo 5 presenta dos mejoras originales para el método de cierre de contornos por tensor voting. Introduce una versión iterativa que se muestra superior al método original particularmente en casos presentando ambigüedades o dificultades tales como desajustes entre niveles de detalles dentro de la misma imagen. La segunda novedad propone el cálculo y el uso de una representación de la curvatura para mejorar la integración de los contornos y la buena continuación de los mismos y para reducir la aparición de artefactos.

La presente tesis incorpora entonces novedades importantes en cada uno de los temas abordados, los cuales forman además un conjunto coherente llegando hasta la constitución de nuevas transformaciones en ondículas sobrecompletas. Estas se muestran eficientes para el análisis y tratamiento de las imágenes naturales.

0.7 Conclusiones

La presente tesis introduce una nueva multiresolución llamada ondículas log-Gabor, inspirada en la fisiología de la corteza visual primaria. La transformación es auto-inversible, y particularmente adaptada para tareas de tratamiento de imágenes como la eliminación de ruido. La transformación a pesar de ser sobrecompleta, incorpora un método

de reducción de la redundancia que autoriza el uso de las ondículas log-Gabor incluso para la compresión de imágenes. Dichas ondículas log-Gabor poco redundantes representan las imágenes a partir de sus bordes multiescala. La última parte de la tesis está dedicada al método de tensor voting que permite completar bordes discontinuos para formar contornos continuos.

La presente tesis deja abierta una gran número de líneas de investigación. Primero para la aplicación y la optimización de los métodos propuestos en tareas de tratamiento de imágenes como la eliminación de ruido, la fusión de imágenes, el análisis de texturas, la extracción de bordes, la compresión, o el reconocimiento de objetos. Segundo en la mejora y el desarrollo de los algoritmos propuestos, por ejemplo la optimización de la transformación lineal, la mejora de la extracción de contornos por la inclusión de detección de uniones. Para la compresión de imágenes, se pueden conseguir importantes mejoras mediante el desarrollo de la codificación predictiva de contornos.

Chapter 1

Introduction

"It's as if the genetic instructions determine a fixed network of principles and a certain few switches that aren't set yet, and the child has to set the switches on the basis of simple data. Once the switches are set, the whole system operates."

Noam Chomsky, in *Evolution of revolution: Chomsky's minimalism*, D. Pacitti interviews N. Chomsky, JUST Response on May 20 2002.

1.1 Wavelet representations

Since twenty years, *wavelets* got increasing importance in signal and image processing. They were introduced for the first time in 1982 by Morlet and coworkers [132], and were further developed in their theoretical and mathematical aspects by Meyer [130], Mallat [117] y Daubechies [33]. Wavelets propose an alternative to the spatial (or temporal) representations, and to the frequencial (Fourier) ones, as they offer a characterization of signals in a space-frequency domain. The present thesis is dedicated to 2D still images, so that the domain of interest is the 2D spatial domain and the wavelet domain is a 4D space-frequency domain. Nevertheless wavelets are also very often used in a temporal-frequency domain. A wavelet transform consists in convolving a signal by translations and dilatations of a *mother wavelet*, which is generally closely localized in space and specific of a frequency band. The space-frequency representations provided by such transforms have been shown to perform efficiently a variety of signal processing tasks. *Dyadic* structures consist in limiting the set of possible dilatations of the mother wavelet to powers of two. Dyadic structures with the use of a critical downsampling allow the construction of (bi-)orthogonal wavelet families [117].

Orthogonal wavelets implement an orthogonal basis of decomposition functions which is used for both analysis (i.e. decomposition) and synthesis (i.e. reconstruction) of the signal. Orthogonality imposes drastic conditions on the shape of the mother wavelet so that *Biorthogonal wavelets* incorporating different basis for the direct and inverse transforms are often preferred. In digital signal processing, those wavelets constitutes today a reference. Nevertheless (bi-)orthogonal wavelets suffer important drawbacks. As a first point they are generally not *translation invariant* which means that a small displacement of the signal in the spatial domain could induce a large variation -out of proportionality- of the wavelet coefficients. This problem is related with the presence of *aliasing*: the wavelet transforms are organized in frequency bands which are not totally separated and intermix together because of the critical downsampling. The aliasing is compensated in the inverse transform and does not prevent the exact reconstruction of the signal. Nevertheless in applications where some coefficients are eliminated, e.g. compression or denoising, aliasing could give rise to artifacts. Moreover the use of different filter sets for direct and inverse transform in the biorthogonal scheme weaken the reconstruction robustness, since some features can appear which do not strictly correspond to the analyzed signal. Additionally in 2D image processing (bi-) orthogonal wavelets are implemented as X and Y separable 1D wavelets, and therefore show a worst behaviour in oblique directions.

1.2 Overcomplete representations

In the last fifteen years a diversity of alternative multiresolution transforms have been proposed to overcome the previously described (bi-) orthogonal wavelet drawbacks at the cost of loosing the properties of (bi-) orthogonality. *Translation invariant wavelets* (also called *undecimated wavelets*) are constituted by the same wavelet functions as the (bi-) orthogonal wavelets but they are not downsampled, thus they do not introduce aliasing and are invariant to translation. They turn out to be more efficient than biorthogonal wavelets particularly in denoising applications [28, 111, 20]. *Steerable pyramids* [182, 159, 160] present the property of translation invariance thanks to a less critical downsampling which avoids the introduction of aliasing. Among other interesting properties steerable pyramids offer an improved behavior in the oblique directions thanks to the implementation of a increased number of orientations. *Complex-valued wavelets* describe signals through both symmetric and antisymmetric functions that are in quadrature of phase. They were successfully employed e.g. for texture analysis [74, 96, 123, 159]. Among other transforms, *curvelets* [46, 189] or *contourlets* [41] were proposed principally for improving the resolution in orientation and for a better matching with edges of the natural images. Other overcomplete multiscale transforms

were developed in parallel or even previously to the wavelets as e.g. the Laplacian pyramids [15] or the cortex transform [216].

On the other hand, *Gabor wavelets* [74, 134, 173] use as mother wavelet a function originally proposed by Gabor in 1946 [68] and used by Morlet and coworkers in 1982 to define the first wavelet transform [132]. A Gabor filter in the spatial domain is a Gaussian multiplied by a complex exponential. Its shape is Gaussian both in the spatial and frequency domain, so that the filter is closely localized both in space and in frequency. Indeed, it can be demonstrated that the filter reaches the theoretically optimal of joint localization in space and frequency [68]. Consequently it has been considered as an optimal filter for signal processing, although its use was always restricted by important drawbacks inherent to Gabor functions, which can be summarized as follows:

1. It is not possible to construct a complete orthogonal basis of Gabor functions.
2. The Gabor filters do not have zero-mean, which induces a DC component in the coefficient of any frequency band. This last trouble can nevertheless be solved by employing log-Gabor functions which are constructed as Gabor functions but on the logarithmic frequency coordinate system [53].
3. It is particularly difficult to cover the highest and lowest frequencies by means of the log-Gabor filters, to achieve a uniform coverage of the frequency domain and to reconstruct exactly the signal (Gabor wavelets are not invertible).

For those reasons, Gabor wavelets have been used mainly in applications which do not require image reconstruction such as texture analysis [27, 158, 74, 173], edge extraction [175, 102], etc, although their good qualities also allowed their use for image restoration and denoising [101].

1.3 Sparse representations

The use of non-critically sampled transforms allows then to overcome some drawbacks of the (bi-) orthogonal wavelets. Those transforms are called *overcomplete* to indicate that they provide a representation whose dimension is superior to the one of the signal: there are more transform coefficients than pixels in the original image. Consequently those transforms introduce *redundancy*. Therefore they are inefficient for applications such as image compression which needs *sparseness*, i.e. a large proportion of zeros within the coefficients. Moreover the transform functions are no more linearly independent and then do not constitute a *basis*, they are referred to as a *dictionary* [118].

Fortunately it is possible to reduce the redundancy of an overcomplete representation through the elimination of most of its coefficients by means of non-linear *sparse coding* algorithms (also called *sparse approximation*, see reviews in [147, 202, 22, 225]). Sparse approximation algorithms take advantage of the linear dependency within the dictionary and exploit the consequent non-uniqueness of the decomposition to represent the signal through a reduced sub-dictionary that best match with it, building then a sparse approximation of the signal. The most popular algorithms are *Matching Pursuit* (MP) [118, 167, 152, 148] and *Basis Pursuit* (BP) [22, 45]. MP is an iterative greedy algorithm which chooses at each step the decomposition function that offer the highest correlation with the signal. It iterates on the residual signal (i.e. the signal minus the chosen decomposition function). BP solves the linear system minimizing a penalizing function corresponding to the sum of the amplitude of all coefficients. Among other alternatives, Olshausen and Field propose a gradient descent technique on a penalizing function based on the RMSE (Root mean square error) and a sparseness condition [139]. Using such algorithms it has been shown that even very redundant dictionaries can be used to achieve an efficient representation in terms of the quantity of information. For example Peotta et al. defined a dictionary of curved Gabor-like functions with a redundancy factor up to 17000 (i.e. there is 17000 times more dictionary functions than image pixels) [148]. Using MP and a wavelet compression on the residual the scheme was shown competitive with JPEG-2000 particularly at high compression rates.

For some particular overcomplete set of decomposition functions (called dictionary), optimally sparse representations can be obtained through dedicated algorithms [29, 107]. For *quasi-incoherent dictionaries*, i.e. dictionaries where the scalar product between decomposition functions is small (in other words the dictionary is "almost orthogonal"), it is possible to define mathematical properties on the optimality and the convergence [72, 202]. Theoretical results can also be derived in the case of sufficiently sparse signals [116]. Nevertheless in the present thesis we will be situated in such particular cases. In the general case and specially when the dictionary has an important overcomplete factor, when working on natural and noisy images (i.e. non-sparse images), and when it is not searched for orthogonal sub-dictionary of decomposition functions, only approximations to the optimal solutions can be found. Moreover the existing algorithms are generally computationally costly, while artifacts are used to appear. Hence, in the current state of the art, the sparse overcomplete representations do not establish yet that they clearly outperform the (bi-) orthogonal wavelet schemes in terms of compression abilities.

Sparse overcomplete wavelets finally do not possess neither the property of orthogonality nor the one of linearity. The optimization of such transforms appears then a particularly difficult and appealing task for the high degrees of freedom that this kind of representation can offer and for the difficulty to manage non-linear transforms. More-

over, it seems particularly difficult to establish the mathematical properties of those non-orthogonal and non-linear transforms.

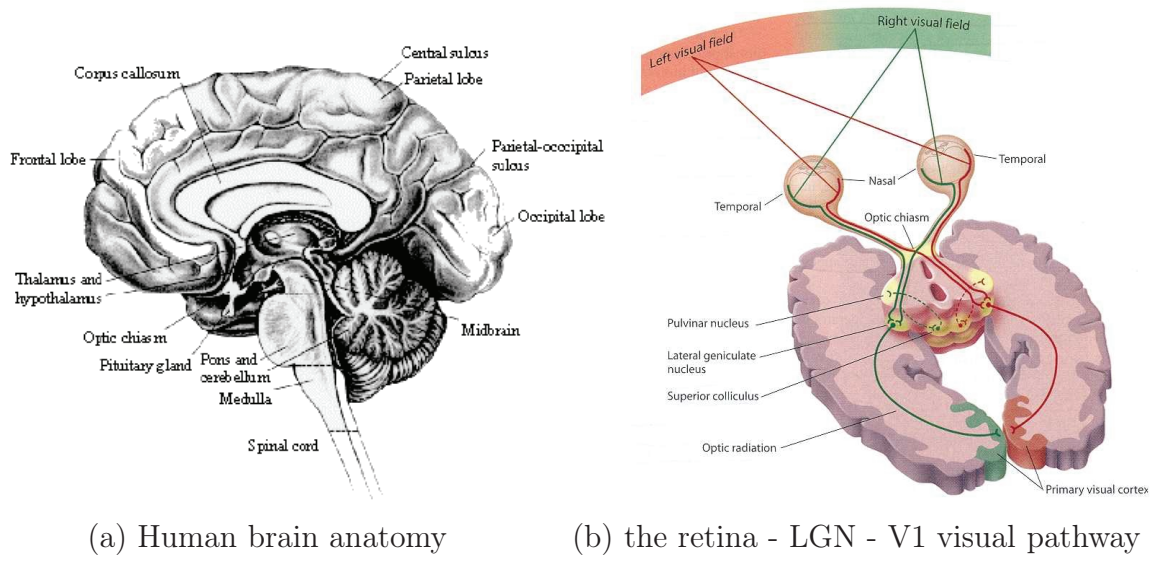
1.4 Visual system physiology

A direct parallelism can be established between image processing multiresolution transforms and biological vision. Space-frequency transformations similar to the wavelets ones have been proposed to model the functional architecture of the first stages of biological vision systems [216], so that the transformation operated by the primary visual cortex is now often considered as a wavelet transform, in the opinion of the wavelet's original developers themselves [117]. The study of biological vision and the image processing developments are then getting closer, so that the knowledge on physiology and modelling of the visual cortex, or psychophysics could, in our opinion, reveal to be useful for computer vision.

In primates, the visual information acquired by the retina is processed through the retina neurons before to be send to the *Lateral Geniculate Nucleus* (LGN) and later reach the *primary visual cortex* (V1) [220, 10, 85] (see in particular [17, 19] for reviews on computational models). Ganglion cells in the retina output and LGN neurons are often modelled as linear spatial bandpass filters, although some non-linearities are also described as e.g. luminance and contrast adaptation. Their response also depends on the temporal frequency of the stimulus [17].

The work initiated by Hubel and Wiesel [88, 87, 172] since 1960 conduced to classify V1 neurons in three classes: *simple cells*, *complex cells* and *hypercomplex cells*. The receptive field of simple cells has been described in a first approximation as a linear transform of the retina input. Each simple cell is selective to visual features closely localized around a particular position of the retina, and having amplitude in a specific frequency band, orientation and phase. Such receptive field of simple cells is usually modeled by Gabor-like functions [124, 35, 36, 53]. For the same retina position, a variety of simple cells appear arranged across various frequency bands [40] and orientations [36, 87] and in quadrature of phase [157]. Thus, such arrangement show important similarity with the complex-valued Gabor wavelets [216]. Non-linearities have also been described in simple cells [17, 177]. They consist first in a contrast normalization. Other important non-linearities consist in lateral interactions between neighboring cells [12, 13, 94, 2, 191, 163]. Those interactions are of two types. The first one consists in an *inhibition* [97] between neighboring cells¹. The second type of non-linearity consists in a *facilitation* of

¹V1 cells are arranged in a *retinotopic* organization which implies that close V1 cells have receptive



(a) Human brain anatomy

(b) the retina - LGN - V1 visual pathway

Figure 1.1: Brain anatomy and the visual pathway. **a.** Crude anatomy of the human brain. **a.** The visual information from the left and right visual fields is first processed in the *retina* before to be send to the *Lateral Geniculate Nucleus*. Afterward it reaches the *primary visual cortex* (from [71]).

cell responses also in a certain proximity although toward a larger distance [194]. The facilitation occurs between cells which for their position and preferred orientation are relatively aligned and co-circular. Non-linearities are also described in terms of non-classical receptive fields [73], center-surround interactions [217, 18, 178], masking [206] or contextual influences [2]. They are not specific to V1 simple cells but seem to appear very frequently in all the neural circuitry.

By its part, complex cells [127, 17] are similarly localized in position, frequency and orientation to simple cells (although slightly coarser in their position selectivity) but they are invariant to phase and are sensible to motion. Hypercomplex cells, also called *end-stopped*, are either simple or complex and they are sensible solely to stimuli limited in width, i.e. stimuli that are not prolonged in one side or the other of the cell receptive field [87]. End-stopped cells appear to be determinant for occluded contour detection [83] and for motion disambiguation [128]. Both complex and hypercomplex cell receptive field are highly non-linear.

In parallel to physiological studies, other types of works are set up in non-invasive ways through psychophysical experiments, e.g. on the receptive fields [156, 52], on lateral interactions and contour perception [55, 84, 122, 138, 194], on masking ef-

fields which are sensitive to close position in the retina. Therefore the inhibition between neighboring cells implies an inhibition of signals originating from neighboring positions of the visual field.

fects [110, 192] or particular features with perceptual significance like curvature extremum [4, 138].

1.5 Biological modeling

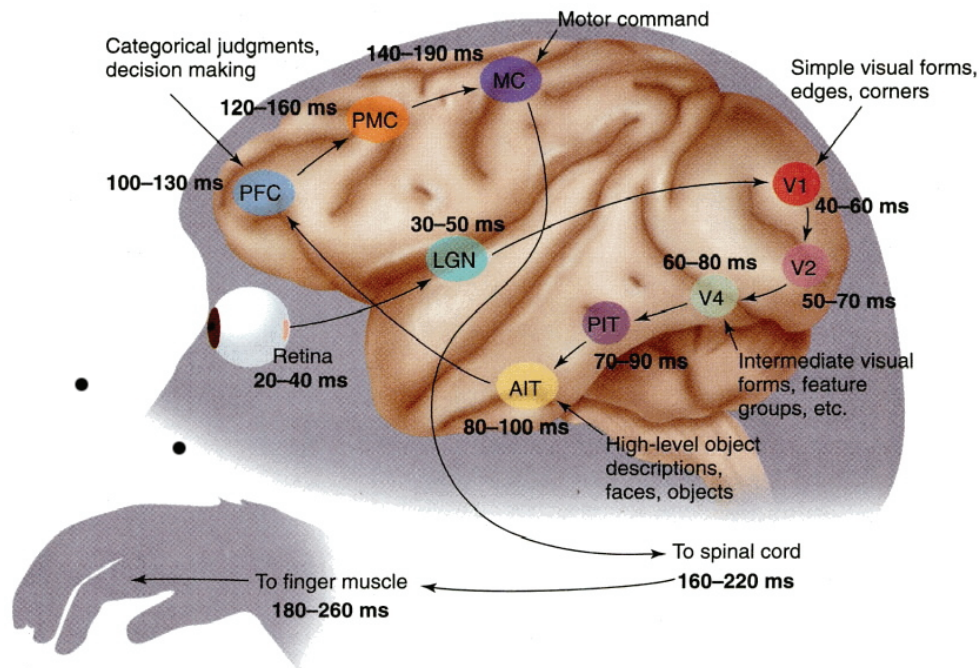


Figure 1.2: A visuo-motor pathway hypothesis including functionality and timing (from Thorpe and Fabre-Thorpe [199]).

Different computational models have been proposed in the literature for explaining the characteristics and abilities of the biological vision architecture. General models incorporate Information Theory concepts for explaining mechanisms of sensory coding [3, 7, 8, 54]. In parallel, dedicated models specifically designed for V1 were proposed. The first proposed schemes model the simple cells receptive field as a linear Gabor filtering [124, 34, 91] or model the multiresolution aspects [108, 216, 37]. Later V1 models states that V1 could perform tasks of the edge extraction and contour integration [73, 79, 81, 83, 99, 112, 131, 133, 175, 201, 208, 221] based on the lateral interactions of local inhibitions and long-range facilitation. Last versions of such models also show how contour integration rules could be learned [76, 23]. Some other models demonstrate the noise robustness and contrast enhancement abilities of V1 architecture [75, 131], or the ability to segregate textures [73, 154]. Finally during last ten years, different models have been proposed for showing how V1 could build a sparse representation of the visual information from an overcomplete set of visual neu-

rons [139, 210, 140, 151]. Energy efficiency criteria were also recently discussed [209]. Reviews of computational models on the early visual system (including contour grouping across occlusions, binocular vision and 3D perception, figure-ground segregation, color or motion perception) can be found in [211, 109, 114, 17]. Further in V1, two interconnected pathways deal with motion perception [128, 9, 142] (through areas MT and MST); and object recognition (through successive visual areas V2, V4 and IT which would construct increasingly integrated shape representations). V4 could code for angles, curves and basic shapes [145, 144]. IT (i.e. *Infero-Temporal cortex*) was shown to deal with object and face recognition [95, 105, 193, 174].

1.6 Statistics of natural images

Statistics of natural images are another available tool for helping us in the design of the overcomplete transforms since multiresolution approaches have to be set up according to the type of data to analyze. Additionally, the features of natural images are directly linked to vision systems because they supposedly have the ability to adapt and to learn the features of the natural environment. This adaptation would occur through separated mechanisms both through the phylogenesis (evolution of the specie) and the ontogenesis (development of the individual).

Studies of the statistics of individual pixels justify e.g. the *gamma correction* of the vision or the *whitening of the histogram*. Both consist in reducing the contrast sensibility in the light colors and increasing it in the dark luminance values [43]. The contrast is nevertheless relatively independent to the luminance, what could explain the separated gain controls for contrast and luminance observed in the visual system [66].

Statistics on Fourier coefficients of natural images show the Fourier amplitude usually decreases inversely with the spatial frequency [53]. Those statistics explain why the visual system is less sensitive to low frequencies than to medium ones. The sensibility is also reduced in highest frequencies where the noise level is usually higher and the features are consequently less significant [3].

Last ten years several methods have been developed for extracting the elementary components of natural images. Such methods are based on sparse coding methods [141] or on *Independent Component Analysis* (ICA) [11, 43, 181]. The independent components are oriented elongated 2D functions and show important resemblance with the receptive fields of V1 simple cells [207, 171]. Similarly, it is also possible to learn the retinotopic topography of cell organization in V1 [89]. Non-classical or non-linear re-

ceptive field properties can also be derived from natural image statistics [223, 121]. Furthermore different classes of probabilistic and computational models of the primary visual system can be derived from statistics of natural images and probabilistic considerations [166, 213], as e.g. the divisive normalization which could model the lateral interactions between cells and simultaneously explain how V1 could decorrelate the visual information.

Statistics of natural images also permit to establish the grouping rules for edges in continuous contours [69, 104, 30, 24]. Those natural image based grouping rules are consistent with the psychophysical studies on contour perception [55, 84] and with the Gestalt psychology [99].

Statistics of wavelets coefficients are also important in many applications. They are related with the type of wavelet transforms but also with the kind of signal to analyze. As an important feature, a high *kurtosis*, i.e. a heavy tail of the histogram of the wavelet coefficients, is important for an efficient and sparse representation of signals [46]. Relations between neighboring wavelet coefficients which are statistically linked can be exploited for applications such as image denoising [160], image compression [14] and even more importantly for texture analysis [159].

1.7 Viewpoints in cognitive sciences

In background of the neural system modeling, there exist vigorous philosophical debates on the nature of the neural system. The presence of a philosophical debate related to scientific questions is not in itself unacceptable or abnormal, since in the history of science many scientific advances have arisen together with new philosophical statements. It is then worth not ignoring the debate underlying the neuroscience research and saying some words about the different opinions. We present four major typical opinions but the list do not pretend to be exhaustive.

1. The first belief, still largely represented, is that the Central Nervous System (CNS) is not understandable by scientific methods. It would be hanging on the "free will" or even be run by an unmaterial entity. Of course the research can hardly be sustained under such assumptions. On the contrary a more scientific approach consists in considering the CNS as entirely material and plausibly understandable (see e.g. [137] for a convincing criticism of the free will).
2. The second opinion, which is perhaps the most widely accepted between re-

searchers consists in considering the CNS as a generic learning mechanism. The CNS would learn the external world so that the complexity of the CNS would come more from the complexity of the external world than from the intrinsic complexity of the system. This viewpoint got importance in the 20th century with the *behaviourist psychology* which demonstrated the CNS capacities for learning and for being conditioned with in particular the previous works of the Pavlov (Nobel Price in 1904) [146] and Skinner [183]. They proposed a scientific methodology considering the CNS as a black box, an unknown system, and studying only the relationships between the inputs and the outputs. This vision got further support by the *connectionists* that propose a plausible neural framework for learning called *neural networks* [48]. Such networks simulating individual neurons are able to classify features even in a unsupervised fashion.

3. A third viewpoint could be the Freudian one which considers the CNS as subject to a unique pulsion or instinct which could be responsible of all the complexity of the human behaviour. This opinion does not seem to be represented in vision modeling.
4. The fourth opinion states that the cognitive processes are supported by complex innate structures which are not learned. The innate basement of the CNS originates in the evolution and determines in a large part which cognitive tasks a given specie is able to fulfill. Such viewpoint is advocated by Konrad Lorenz, Nobel Price in 1973 for the foundation of the *Ethology* (i.e. the study of animal behavior under natural conditions): "Below the diverse types of behaviour that the animals learn, underlie invariable motor rules they inherit. Those behavioural features constitute a specific characteristic [of the given specie], similar to the structure or the shape of the body" [115]. The same opinion is advocated by Noam Chomsky in linguistic "The evidence seems compelling, indeed overwhelming, that fundamental aspects of our mental and social life, including language, are determined as part of our biological endowment, not acquired by learning" ([25], p. 161), and also by Hubel and Wiesel Nobel Prices in 1981 for their physiological study of the primary visual cortex "most of the connections [in the primary visual cortex] seem to be wired up and ready to go at birth" [87]. Although this viewpoint is supported by various of the most important scientific personalities in different cognitive science fields, it not accepted by the majority of researchers and is still in debate. Moreover it is largely unknown from non-specialists (and often superficially known by researchers).

Such a debate between the "ambientalists" (behaviorists and connectionists) and the "innatists" has a large scientific relevance. It is also particularly animated because beyond the technical questions underly philosophical ideas that interfere with the debate

in a more importantly way that one could firstly suppose (the philosophical aspects underlying both viewpoints are summarized in the Appendix A). Both parts achieved decisive results, in particular concerning cortical functionalities for innatists and learning mechanisms for connectionists. This debate is productive and yields key findings within and outside the cognitive science community. It even influences our perception of the human nature, such as it seems that a very important scientific step will be passed when both viewpoints will be concealed.

The diversity of viewpoints gives rise to important differences in the modelling of the cognitive systems. The *ambientalists* (i.e. the second viewpoint) tend to insist in the ideas of the plasticity of the cortex (for example for recovering abilities after lesions) and in the adaptability to the environment [161]. They center their study on the different aspects of learning, and are supporters of building models including as less quantity of prior knowledge as possible, the systems being theoretically able to learn everything they need from the external world. They do not like to optimize the algorithms for a specialized task (this kind of implementation is called *soft programming*). Many researchers are seduced by this viewpoint because it let's hope to find a general and simple rule responsible for all the behaviour of the CNS. See e.g. [23] for a recent example of such an approach for contour integration in images. *Innatists* (i.e. the fourth viewpoint) proposed that the cortex is organized in fixed functional modules or areas [114], and that there exist multiple innate mechanisms underlying each individual neural process including the learning processes. They stress the limits of neural plasticity [185]. They are prone to include priory knowledge within biological models, so that the models can appear closer to classical image processing methods. See e.g. [83] for contour extraction based on a *hard wired* (i.e. without learning nor plasticity) biological model.

1.8 Novelty of the approach

In this thesis we adopt an apparently naive approach that consists in considering the visual system of primates as a nearly optimal system and to try to reproduce the cortical transformations as closely as possible for building efficient image processing algorithms. This approach is not the most common in image processing but it could provide good results first because of its novelty and second because it reverses another naive approach more commonly adopted: the technology is often perceived as a "progress" able to correct the defaults of the nature and to "improve" it. In other words the technology is ideologically considered as more perfect than the nature. It can be considered instead that the technology is highly imperfect in the sense that it brings many new

troubles and unbalances for each solution it proposes; and in opposition the nature, and particularly the visual cortex, can be considered as nearly optimal for the millions of years of evolution pressure, for its capacity of adaptation and because the biological visual system still offers much superior performances than any artificial vision system designed up to today.

The present work does not investigate deeply on the unification of multiresolution, edge extraction nor sparse approximation. It can be considered (and as been described so by various referees) as an experimental and descriptive study rather than a theoretical work. I desire to argue in favor of the present approach and rather to call into question the tendency to consider theorization and unification as the most worthy scientific objectives. Firstly because before unification and theorization, a work of description is necessary. A work of description of the functionalities of the primary visual areas is even more justified that we are just in the beginning of the understanding of the cortical circuitry so that the eventual phase of "unification" could still remain far. And secondly because there is no proof that any complex system can be sufficiently simplified to be solved analytically or even by simulation. Some complex and non-simplifiable system could exist and stay beyond the limits of formalization. It is then worth giving value to a descriptive work rather than uniquely waiting for a hypothetical unified theory of the visual system.

1.9 Organization of the thesis and description of the new contributions

The thesis is organized as follows. Chapter 2 presents methods for building self-invertible linear 2D log-Gabor wavelets that model the distribution of receptive fields of simple cortical cells. The efficiency of those wavelets is illustrated in an image denoising scenario. The present log-Gabor wavelets are original in two important aspects: first it includes localized oriented high-pass filters, and secondly, they are self-invertible. This is an important improvement since no exact reconstruction Gabor-like multiresolutions were previously proposed in the literature. Chapter 3 details energy methods for edge and line extraction based on 2D log-Gabor wavelets. Such feature extraction method is inspired from complex cortical cell models. It proposes a new method as a synthesis of different methods previously proposed in the literature. Chapter 4 introduces a new sparse approximation algorithm based on the edge/line extraction of the chapter 3. The proposed sparse approximation is original since it is the first sparse approximation method based on edge/line extraction. The scheme can be considered as similar to the Matching Pursuit method but incorporating a non-iterative process for coefficient

selection. Chapter 5 presents a separate work on the grouping of edge segments which is not based on log-Gabor wavelets but on an alternative representation called *tensor voting*. It introduces new contributions to tensor voting by proposing an iterative version and the use of a curvature representation for improving the integration of contours. Chapter 6 gathers the conclusions of the thesis.

Chapter 2

Self-invertible log-Gabor wavelets

"The most important and urgent problems of the technology of today are no longer the satisfactions of the primary needs or of archetypal wishes, but the reparation of the evils and damages by technology of yesterday."

Dennis Gabor, *Innovations: scientific, technological and social.*

Orthogonal and biorthogonal wavelets became very popular image processing tools but exhibit major drawbacks, namely a poor resolution in orientation and the lack of translation invariance due to aliasing between subbands. Alternative multiresolution transforms which specifically solve these drawbacks have been proposed. These transforms are generally overcomplete and consequently offer large degrees of freedom in their design. At the same time their optimization gets a challenging task. We propose here the construction of log-Gabor wavelet transforms which allow exact reconstruction and strengthen the excellent mathematical properties of the Gabor filters. Two major improvements on previous Gabor wavelet schemes are proposed: first the highest frequency bands are covered by narrowly localized oriented filters. Secondly, the set of filters cover uniformly the Fourier domain including the highest and lowest frequencies and thus exact reconstruction is achieved using the same filters in both the direct and the inverse transforms (which means that the transform is self-invertible). The present transform not only achieves important mathematical properties, it also follows as much as possible the knowledge on the receptive field properties of the simple cells of the Primary Visual Cortex (V1) of primates and on the statistics of natural images. Compared to the state of the art, the log-Gabor wavelets show excellent ability to segregate the image information (e.g. the contrast edges) from spatially incoherent Gaussian noise by hard thresholding and then to represent image features through a reduced set of large magnitude coefficients. Such characteristics make the transform a promising tool

for processing natural images. This study was previously communicated in parts in [65, 62, 60, 63, 59, 169].

2.1 Introduction

After the development and rapid successes of the wavelet transforms in image processing, alternative multiresolutions have been proposed mainly for a better resolution in orientation and for avoiding aliasing effects. Last fifteen years oriented band-pass multiresolution transforms have arisen with increased importance thanks to the development of steerable pyramids [182], Gabor-like multiresolutions [74, 134, 173, 180], complex-valued wavelets [106, 74, 96, 159], curvelets [189] and contourlets [41], to name a few.

A Gabor function is a Gaussian multiplied by a complex exponential. Thus, in the Fourier domain, it is a Gaussian shifted from the origin. Gabor functions gather a number of interesting mathematical properties: first they have a smooth and infinitely differentiable shape. Second, their modulus are monomodal, i.e. they have no side lobes neither in space nor in the Fourier domain. Third, they are optimally joint-localized in space and frequency [68]. For such good properties they have been proposed as ideal functions for signal processing.

Similarly, 2D Gabor functions are highly jointly localized in position, orientation and spatial frequency. Neuroscience studies have shown that the receptive fields of simple cells of the Primary Visual Cortex (V1) of primates can be modelled by Gabor functions [36, 124]. Considering natural vision as optimized by the millions of years of evolutionary pressure, the recruitment of Gabor functions by V1 can be seen as an additional argument in favor of their adequacy for image processing.

Nevertheless Gabor functions present some important drawbacks. First, it is not possible to build a complete orthogonal basis of Gabor functions, therefore non-orthogonal dictionaries have to be employed. Non-orthogonality implies that exact reconstruction using the same filters for analysis and synthesis will not be possible unless an over-complete dictionary is considered. Secondly, Gabors are bandpass filters, they are consequently inadequate for covering the lowest and highest frequencies. Thirdly, it is particularly difficult to cover up the mid frequencies with sufficient uniformity.

Gabor multiresolutions have been successfully used for image analysis and applications where exact reconstruction is not required, such as texture analysis [27, 173],

texture synthesis [158], edge/contour extraction [83, 102, 73], or object recognition [164, 103, 92]. And, even without exact reconstruction they have been shown useful for image restoration applications [32, 101, 131, 26]. In parallel, different methods for reconstruction improvement have been proposed to recover the highest frequencies [134], to avoid excessive low-pass overlapping [53], to improve the reconstruction [37, 108] or to cover more uniformly the Fourier domain [74].

One of the important applications where oriented multiresolution schemes appeared to be particularly efficient is image denoising. In this field (bi-)orthogonal wavelets are themselves one of the most popular and optimal tools [44]. Nevertheless it has been shown that undecimated wavelets provide better results than the critically sampled wavelets [28, 20, 111]. Latter works have shown the interest of redesigning the Fourier domain tiling, particularly for increasing the number of orientations [189, 41, 160]. The advantage of using complex-valued wavelets were discussed in [6, 101]. Indeed the relaxation of the critical sampling constraint provides high degrees of freedom in the construction of the multiresolution scheme and it is worth taking advantage of this additional flexibility for choosing the filter shape, the bandwidths, the degree of overlapping between filters, the number of orientations, or for choosing complex-valued filters, etc. We propose here a transform optimized following biological, perceptual and natural image statistics criteria besides the mathematical ones. The rationale for taking into account those criteria is: (1) the number of free parameters on the filter shape and multiresolution arrangement is prohibitive for a systematic study. (2) The Primary Visual Cortex supposedly evolved towards an efficient and robust image processing system adapted to natural images. Considering the important similarity between V1 simple cells and oriented multiresolution transforms, we hypothesize that biological knowledge on V1 could serve as an useful guide for the choice of the free parameters. Moreover as an interesting feedback, the implementation of biologically inspired wavelet transforms could help for the understanding of V1. (3) It is important to provide good image quality from the human perceptual point of view because digital images are almost exclusively used by humans. It would then be of great help to construct methods limiting the appearance of perceptually salient artifacts. The similarity with the biological models could serve such objective. (4) The multiresolution must be built according to the type of signal to analyze. Here as we focus on natural images, studies on their statistics can also be used as guidelines. Moreover it is worth noting that there are direct relationships between receptive fields of V1 and statistics of natural images [53, 141, 43].

We propose then here an implementation of a multiresolution transform fulfilling the following constraints: (1) an optimal localization in space, frequency and orientation through the use of Gabor filters; (2) an augmented number of orientations; (3) a resemblance to the receptive field of V1 simple cells; (4) an exact reconstruction

by a self-invertible transform (the same transform functions are used for the direct and inverse transforms) and (5) complex-valued filters sensitive to both symmetric and antisymmetric features. We pay special attention to the different aspects of the multiresolution design and we propose a series of solutions for improving the efficiency of the Gabor wavelets scheme. In comparison with previous implementations of Gabor multiresolution ([74, 134, 173, 158]), the most important novelties of the present work consists in incorporating complex-valued oriented high-pass filters, and achieving the exact reconstruction and self-invertibility properties.

The present chapter fulfills two objectives: first to describe in detail in the Section 2.2 the proposed self-invertible log-Gabor wavelet transform and second to illustrate (in Section 2.3) the efficiency of the method in an image denoising scenario. Finally, we conclude in Section 2.4.

2.2 Method

Let's consider a $n \times n$ square image \mathbf{x} of $N = n^2$ pixels. The Gabor wavelets \mathbf{W} consist in filtering the input image $\mathbf{x} \in \mathbb{R}^N$ by a set of filters $(\mathbf{G}_r)_{r \in \Omega}$. The band-pass, high-pass and low-pass filters are described respectively in Sections 2.2.1, 2.2.2 and 2.2.3. Direct and inverse transforms are defined in Sections 2.2.4 and 2.2.5, respectively. Section 2.2.6 deals with the matrix notation and Section 2.2.8 presents the border strategy. Section 2.2.7 describes the downsampling method. Methods for obtaining exact reconstruction are described in Section 2.2.9.

2.2.1 Bandpass log-Gabor filters

Classical Gabor filters give rise to important difficulties when implemented in multiresolution. Filters overlap more importantly in the low frequencies than in the higher ones yielding a non-uniform coverage of the Fourier domain. Moreover Gabor filters have not zero mean, they are then affected by DC components. For those reasons log-Gabor filters are used in the present implementation instead of Gabor filters. The log-Gabor filters lack DC components and can yield a fairly uniform coverage of the frequency domain in an octave scale multiresolution scheme [53]. The log-Gabor filters are defined in the log-polar coordinates of the Fourier domain as Gaussians shifted from the origin (see the filter shapes for different number of orientations Fig. 2.1, Fig. 2.2, Fig. 2.3 and

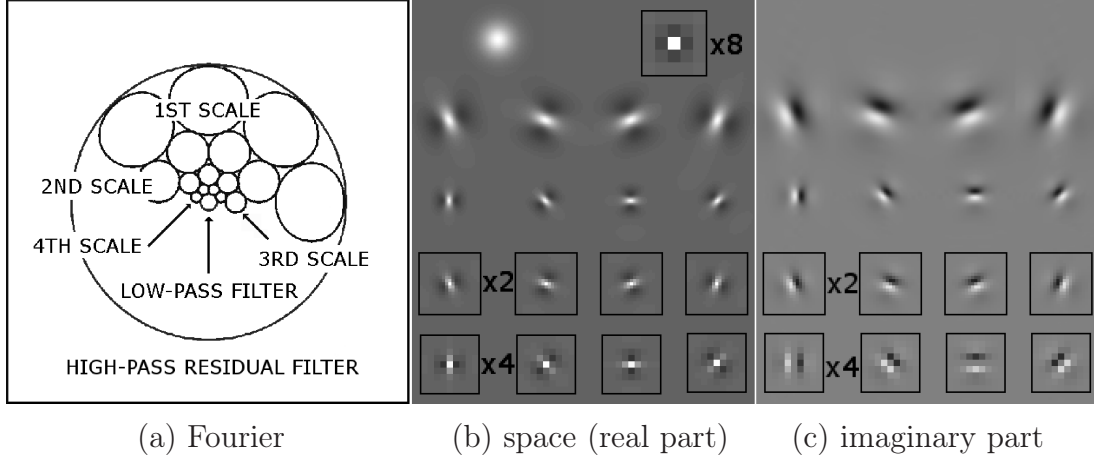


Figure 2.1: A log-Gabor wavelet scheme including 4 scales, 4 orientations and isotropic low-pass and high-pass filters. **a.** The Fourier domain is covered by log-Gabor filters with 4 scales, 4 orientations, a low-pass filter and a high-pass filter. Only the contours at 78% of the maximum of amplitude are depicted. The high-pass residual filter is represented by the region outside the bigger circle (see [134, 60] for possible implementations of such a filter). **b.** Real parts of log-Gabor filters on 4 scales and 4 orientations are drawn in the space domain. The low-pass and high-pass filters can also be seen on the top left and top right of the image, respectively. 2nd scale, 1st scale and high-pass filters are magnified respectively by factors of 2, 4 and 8 for a better visualization. **c.** Imaginary parts of log-Gabor filters in the spatial domain. Note low-pass and high-pass filters do not have imaginary part.

Fig. 2.4.a):

$$\mathbf{G}_{(s,t)}(\rho, \theta) = \exp\left(-\frac{1}{2}\left(\frac{\rho-\rho_s}{\sigma_\rho}\right)^2\right) \exp\left(-\frac{1}{2}\left(\frac{\theta-\theta_{(s,t)}}{\sigma_\theta}\right)^2\right)$$

$$\text{with } \begin{cases} \rho_s = \log_2(n) - s \\ \theta_{(s,t)} = \begin{cases} \frac{\pi t}{n_t} & \text{if } s \text{ is odd} \\ \frac{\pi}{n_t}(t + \frac{1}{2}) & \text{if } s \text{ is even} \end{cases} \\ (\sigma_\rho, \sigma_\theta) = 0.996(\sqrt{\frac{2}{3}}, \frac{1}{\sqrt{2}}\frac{\pi}{n_t}) \end{cases} \quad (2.1)$$

where (ρ, θ) are the log-polar coordinates (in \log_2 scale, indicating the filters are organized in octave scales); $n_s = 5$ is the number of scales of the multiresolution scheme and n_t is the number of orientations (n_t will range between 3 to 20. 8 orientations will be used as a typical value). $s \in \{1, \dots, n_s\}$ and $t \in \{1, \dots, n_t\}$ indexes the scale and the orientation of the filter, respectively; $(\rho_s, \theta_{(s,t)})$ are the coordinates of the center of the filter; and $(\sigma_\rho, \sigma_\theta)$ are the bandwidths in ρ and θ , common for all filters.

It is highly justified here to consider the Fourier domain through the log-polar coordinates. Indeed in such coordinate system, the octave distribution of the filters

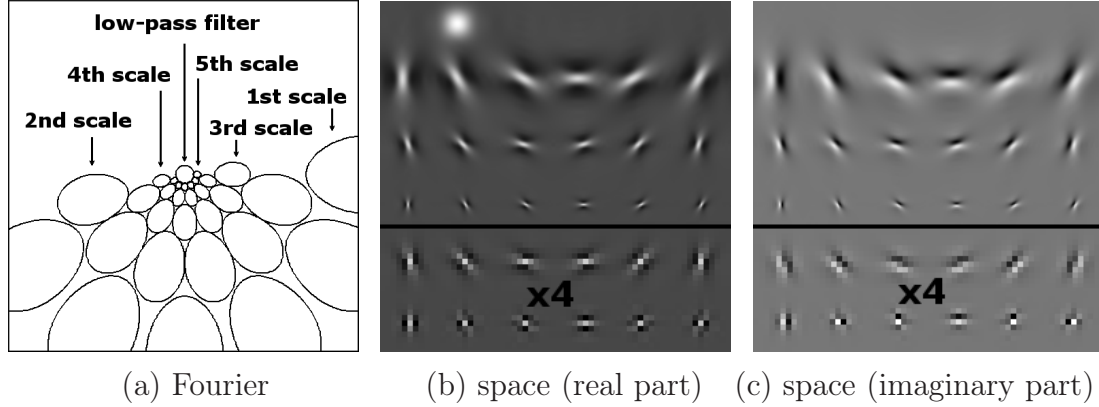


Figure 2.2: Multiresolution scheme with 5 scales and 6 orientations. **a.** Schematic contours of the filters in the Fourier domain. **b.** Real part of the filters. The two first scales are drawn at the bottom magnified by a factor of 4. Scales are arranged in rows and orientations in columns. The low-pass filter is drawn in the upper-left part. **c.** The imaginary part of the filters. The low-pass filter does not have imaginary part.

constitutes a regular grid, the center of the filters defined by Eq. 2.1 laying on an uniform hexagonal lattice. Moreover in such coordinates the filters are purely Gaussian and their bandwidths are constant across orientations and scales.

The filters defined by Eq. 2.1 cover only one side of the Fourier domain (see Fig 2.3.a). It is worth considering them in Fourier as the sum of a symmetric and an antisymmetric components. Thus, those components sum their amplitude in one side of the Fourier domain and cancel themselves in the other side. This explains that in the spatial domain, the filters have both a real part (with cosine shape due to the symmetric component) and an imaginary part (with sine shape due to the antisymmetric component). A single log-Gabor filter defined in Fourier by Eq. 2.1 yields then both a real and an imaginary part in the spatial domain. Both parts can be seen for the 4 scale and 4 orientation scheme in Fig. 2.1.b and c respectively (see also in Fig. 2.2.b and c and Fig. 2.3.b and c for other schemes). They can be compared with the filters of other multiresolution transforms (orthogonal wavelets 'Db4' [33] in Fig. 2.4.b and steerable pyramids [160] in Fig. 2.4.c).

One objective of this chapter is to choose the transform parameters as close as possible of the known physiology of simple cortical cells. According to such purpose, log-Gabor filters are chosen for modeling the receptive field of the simple cortical cells as proposed in [36, 124, 53]. Simple cells are known to be organized in pairs in quadrature of phase [157], justifying the choice of complex-valued filters. The choice of the bandwidth in orientation is motivated by the simple cells orientation resolution which has been evaluated as around 20-40 degrees of full bandwidth at half response [87, 36].

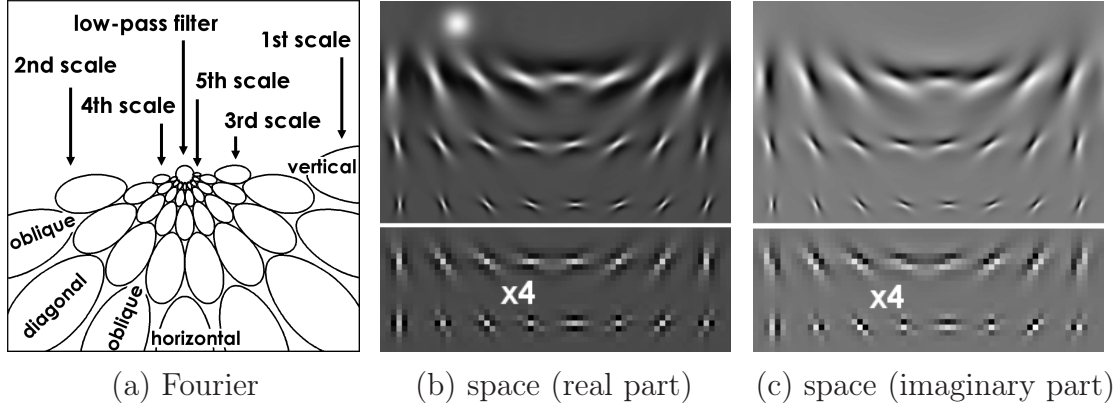


Figure 2.3: Multiresolution scheme for 5 scales and 8 orientations. **a.** Schematic contours of the log-Gabor filters in the Fourier domain with 5 scales and 8 orientations (only the contours at 78% of the filter maximum are drawn). **b.** The real part of the corresponding filters is drawn in the spatial domain. The two first scales are drawn at the bottom magnified by a factor of 4 for a better visualization. The different scales are arranged in lines and the orientations in columns. The low-pass filter is drawn in the upper-left part. **c.** The corresponding imaginary parts of the filters are shown in the same arrangement. Note that the low-pass filter does not have imaginary part. Insets (b) and (c) show the final filters built through all the processes described in Section 2.2.

Such orientation bandwidth would require around 6 to 13 orientations to cover the 180 degrees of the plane. For the proposed scheme using the typical value of 8 orientations we obtain filters with 31.8 degree full bandwidth at half response. The bandwidth in scale of the simple cells has been evaluated between 0.6 and 3.2 octaves [40], and around 1.3 octaves in mean [36] (at half response). The present filters have a 1.43 octave bandwidth. De Valois and al. [40] reported the existence of cells covering different scales over at least 4 octaves for each retina location. At least 3 scales for each retina location have been encountered, which justifies the choice of a multiscale transform. Additionally the shape of the 8 orientation-scheme filters shown Fig. 2.3.b-c appears close to the independent components learned by sparse coding or ICA techniques on natural images [141, 43], confirming the adequacy of the filters to match natural image features. By its parts the modulus operation on real and imaginary parts which will be exploited for denoising in Section 2.3 and the resulting independency toward phase are characteristic of the complex cortical cells [17]. The thresholding operation used in the denoising method has also a biological justification [153]. The present scheme gets then close to the V1 physiology particularly in terms of receptive field structures and multiscale and multiorientation arrangement.

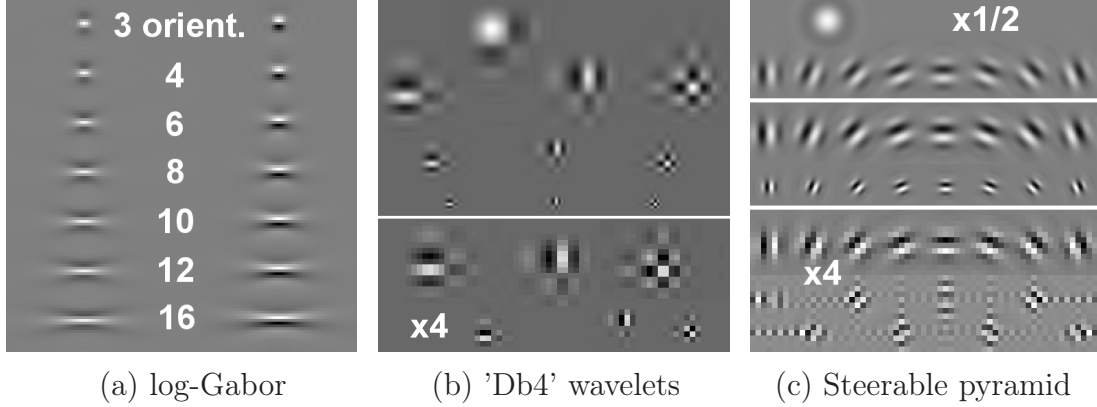


Figure 2.4: Filters from different multiresolution schemes. **a.** In the proposed scheme the elongation of log-Gabor wavelets increases with the number of orientations n_t . Here the real parts (left column) and imaginary parts (right column) are drawn for the 3, 4, 6, 8, 10, 12 and 16 orientation schemes. **b.** As a comparison with log-Gabor scheme, orthogonal wavelets filters 'Db4' are shown. Horizontal, vertical and diagonal wavelets are arranged on columns (low-pass on top). **c.** As a second comparison, steerable pyramid filters [160] are shown. The arrangement over scales and orientations is the same as for the log-Gabor scheme.

2.2.2 High-pass oriented filters

An important problem appears in the first scales where from Eq. 2.1, \mathbf{G}_r would have significant amplitude above the Nyquist frequency ($\rho \geq \log_2(\frac{n}{2})$). Cutting off abruptly the filter response above the Nyquist frequency strongly distorts the filter shape in the spatial domain (i.e. causes the appearance of side lobes or ringing). For this reason in many implementations high frequencies are not covered, e.g. in [173]. Alternatively, to not discard this part of the spectrum Nestares et al. included a non-oriented high-pass filter [134]. Nevertheless when it is used in denoising or compression, a non-oriented high-pass filter introduces cross-shaped artifacts, which are very salient and artificial-looking. We propose here several solutions to design oriented high-pass filters having smooth shape without extra side lobes. They consist first in incorporating a half-pixel shift in the spatial position of the imaginary part of the filters. This shift allows the first scale filters to capture much more adequately the antisymmetric features, as it will be illustrated Section 2.2.2.C, but consequently the real and imaginary parts of the filters have to be separately defined. Also vertical and horizontal filters are designed in a different way than the other filters, i.e. the oblique ones (Section 2.2.2.A and B). Note that, because the definition of the first scale filters differs from the other scales, the transform is no more strictly a "wavelet" one, but it can be considered as a wavelet-like transform since the general shape of the decomposition functions is obtained by translation, dilatation, and rotation of a mother function if we except the modifications

proposed here for improving the reconstruction performance. Moreover the proposed Gabor wavelets build directly on the original Morlet and coworkers proposition of using Gabor functions for a cycle-octave multiresolution [132].

A. Real part of the horizontal and vertical filters

The real part of horizontal and vertical filters is defined by central symmetry as shown in Fig. 2.5. Thus they are continuous across the periodicities of the Fourier domain which is important since strong discontinuities in the Fourier domain create side lobes and a worst localization in the spatial domain. They are also well localized and without extra side lobes in the space domain (see Fig. 2.3.b, bottom row, 1st and 4th columns).

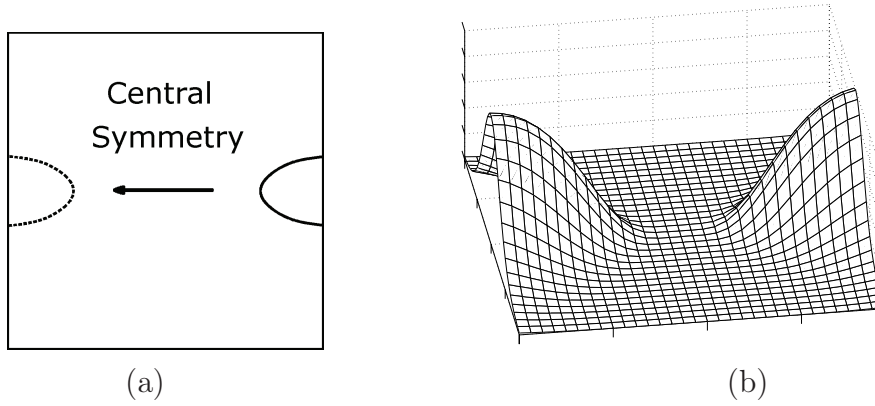


Figure 2.5: The real part of the first scale vertical filter is constructed in Fourier by central symmetry. **a.** The contours of the filters are drawn in the Fourier domain at 78% of the filter maximum. The filter is build up by summing its central symmetry. **b.** Resulting filter shown in the Fourier domain (horizontal axis represent the Fourier space having its origin in the center and the vertical axis represents the amplitude). This filter is shown in spatial domain in Fig. 2.3.b bottom row, left column.

B. Real part of oblique filters

The real part of oblique filters (i.e. filters which are not vertical nor horizontal) is also defined by central symmetry. But this is not sufficient to maintain the Fourier domain continuous (across periods) and to keep a good localization in the space domain. We propose then to fold up those filters by considering them as periodic with periodicity n . Nevertheless when the part beyond the Nyquist frequency ($|u| \geq \frac{n}{2}$ or $|v| \geq \frac{n}{2}$, where (u, v) are the Cartesian coordinates of the Fourier domain) is folded up by periodicity, it covers the whole spectrum with significant amplitude. To maintain the filter selectivity

to high-frequencies, it is then necessary to filter down the induced lowest frequencies by multiplying the folded part by an attenuation factor α defined here as a raised cosine function (see also Fig. 2.6 for an illustration of the construction of such filters):

$$\alpha = \begin{cases} \frac{1}{2} \cos(\pi \frac{d}{n/3}) + \frac{1}{2} & \text{if } d < n/3 \\ 0 & \text{if } d > n/3 \end{cases} \quad (2.2)$$

where d is the Euclidian distance to the closest frequency inside the Nyquist range (i.e. u and v within $[-n/2, n/2]$).

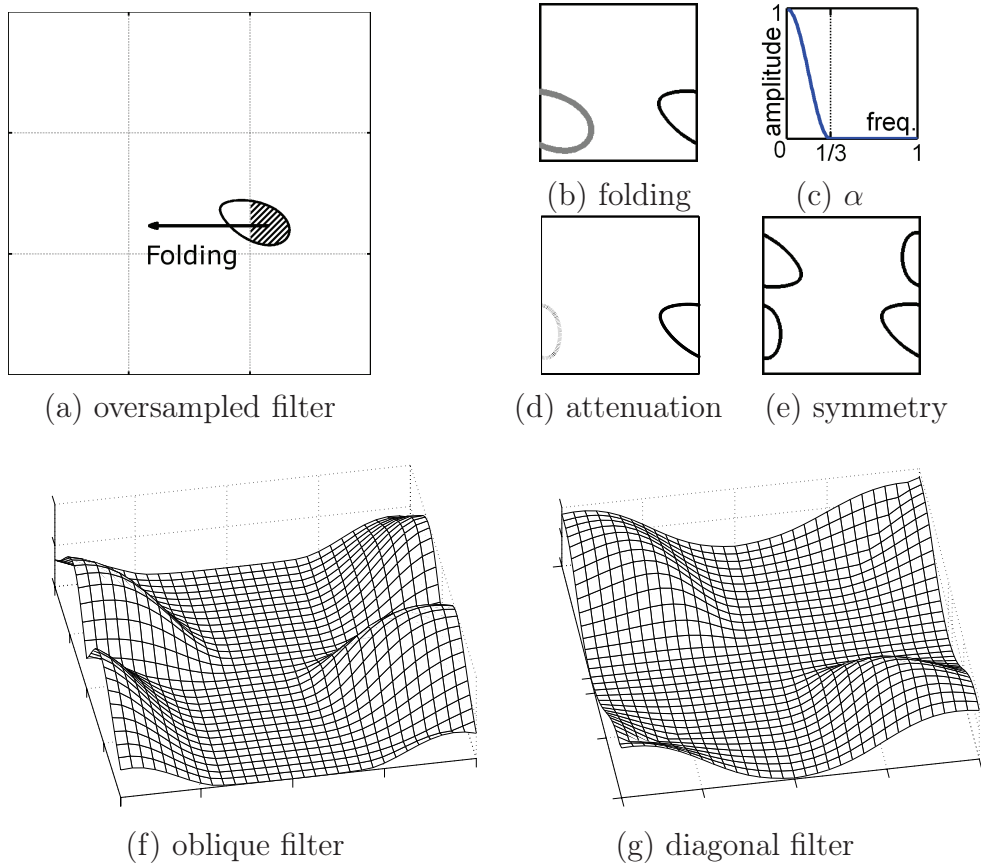


Figure 2.6: Construction of the real part of diagonal and oblique first scale filters in the Fourier domain. **a.** The frequency domain is oversampled 3 times. **b.** The frequencies above the Nyquist frequency are folded up by periodicity. **c.** Raised cosine function α . **d.** The folded part of the filter is attenuated multiplying by the raised cosine α . **e.** The central symmetry is summed up for the construction of a real-valued filter. **f.** Resulting oblique filter. **g.** Resulting diagonal filter. These filters are shown in the spatial domain in Fig. 2.3.b bottom row.

C. Imaginary part of the horizontal and vertical filters

The imaginary part is build in Fourier similarly to the real part (Section 2.2.2.1) but antisymmetrically instead of symmetrically. In the spatial domain the imaginary part have to be antisymmetric, but here the axis of central antisymmetry will not be localized in a pixel but in between two pixels, this for two related reasons. First, in the spatial domain, an antisymmetric filter would have a zero as central coefficient. And, as first scale filters have a frequency close to the Nyquist frequency, i.e. an approximately two pixel period, the filter would be almost null in its central part. The filter would then be very coarsely localized, having most of its amplitude far from its center. Consequently the filter must be centered in between two pixels (see Fig. 2.7.a). Second, in the Fourier domain, an antisymmetric high-pass filter is not continuous across periods while a half-pixel shifted version is (see Fig. 2.7.b). For those reasons the filter is multiplied by a $e^{i\pi\frac{u}{n}}$ or $e^{i\pi\frac{v}{n}}$ function in the Fourier domain (respectively for horizontal and vertical filters) which induces a half pixel displacement in the space domain. As a consequence the Fourier coefficients are now complex-valued and both the real and imaginary Fourier parts of the filters become continuous across periods (see Fig. 2.7.b).

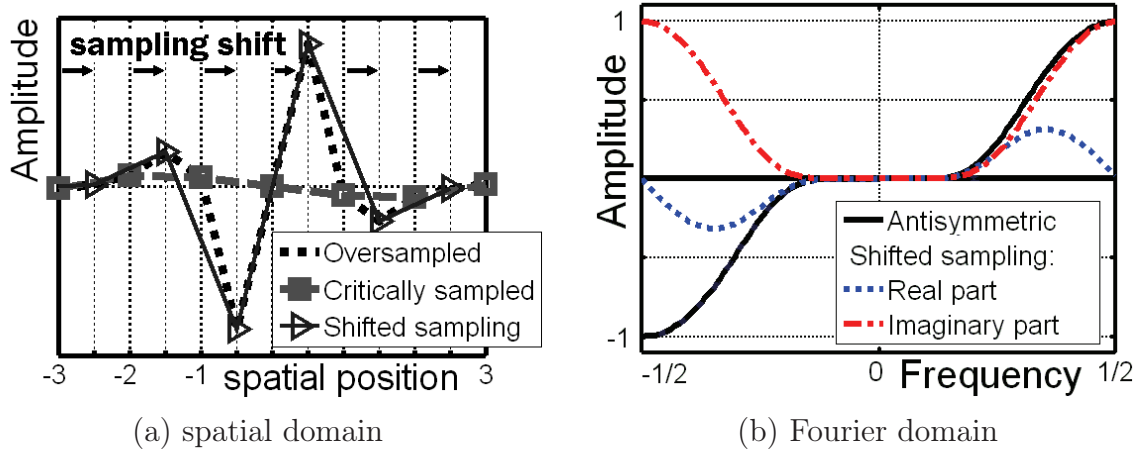


Figure 2.7: Imaginary part construction of horizontal and vertical first scale filters. **a.** In the spatial domain it is necessary to shift the sampling grid by half a pixel. A non-shifted sampling yields a function almost null in the central coefficients, most of the amplitude being localized far from the center inducing a very poor localization of the filter. The same antisymmetric filter built on a half-pixel shifted grid is much more closely localized. **b.** Shifting the sampling grid is realized by multiplying the frequencies by $e^{i\pi\frac{u}{n}}$ (inducing the filter coefficients are complex-valued also in the Fourier domain). It makes the filter continuous in its frequency coverage across periods.

D. Imaginary part of oblique filters

Those filters are defined antisymmetrically and are folded by periodicity with the same attenuation function α described in Section 2.2.2.B. Moreover they are also shifted half the interpixel distance perpendicularly to the preferred direction of the filter by multiplication in Fourier by $e^{\frac{i\pi}{n}(u \cdot t_u + v \cdot t_v) \cdot \max(|t_u|, |t_v|)}$, where (t_u, t_v) is the normalized preferred direction vector of the filter.

E. Second scale filters

Second scale filters are also folded by periodicity with attenuation α (see Section 2.2.2.B). Here nevertheless it is not necessary to shift half-pixel in position the antisymmetric filter and the same definition is used for all orientations. An example of second scale filter is shown in Fig. 2.8. In the coarser scales (i.e. $s \geq 3$), the filter part beyond the Nyquist frequency has sufficiently low amplitude to be cut off.

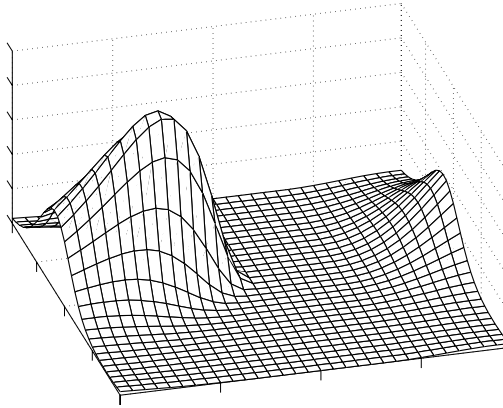


Figure 2.8: Second scale filters are constructed in Fourier by folding the filters by periodicity and applying the attenuation function α on the folded part. (Second scale filters are shown in the spatial domain in Fig. 2.3.b,c next to last rows.)

2.2.3 Low-pass filter

The multiresolution scheme is completed with a low-pass filter (DC) for recovering the luminance information. The low-pass filters could be defined simply as a Gaussian low-pass filter as in [158]. Nevertheless for a better filling-in of the low-pass residual frequencies not covered by the log-Gabor filters, it is defined in a different way. Two

additional scales above the number of scales n_s deployed are built and summed up together (as the root sum squared). Moreover the part inside the highest additional scale is set up to one. In practice if 5 scales are deployed, the filters that would correspond to the 6th and 7th scales are summed and additionally the space inside the 7th scale ($\rho < \log_2 n - 7$) is set up to one (the resulting filter can be seen in the spatial domain in Fig. 2.3.b, upper left part.)

2.2.4 Direct log-Gabor wavelet transform

The whole set of bandpass filters, high-pass filters and the low-pass filter are indexed by $\Omega = \{1, \dots, n_s n_t + 1\}$. The filter set defined in the Fourier domain is then refereed as $(\mathbf{G}_r)_{r \in \Omega}$. The convolution of the image \mathbf{x} by any filter \mathbf{G}_r is \mathbf{h}_r which is called a *channel*. It is computed in the Fourier domain as:

$$\mathbf{h}_r = \mathcal{F}^{-1}(\mathbf{G}_r \cdot \mathcal{F}(\mathbf{x})) \quad (2.3)$$

where \mathcal{F} and \mathcal{F}^{-1} are respectively the direct and inverse discrete Fourier transform. The whole set of channels, $(\mathbf{h}_r)_{r \in \Omega}$, is called a *pyramid* (see Fig. 2.9 for an example of such a pyramid for a 5 scale, 6 orientation scheme). Finally, the log-Gabor wavelet transform of \mathbf{x} is defined by:

$$\mathbf{W}(\mathbf{x}) = (\mathbf{h}_r)_{r \in \Omega}. \quad (2.4)$$

2.2.5 Reconstruction transform

The reconstruction transform \mathbf{W}^\dagger consists in filtering in Fourier each channel \mathbf{h}_r by the corresponding filter $\overline{\mathbf{G}_r}$ (where $\overline{\mathbf{G}_r}$ is the complex conjugate of \mathbf{G}_r . Note that in most cases the filter \mathbf{G}_r in Fourier is real-valued, so that $\overline{\mathbf{G}_r} = \mathbf{G}_r$):

$$\mathbf{z}_r = \mathcal{F}^{-1}(\overline{\mathbf{G}_r} \cdot \mathcal{F}(\mathbf{h}_r)) \quad (2.5)$$

Each \mathbf{z}_r is called a *reconstructed channel*. All reconstructed channels are finally summed up to obtain the reconstruction:

$$\mathbf{W}^\dagger(\mathbf{h}) = \sum_{r \in \Omega} \mathbf{z}_r = \mathcal{F}^{-1} \left(\sum_{r \in \Omega} \overline{\mathbf{G}_r} \cdot \mathcal{F}(\mathbf{h}_r) \right) \quad (2.6)$$

Because the image \mathbf{x} is real, its Fourier transform has hermitian symmetry. For this reason, we used band-pass Gabor filters covering only half of the Fourier plane,

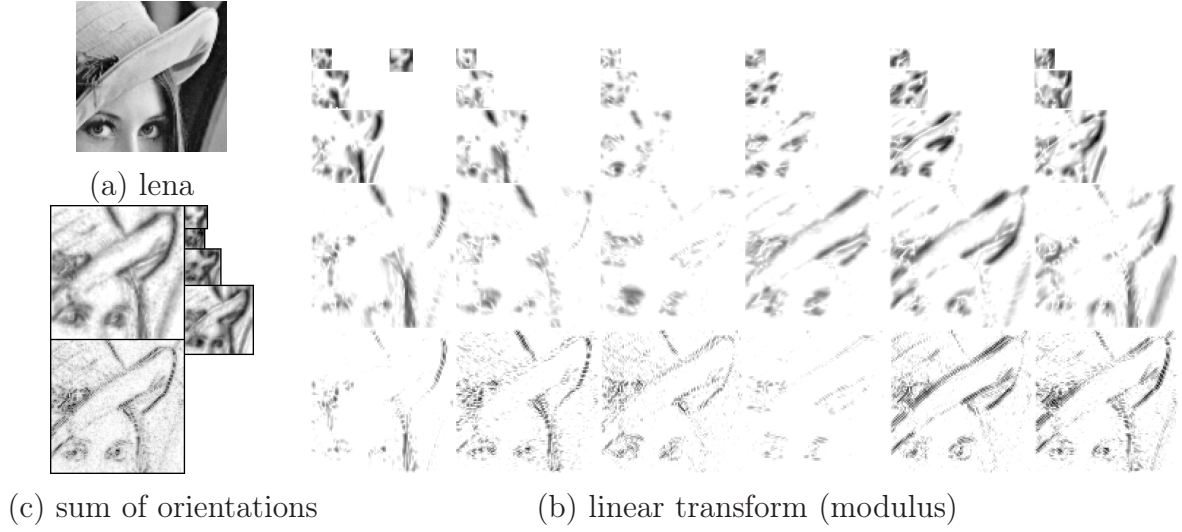


Figure 2.9: 5 scales, 6 orientations pyramid of a 96×96 detail of the image 'Lena' ('lena' is shown in the inset (a)). In (b) the 5 scales are arranged in rows. Their different sizes are due to the different downsampling applied to each scale (see Section 2.2.7). Along columns are arranged the six orientation channels. The low-pass filter is shown in the first row between the two first orientations. Only the moduli of the coefficients are shown here. Black colors indicates high moduli values, and white colors low coefficient values. In (c) all the orientations are summed up for each scale. The largest insets represent the first and second scale (respectively for the bottom and upper insets), then from the largest to the smallest, the 3rd, 4th, low-pass and 5th scales are drawn.

while the other half plane is further completed at the reconstruction using hermitian symmetry (this completion is obviated in the equations).

Because the transform is overcomplete, it can not be strictly bijective nor invertible: the transform domain has not the same dimension as the original domain. But, under the conditions that will be further exposed in Section 2.2.9, \mathbf{W}^\dagger can be the pseudo-inverse of \mathbf{W} so that $\mathbf{W}^\dagger(\mathbf{W}(\mathbf{x})) = \mathbf{x}$, $\forall \mathbf{x} \in \mathbb{R}^N$.

2.2.6 Matrix notation

Because the transform is linear, it can also be viewed as a matrix operation $W \in \mathbb{R}^{N \times M}$. \mathbf{W} can then be described as the scalar product of $\mathbf{x} \in \mathbb{R}^N$ with a family of decomposition functions $(\mathbf{w}_k)_{k \in \{1, \dots, M\}}$ (the spatial log-Gabor functions), with each $\mathbf{w}_k \in \mathbb{R}^N$:

$$\mathbf{W}\mathbf{x} = \mathbf{h} = (h_k)_{k \in \{1, \dots, M\}} = (\langle \mathbf{w}_k, \mathbf{x} \rangle)_{k \in \{1, \dots, M\}} \quad (2.7)$$

Because we use the same filters \mathbf{w}_k for reconstruction, the reconstruction transform is W^T (where T refers to the transposed matrix) and for all $\mathbf{h} \in \mathbb{R}^M$:

$$W^T \mathbf{h} = \sum_{k=1}^M h_k \mathbf{w}_k \quad (2.8)$$

Under the conditions defined in Section 2.2.9, W^T provides exact reconstruction. Thus the transform is said *self-invertible* (i.e. exact reconstruction is achieved using the same filters both for the decomposition and the reconstruction) [182].

2.2.7 Sparse downsampling

To limit the expansion of the transform domain it is desirable to downsample the channels as much as possible, but we desire to avoid at the same time any kind of aliasing. The *sparse downsampling* method [219] can be applied advantageously here for downsampling 3rd, 4th, 5th and low-pass filters both in u and v by factors of 2, 4, 8 and 8, respectively. The procedure is implemented in the Fourier domain as follows: a window of the defined downsampling size is centered on the filter and its position is adjusted so as to retain inside the maximum quantity of the filter amplitude (this operation is achieved by displacing iteratively the window in all the possible directions). Filter coefficients outside the window are zeroed out. In this way, only filter coefficients smaller than 5% of the filter maximum are cut off which preserves the smooth shape of the filters in the spatial domain. Using those filters, the channels can then be downsampled in an invertible manner without mixing frequencies i.e. without introducing aliasing: any frequency of the downsampled channel corresponds bijectively to one unique non-zeroed frequency in the upsampled version of the channel (the sparse downsampling procedure is illustrated in Fig. 2.10).

The present log-Gabor filters which are complex-valued in the spatial domain bring then an additional advantage: the filters lie only in one side of the Fourier domain allowing the implementation of the sparse downsampling method. In comparison, pure real or pure imaginary spatial filters are symmetric or antisymmetric in Fourier and compel to downsample at Nyquist frequency, that is twice the highest frequency. The sparse downsampling compensate largely that two real values (i.e. a complex one) are obtained for each coefficient: complex-valued log-Gabor filters can still be downsampled by a factor around two times larger than real-valued filters. Consequently, the overcompleteness factor measured as the ratio M/N , where M is the dimension of the transform domain and N is the dimension of the image, is reduced to $M/N \simeq 40$ (it can be approximated as $\frac{14}{3}n_t$ where n_t is the number of orientations).

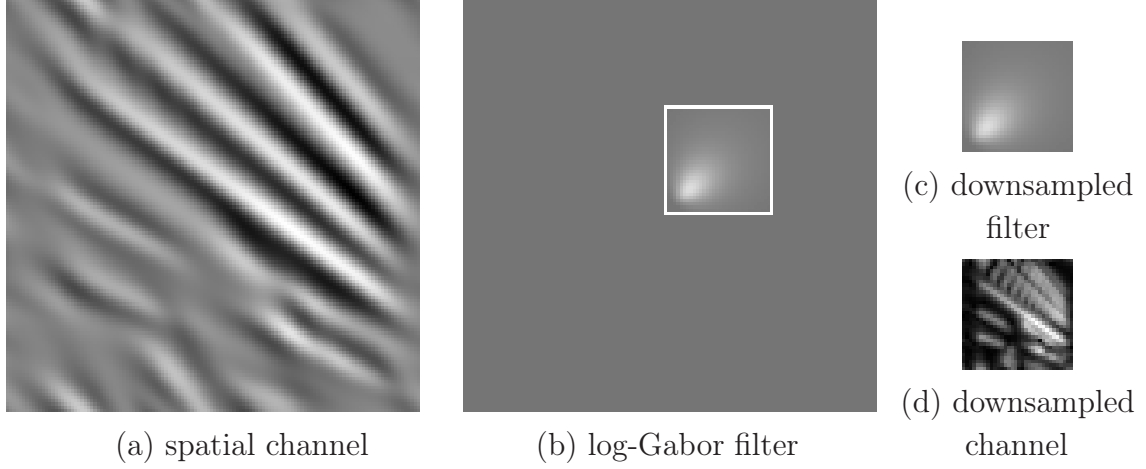


Figure 2.10: Sparse downsampling method. **a.** 4th scale 2nd orientation channel of the image "MIT" (the original can be seen in Fig. 2.13). Only the moduli of the complex coefficients are shown. **b.** Corresponding log-Gabor filter (scale 4, orientation 2) represented in the Fourier domain (Fourier origin is located in the center of the inset). A window of the desired downsampling size is adjusted on the filter so as to retain inside the maximum of the amplitude. Filters coefficients outside the window are zeroed out (only coefficients lower than 5% of the filter maximum are eliminated). **c.** The filter can be downsampled in an invertible manner by a factor 4×4 by removing the zero coefficients. **d.** The resulting downsampled channel obtained after inverse Fourier transform (as in (a) only the moduli of the coefficients are shown).

In parallel, it is straightforward to prove that exact reconstruction using the same non-orthogonal decomposition functions for the analysis and the synthesis requires the dictionary to be overcomplete.

Proof: Consider a signal $\mathbf{x} \in \mathbb{R}^N$, and a non-orthogonal dictionary $(\mathbf{w}_i)_{i \in \{1, \dots, M\}}$ of \mathbb{R}^N . The linear transform is $(h_i)_{i \in \{1, \dots, M\}}$ with $h_i = \langle \mathbf{x}, \mathbf{w}_i \rangle$, where \langle, \rangle represents the scalar product. The exact reconstruction using the same dictionary induces:

$$\mathbf{x} = \sum_{i=1}^M \langle \mathbf{y}, \mathbf{w}_i \rangle \mathbf{w}_i \quad \forall \mathbf{y} \in \mathbb{R}^N. \quad (2.9)$$

Since the dictionary is non-orthogonal, there exist p and $q \in \{1, \dots, M\}$ that verify $\langle \mathbf{w}_p, \mathbf{w}_q \rangle \neq 0$. From Eq. 2.9 we have:

$$\mathbf{w}_p = \sum_{i=1}^M \langle \mathbf{w}_p, \mathbf{w}_i \rangle \mathbf{w}_i \quad (2.10)$$

Then:

$$(1 - \|\mathbf{w}_p\|^2) \mathbf{w}_p = \sum_{i=1, i \neq p}^M \langle \mathbf{w}_p, \mathbf{w}_i \rangle \mathbf{w}_i \quad (2.11)$$

Hence:

1. if $\|\mathbf{w}_p\|^2 = 1$: As $\langle \mathbf{w}_p, \mathbf{w}_q \rangle \neq 0$, there exists a sum of non-zero vectors of the dictionary $(\mathbf{w}_i)_{i \in \{1, \dots, M\}}$ which is null.
2. if $\|\mathbf{w}_p\|^2 \neq 1$: \mathbf{w}_p is a linear combination of $(\mathbf{w}_i)_{i \neq p}$.

In both cases the vectors $(\mathbf{w}_i)_{i \in \{1, \dots, M\}}$ are not linearly independent. The dictionary is at least complete so that the exact reconstruction property is verified. Hence, the dictionary is overcomplete.

Moreover to preserve the smooth shape of the filters, M must be several times larger than N . Here a higher downsampling would distort the filters or induce aliasing effects. The above mentioned overcomplete factor of 40 is then necessary. It is also consistent with the redundant number of simple cortical cells compared with the number of photoreceptor neurons [87] which can be evaluated around 25.

The lack of aliasing guaranties the *shiftability in space* (or *translation invariance*) [182] which is of special importance for image processing tasks such as image denoising, image fusion, edge extraction, etc: the transform coefficients do not change drastically for a small displacement in space.

2.2.8 Border strategy

For a better response in the regions close to the border of the image, it is advantageous adding around the image a few extra rows and columns built as a smoothed version of the closest regions of the image. This would avoid the appearance of high-amplitude coefficients due to the sharp luminance transitions from one side of the image to the other one. The number of pixels added can be e.g. $2n_s$ in each side of the image.

2.2.9 Methods for reconstruction improvement and exact reconstruction

The condition of exact reconstruction comes from Eq. 2.3 and Eq. 2.6 as:

$$\sum_{r \in \Omega} |\mathbf{G}_r|^2 = 1 \quad (2.12)$$

Lets defined $\mathbf{S}(u, v) = \sum_{r \in \Omega} |\mathbf{G}_r|^2(u, v)$. In the sequel we propose different methods for improving the reconstruction, i.e. making \mathbf{S} closer to $\mathbf{1}$.

Hexagonal Fourier lattice by shifting even scales in orientation

As already included in Eq. 2.1 in the definition of $\theta_{(s,t)}$, every second scale is shifted in orientation by $\frac{1}{2}\frac{\pi}{n_t}$ angle. This shift, originally proposed in [74], induces a hexagonal configuration which provides a more uniform coverage of the Fourier domain (see also Fig 2.3.a).

Bandwidth adjustment

The bandwidths defined by Eq. 2.1 have been empirically adjusted. They allow a fairly uniform coverage of the Fourier domain, the variations of \mathbf{S} measured as $2(\mathbf{S}_{max} - \mathbf{S}_{min})/(\mathbf{S}_{max} + \mathbf{S}_{min})$ being lower than 0.6% between the 2nd and 5th scale.

Exact reconstruction

Each filter coefficient $\mathbf{G}_r(u, v)$ is finally normalized by $\sqrt{\mathbf{S}(u, v)}$, thus from Eq. 2.12 we have exact reconstruction filters. Because \mathbf{S} is already very close to the flat response before the division, the deformation introduced is small and does not produce appreciable distortions in the spatial domain (in particular side lobes are sufficiently small in amplitude to not be perceptible). We have now $\forall \mathbf{x} \in \mathbb{R}^N$, $W^T W \mathbf{x} = \mathbf{x}$. Note nevertheless that, due to the overcompleteness, in general $W W^T \mathbf{h}$ is not equal to \mathbf{h} : $\exists \mathbf{h} \in \mathbb{R}^M$, $W W^T \mathbf{h} \neq \mathbf{h}$.

It is then straightforward to show that W is a *tight frame* [117]. We have from Eq. 2.3 (where x are the 2D spatial coordinates and v the 2D frequency coordinates):

$$\sum_x |\mathbf{h}_r(x)|^2 = \sum_x |\mathcal{F}^{-1}(\mathbf{G}_r \cdot \mathcal{F}(\mathbf{x}))|^2 \quad (2.13)$$

By applying the Parseval theorem on the second member of Eq. 2.13:

$$\sum_x |\mathbf{h}_r(x)|^2 = \sum_v |[\mathbf{G}_r \cdot \mathcal{F}(\mathbf{x})](v)|^2 = \sum_v |\mathbf{G}_r(v)|^2 \cdot |[\mathcal{F}(\mathbf{x})](v)|^2 \quad (2.14)$$

Summing over all the filters r we obtain through the Eq. 2.12 and the Parseval theorem:

$$\sum_r \sum_x |\mathbf{h}_r(x)|^2 = \sum_v |[\mathcal{F}(\mathbf{x})](v)|^2 \cdot \sum_r |\mathbf{G}_r(v)|^2 = \sum_v |[\mathcal{F}(\mathbf{x})](v)|^2 = \|\mathbf{x}\|_2^2 \quad (2.15)$$

where $\|\mathbf{x}\|_2$ is the norm-2 of the vector \mathbf{x} . Moreover using the matrix notation, $\sum_{r \in \Omega} \sum_x |\mathbf{h}_r(x)|^2 = \sum_{k=1}^M | \langle w_k, \mathbf{x} \rangle |^2$.

This demonstrates that the dictionary of functions $(w_k) \in \{1, \dots, M\}$ forms a tight frame, since for any $\mathbf{x} \in \mathbb{R}^N$,

$$\sum_{k=1}^M |\langle \mathbf{x}, w_k \rangle|^2 = \|\mathbf{x}\|_2^2 \quad (2.16)$$

2.3 Application to image denoising

The present log-Gabor wavelet scheme can be used efficiently in many classical image processing applications such as edge extraction (see Chapter 3 and [63, 64]), image fusion [170], image denoising (see Chapter 4 and [63, 169]), texture analysis [31] and also image compression (see Chapter 4 and [60, 62, 64]). Here to illustrate the efficiency of the transform we propose to compare its ability to segregate the noise from the signal in an image denoising scenario.

Most currently used denoising methods are based on anisotropic diffusion [203, 188, 78, 70, 113] or wavelet thresholding techniques [44, 28, 20, 160]. Wavelet or multiresolution image denoising applications usually proceed in three stages: first a transformation, then a thresholding operation and finally the inverse transform for reconstructing the image. The transform aims at describing the signal in a domain where image content (principally the contrast edges) has statistically different amplitude than the noise: edges induce high amplitude coefficients while spatially incoherent noise produces a low level of amplitude spread in all the coefficients. A basic thresholding permits then to segregate most of the signal from the noise. There exist many methods for determining the optimal threshold [44, 101, 224, 20] (see reviews in [197, 165]). Moreover more elaborated methods [21, 160] also use the context (i.e. they take into account the neighborhood of each coefficient instead of solely the coefficient in order to decide if it represents signal features or for noise).

Here we aim at comparing only the efficiency of the log-Gabor wavelet transform in comparison with other multiresolutions independently of the method for threshold determination. Therefore we will test the transform in a series of classical images (128×128 tiles will be used) where Gaussian noise has been added. A gradient descent technique is applied to find the best threshold, i.e. the one giving the highest PSNR (*Peak Signal-to-Noise Ratio*) given the original image. The PSNR is calculated in dB as $-20 \log_{10}(\sigma_e)$ where σ_e is the *Root Mean Square Error* (RMSE) between the original and the denoised image. Thus, the experiment finds the best result achievable if the threshold determination technique would be optimal. It measures then the ability of each transform to separate the signal from the noise by thresholding. Note that these

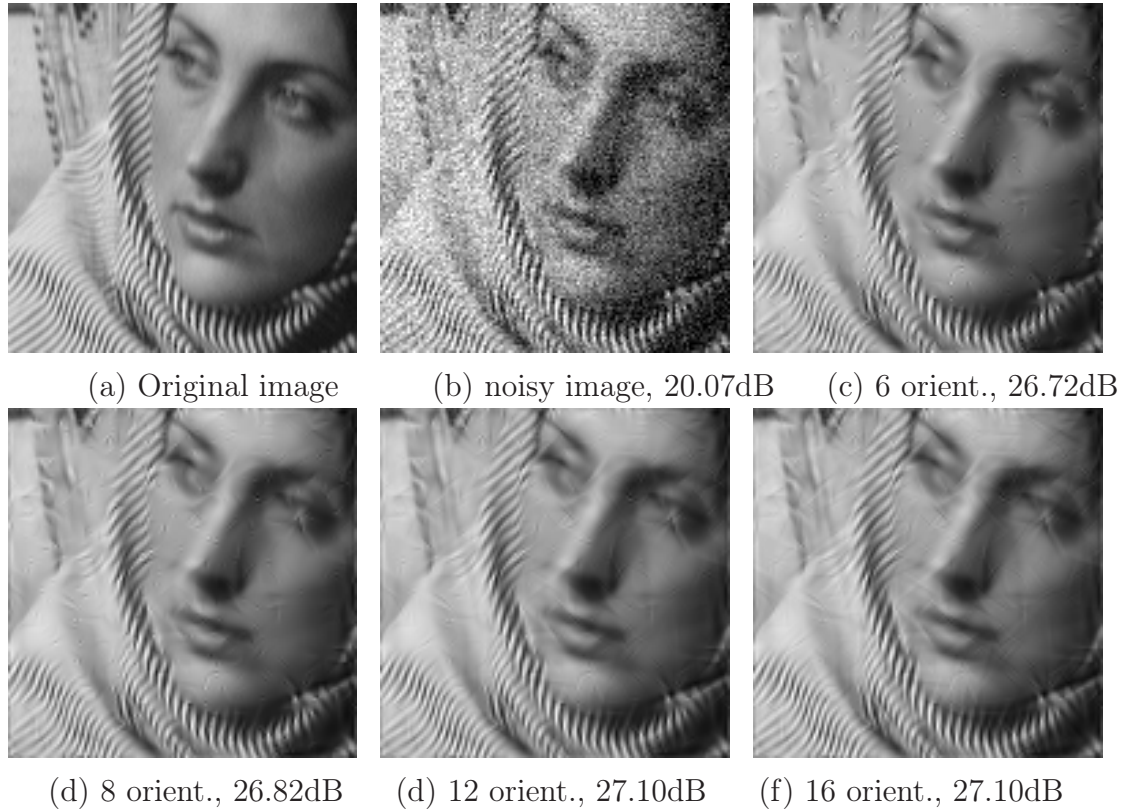


Figure 2.11: Denoising results through log-Gabor wavelets using different numbers of orientations. **b.** The noisy image has a PSNR equal to 20.07dB. **c.** In the case of denoising using 6 orientations a significant number of high amplitude noise points are still visible. The PSNR is equal to 26.72dB. **d.** Using 8 orientations, the number of noise points is reduced. The larger size of the filters could explain in part the better segregation between edges and noise. The PSNR improves to 26.82dB. **e.** Using 12 orientations the PSNR is 27.099dB. Almost all isolated noise points are removed but some elongated artifacts appear. **f.** Although the best PSNR result is obtained here for 16 orientations (PSNR=27.100dB), many salient elongated artifacts appear.

experiments are not only significant for image denoising but also for many other image processing tasks such as image compression or edge and feature extraction for which the possibility to code the image content through a reduced set of large coefficients is of primary importance.

For each experiment all the multiresolutions are implemented with five scales. A vector of five thresholds (one for each scale) is determined by gradient descent searching for the best PSNR to the original image. In those experiments no strategy has been set up for improving the response close to the borders of the image for any multiresolution. This allows to visualize the artifacts (e.g. ringing) due to the abrupt edges in those regions.

2.3.1 Number of orientations

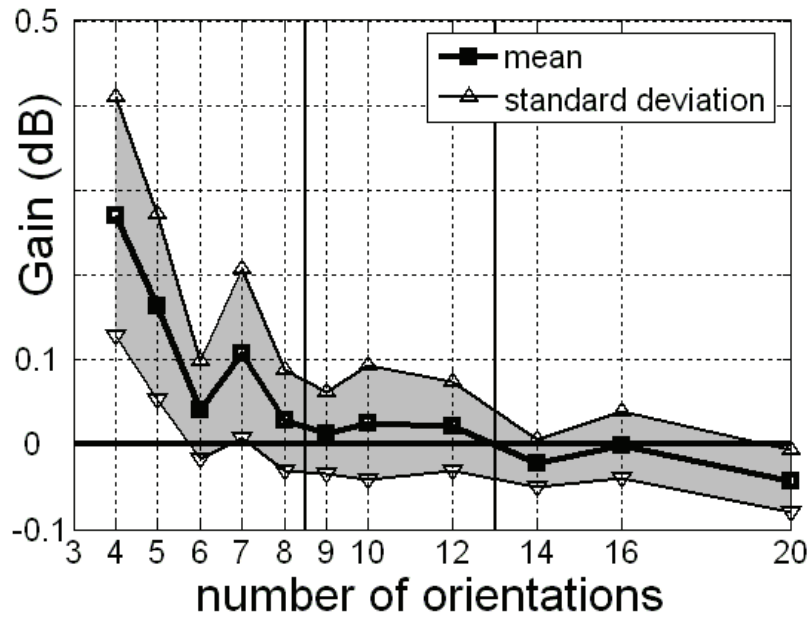


Figure 2.12: Denoising results as a function of the number of orientations. The denoising results are calculated over a set of 6 images (see the detailed results in Table 2.1), each of them tested with 3 different levels of noise (the PSNR is around 14, 20 and 28 dB). To reduce the influence of the large variability of conditions across experiments, the mean and standard deviation (in dB) are calculated taking as reference the previous number of orientations tested. Results show then how the denoising improves when the number of orientations is incremented from 3 to 4 orientations, from 4 to 5, from 5 to 6, etc. Results improve clearly from 3 to 8 orientations, they then achieve a maximum around 12 orientations and latter decrease slowly when more orientations are employed (see also the filter shapes as a function of the number of orientations in Fig. 2.4.a).

The first experiment aims at determining the optimal number of orientations to be used with log-Gabor wavelets. The question of the number of orientations is important since (bi-) orthogonal wavelets deploy just 3 orientations whereas more recent studies claimed the necessity of 6 orientations [173, 96], 8 orientations or more [160, 63], a variable number of orientations depending of the scale [41], or even up to 64 orientations [189].

Results gathered in Table 2.1 and Fig. 2.12 are calculated over a set of 6 images each of them tested with 3 levels of noise. They show that the denoising improves importantly between 3 and 8 orientations which confirms the necessity of using more orientations than (bi-) orthogonal wavelets do. The better results are obtained around 8 to 12 orientations but with variability, since in particular cases the best results are obtained with as few as 5 or 6 orientations or as many as 16 or 20 orientations. Visual inspection of Fig. 2.11 allows to see first that the number of remaining noise points decreases when the number of orientations increases particularly from 6 to 8 orientations and from 8 to 12 orientations. And secondly that 12 or 16 orientations yield elongated artifacts. Such artifacts do not appear, or at least are much less salient when using 8 orientations or less. All those results support the choice of 8 orientations as a good compromise between a low mathematical error and a high perceptual quality. This choice is also consistent with biological models of simple cells described in Section 2.2.1. Moreover it will allow to fairly compare with the steerable pyramids used in [160], since they also implement 8 orientations.

Note also that in the present scheme (as in other studies, see e.g. [189, 41]) the bandwidth varies depending on the number of orientations. Nevertheless it would be interesting to test both the number of orientations and the bandwidth parameters separately.

2.3.2 Comparison between different multiresolution schemes

In Table 2.2 log-Gabor wavelets with 8 orientations are compared with orthogonal wavelets 'Db4' [44], undecimated wavelets 'Db4' [28, 20, 111], and steerable pyramids with 8 orientations [160, 182] (see the shape of the filters Fig. 2.3 and Fig. 2.4). *Soft thresholding* is applied in the first two methods. Given a threshold, it consists in the diminishing (or *shrinkage*) of all coefficients by the threshold value (and consequently it zeroes out all the coefficients which amplitude lies below the threshold). Table. 2.3 shows that for almost all the experiments soft thresholding provides better results than *hard thresholding* (i.e. zeroing out the coefficients lying below the threshold and maintaining unchanged the coefficients exceeding it) for critically sampled wavelets and also,

Table 2.1: Log-Gabor denoising results are given as PSNR (in dB) in function of the number of orientations deployed. In the last row, mean and standard deviations are given for each increase of the number of orientations (see also Fig. 2.12). For each experiment the best results are indicated through boldface characters. The original images can be seen in Fig. 2.11.a, Fig. 2.15.a, Fig. 2.16.a, Fig. 2.17.a, Fig. 2.14.a, Fig. 2.13.a.

Image		Number of orientations											
PSNR	3	4	5	6	7	8	9	10	12	14	16	20	
barbara	28.17	31.38	32.01	32.34	32.45	32.71	32.73	32.82	32.79	32.79	32.76	32.75	32.67
	20.07	25.58	25.96	26.36	26.42	26.78	26.82	26.94	26.95	27.1	27.08	27.1	27.05
	14.25	22.03	22.23	22.47	22.51	22.75	22.77	22.71	22.99	22.97	22.98	22.92	
teapot	28.12	31.48	31.75	31.82	31.79	31.83	31.77	31.75	31.69	31.63	31.56	31.5	31.41
	20.19	25.87	25.9	25.94	25.87	25.88	25.82	25.83	25.78	25.73	25.69	25.65	25.59
	14.16	21.77	21.78	21.81	21.79	21.79	21.77	21.75	21.72	21.7	21.67	21.64	21.6
mandrill	28.11	28.67	28.89	29.02	29.02	29.09	29.11	29.14	29.13	29.18	29.18	29.2	29.21
	20.16	22.7	22.97	23.12	23.17	23.24	23.35	23.33	23.43	23.5	23.51	23.54	23.55
	14.2	19.51	19.8	19.83	19.99	19.94	20.12	20.05	20.19	20.25	20.16	20.28	20.3
house	28.09	32.35	32.66	32.68	32.77	32.79	32.9	32.89	32.96	33	33.01	33.03	33.05
	20.36	27.99	28.35	28.54	28.64	28.73	28.78	28.81	28.84	28.89	28.90	28.86	28.83
	14.08	24.85	25.01	25.31	25.37	25.47	25.46	25.47	25.55	25.53	25.52	25.51	25.44
lena	28.1	30.5	30.86	31.01	31.05	31.14	31.16	31.16	31.17	31.18	31.14	31.13	31.06
	20.23	25.39	25.67	25.83	25.9	26.02	26.06	26.11	26.1	26.13	26.13	26.11	26.03
	14.18	22.14	22.29	22.45	22.46	22.54	22.54	22.57	22.57	22.55	22.55	22.53	22.49
mit	28.1	31.44	31.83	32.1	32.16	32.31	32.34	32.38	32.38	32.37	32.34	32.31	32.23
	20.17	25.32	25.66	25.78	25.76	25.95	25.96	26.02	26.01	26.07	26.03	26.05	26.02
	14.1	21.22	21.42	21.56	21.6	21.66	21.7	21.69	21.7	21.69	21.69	21.67	21.63
Gain	mean	-	0.270	0.163	0.041	0.108	0.029	0.014	0.026	0.021	-0.022	-0.0004	-0.043
	dev.	-	±.141	±.110	±.058	±.101	±.059	±.049	±.067	±.053	±.028	±.040	±.037

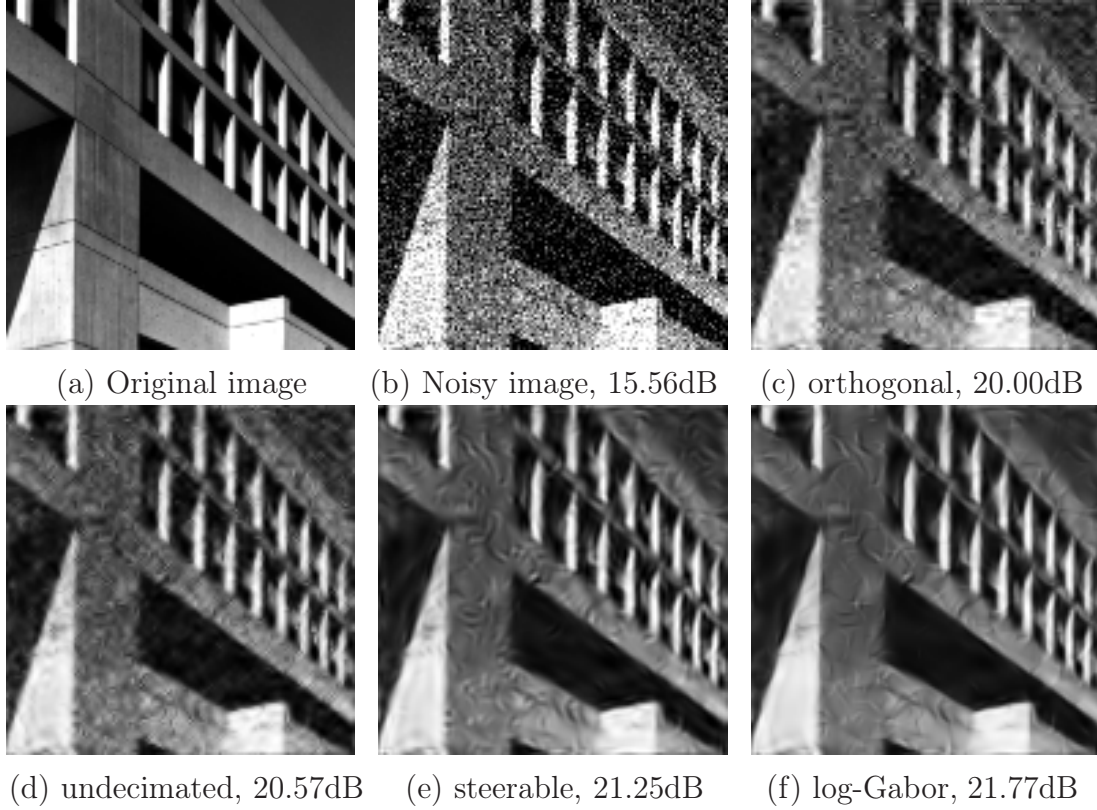


Figure 2.13: Denoising results. **a.** 128×128 pixel detail of the image "MIT". **b.** The noisy version has a PSNR of 15.56dB. **c.** Denoising by orthogonal wavelets provides a 20.00dB denoised image. **d.** Undecimated wavelets yield a 20.57dB denoised image. **e.** Steerable pyramid yields a 21.25dB denoised image. **f.** Log-Gabor wavelets achieve a 21.77dB denoised image.

although in a lower amount, for undecimated wavelets. This result is consistent with previous studies, see e.g. [44]. On the contrary hard thresholding yields better results for steerable pyramids and log-Gabor wavelets. Hard thresholding is then used for both those methods. For log-Gabor wavelets the threshold is applied in the modulus of the complex coefficients with exception of the first scale where it is applied separately on the real and imaginary parts of the coefficients because, as exposed in Section 2.2.2, they do not correspond exactly to the same position. It is not totally surprising that hard thresholding performs better here than soft thresholding since it preserves better the high amplitude coefficients which are the ones representing the image features. On the contrary the fact that soft thresholding provides better results for orthogonal wavelets could be due to the lower degree of matching between decomposition functions and image features. This would induce that each image feature is represented by a sum of transform coefficients having very different amplitude and being consequently less adequate for the segregation by hard thresholding. The lack of translation invariance would also enhance this effect (this would explain that soft thresholding does not improve so

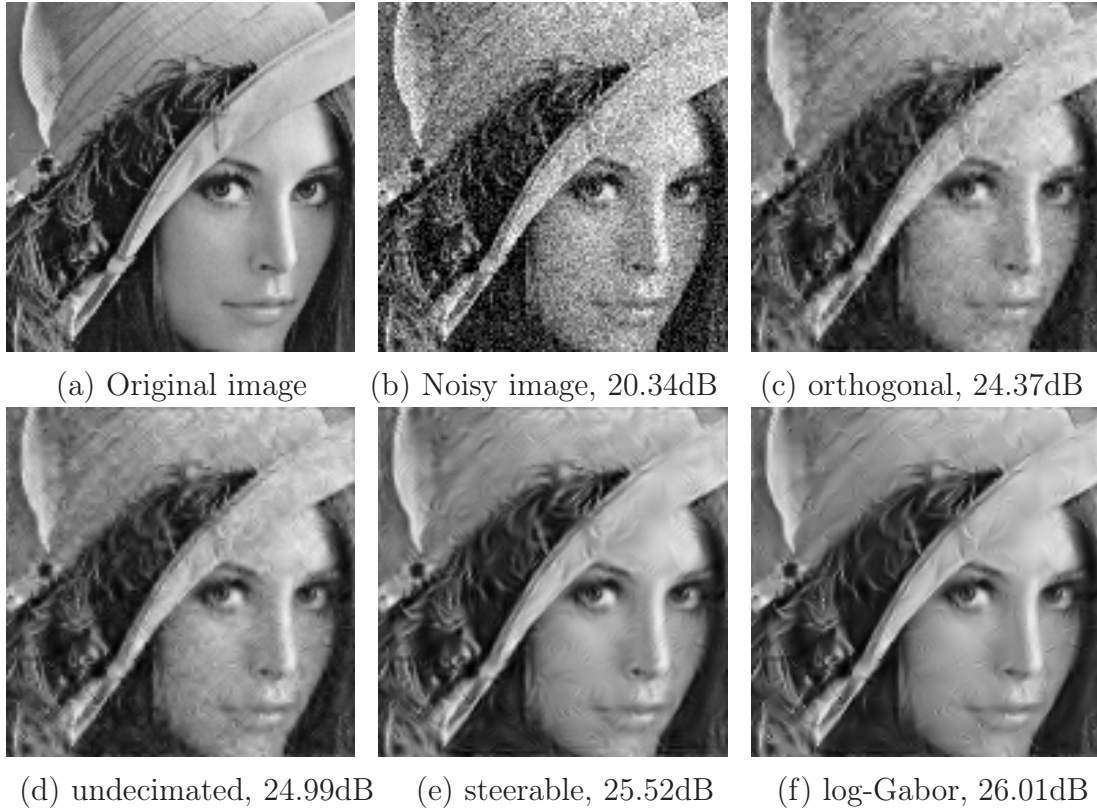


Figure 2.14: Denoising results using different multiresolution schemes. **a.** 128×128 detail of the image "Lena". **b.** The noisy version has a PSNR equal to 20.34dB. **c.** Denoising by wavelets provides a 24.37dB denoised image. **d.** Undecimated wavelets yield a 24.99dB denoised image. **e.** Steerable pyramid yields a 25.52dB denoised image. **f.** Log-Gabor achieves a 26.01dB denoised image. (see also in Table 2.2 for the entire result set).

much the results of undecimated wavelets).

It appears from Table 2.2 that apart from very few exceptions, undecimated wavelets always provide better results than critically sampled wavelets ($+0.51$ dB of improvement in mean, the standard deviation being $\sigma=0.18$ dB). Steerable pyramids provide better results than undecimated wavelets ($+0.40 \pm 0.36$ dB of improvement) and log-Gabor wavelets outperform all the three former methods ($+0.38 \pm 0.26$ dB of improvement compared with steerable pyramids. Improvements in terms of PSNR are observed in each of the 30 cases tested). Visual inspection on the Fig. 2.13, 2.14, 2.15, 2.16 and 2.17 show that denoised images contain artifacts which shape is characteristic of the multiresolution employed. Indeed the shape of the artifacts corresponds to the shape of the decomposition functions shown Fig. 2.3 and Fig. 2.4. Residual features due to noise also remain, their quantity could be reduced by a more severe thresholding. Here the thresholds are adjusted for the best PSNR. The thresholding should be a bit

Table 2.2: Denoising results given as PSNR in dB. The mean μ and standard deviation σ are given in each case in comparison with the previous method. Best results are indicated by boldface characters.

Original image	noise level	orthogonal soft	undecimated soft	steerable hard	log-Gabor hard
bar- bara	34.10	35.42	35.88	36.54	37.12
	28.11	30.36	30.99	32.16	32.57
	20.25	24.6	25.27	26.70	26.72
	14.74	21.08	21.64	22.69	22.92
	10.25	18.52	18.82	19.14	19.28
tea- pot	34.33	35.28	35.83	36.01	36.45
	28.41	30.12	30.90	31.24	31.63
	20.71	24.44	25.28	25.62	25.79
	15.06	20.75	21.45	21.69	21.77
	10.41	17.48	17.91	17.96	18.06
man- drill	34.12	34.29	34.35	34.30	34.35
	28.09	28.65	28.83	28.96	29.06
	20.3	22.24	22.77	23.26	23.42
	14.75	18.64	19.25	19.73	20.01
	10.28	16.22	16.67	16.98	17.16
house	34.15	35.66	36.13	36.09	36.52
	28.2	31.31	32.02	32.36	32.94
	20.38	26.86	27.59	28.28	29.10
	14.78	23.58	24.09	24.83	25.37
	10.2	19.94	20.27	20.72	20.89
lena	33.97	34.83	35.10	35.01	35.45
	28.15	29.79	30.23	30.54	31.11
	20.34	24.37	24.99	25.52	26.01
	14.89	21.27	21.80	22.16	22.59
	10.26	18.32	18.67	18.78	18.94
mit	34.82	35.73	36.12	36.06	36.94
	28.79	30.51	31.10	31.22	32.29
	20.98	24.10	24.85	25.29	25.95
	15.56	20.00	20.57	21.25	21.77
	10.64	16.21	16.56	16.86	17.10
mean improvement ($\mu \pm \sigma$)	34	+0.95 \pm 0.42	+0.37 \pm 0.16	+0.10 \pm 0.26	+0.47 \pm 0.25
	28	+1.83 \pm 0.76	+0.55 \pm 0.20	+0.40 \pm 0.35	+0.52 \pm 0.29
	20	+3.94 \pm 1.38	+0.69 \pm 0.096	+0.65 \pm 0.36	+0.39 \pm 0.29
	15	+5.92 \pm 1.58	+0.58 \pm 0.062	+0.59 \pm 0.27	+0.35 \pm 0.16
	10	+7.44 \pm 1.43	+0.37 \pm 0.053	+0.26 \pm 0.14	+0.16 \pm 0.044
	all	+4.02 \pm 2.71	+0.511 \pm 0.180	+0.402 \pm 0.355	+0.378 \pm 0.259

Table 2.3: Comparison between hard and soft thresholding. Denoising results are given as PSNR in dB and in last row as mean and standard variation of the Gain obtained by the use of soft thresholding instead of hard thresholding.

image	noise level	orthogonal wavelets		undecimated wavelets		steerable pyramids		log-Gabor wavelets	
		hard	soft	hard	soft	hard	soft	hard	soft
bar- bara	34.1	34.55	35.42	35.62	35.88	36.54	36.08	37.12	36.31
	28.11	29.33	30.36	30.95	30.99	32.16	31.37	32.57	31.45
	20.25	23.7	24.6	25.43	25.27	26.7	25.83	26.72	25.68
	14.74	20.22	21.08	21.52	21.64	22.69	22.18	22.92	22.15
	10.25	18.35	18.52	18.73	18.82	19.14	19.03	19.28	19.09
tea- pot	34.33	34.47	35.28	35.3	35.83	36.01	35.84	36.45	36.08
	28.41	29.07	30.12	30.42	30.9	31.24	31	31.63	31.13
	20.71	23.57	24.44	25.04	25.28	25.62	25.43	25.79	25.46
	15.06	20	20.75	21.28	21.45	21.69	21.65	21.77	21.66
	10.41	17.21	17.48	17.83	17.91	17.96	18	18.06	18.02
man- drill	34.12	34.12	34.29	34.12	34.35	34.3	34.44	34.35	34.45
	28.09	28.09	28.65	28.15	28.83	28.96	29.05	29.06	29.07
	20.3	20.77	22.24	22.11	22.77	23.26	23.18	23.42	23.21
	14.75	17.62	18.64	18.89	19.25	19.73	19.65	20.01	19.73
	10.28	15.61	16.22	16.54	16.67	16.98	16.89	17.16	16.94
house	34.15	34.38	35.66	35.4	36.13	36.09	36.15	36.52	36.37
	28.2	30.09	31.31	31.57	32.02	32.36	32.2	32.94	32.56
	20.38	26.1	26.86	27.74	27.59	28.28	27.84	29.1	28.24
	14.78	23.02	23.58	24.39	24.09	24.83	24.3	25.37	24.53
	10.2	19.69	19.94	20.5	20.27	20.72	20.41	20.89	20.43
lena	33.97	33.97	34.83	34.42	35.1	35.01	35.08	35.45	35.37
	28.15	28.52	29.79	29.73	30.23	30.54	30.32	31.11	30.66
	20.34	23.16	24.37	24.62	24.99	25.52	25.24	26.01	25.48
	14.89	20.56	21.27	21.69	21.8	22.16	22	22.59	22.2
	10.26	18.03	18.32	18.61	18.67	18.78	18.74	18.94	18.8
mit	34.82	35	35.73	35.85	36.12	36.06	35.79	36.94	36.45
	28.79	29.65	30.51	30.91	31.1	31.22	30.79	32.29	31.43
	20.98	23.32	24.1	24.98	24.85	25.29	24.8	25.95	25.12
	15.56	19.62	20	20.68	20.57	21.25	20.77	21.77	20.95
	10.64	15.98	16.21	16.73	16.56	16.86	16.68	17.1	16.73
Gain	-	+0.76	± 0.35	+0.21	± 0.29	-0.24	± 0.25	-0.45	± 0.32

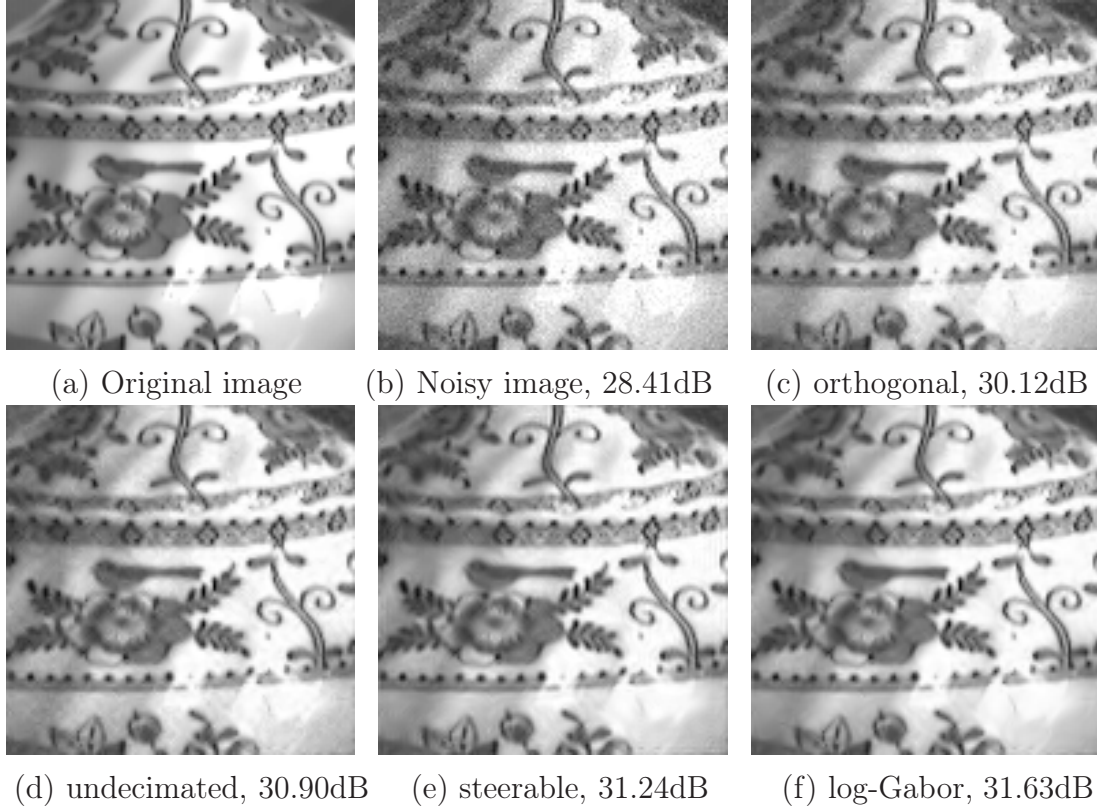


Figure 2.15: Denoising results on a 128×128 detail of the image "teapot".

higher for avoiding most of the residual noise at the cost of losing an additional part of the signal. This would be perceptually more pleasant even if the PSNR would be worst. It is remarkable that the log-Gabor method generally induces less amount of artifacts (particularly when compared with orthogonal and undecimated wavelets but also in comparison with steerable pyramids) which would indicate an augmented statistical separation in the transform domain between edge and noise features. In addition the artifacts can appear more natural looking so that they can result less annoying and be more easily perceived as image features (specially for images without human features, see Fig. 2.13, Fig. 2.15, Fig. 2.16 and Fig. 2.17).

It has already been established that the translation invariance property improves the results of overcomplete representations in comparison to critically sampled ones [182, 28, 20, 111], also because of the aliasing present in critically sampled wavelets. The better resolution in orientation can explain the better results of steerable pyramids and log-Gabor wavelets on undecimated wavelets. Note moreover that the finer bandwidth in orientation yields larger filters in the space domain, consequently it involves more pixels in the calculation of the coefficients: the noise is averaged on more pixels what leads to a better noise robustness (Fig. 2.4.a shows how the filter size increases with the number of orientations).

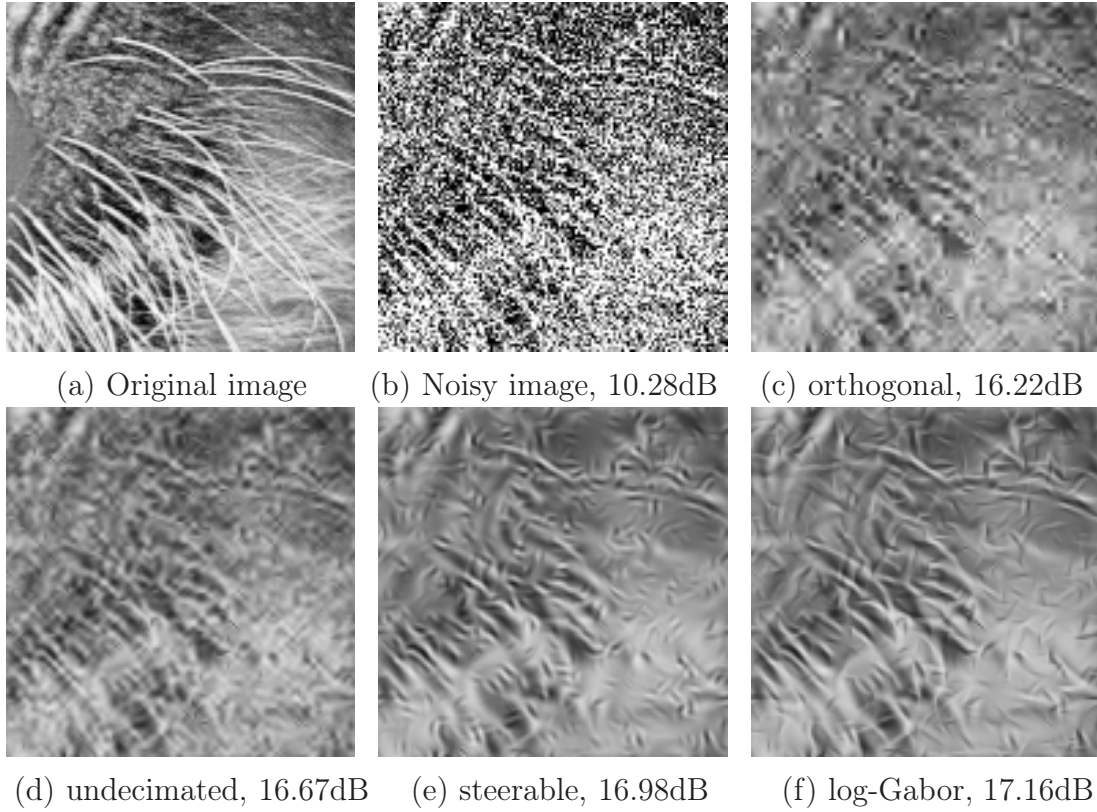


Figure 2.16: Denoising results on a 128×128 detail of the image "mandrill".

The better performances of the log-Gabor representation in comparison to steerable pyramids could be explained by four factors: (1) the Gabor (Gaussian) filter shape offers an optimal joint localization in frequency and space, which is an improvement in comparison with the raised cosine shape of steerable pyramids. (2) Log-Gabor functions are complex-valued with odd and even parts which permit a better capture of both edges and ridges [101, 63]. (3) The oriented high-pass log-Gabor filters are smooth and without extra side-lobes in space (while the high-pass steerable filters [160] shown in Fig. 2.4.c, bottom row seem to be constructed without the precautions described in section 2.2.2). (4) The proposed log-Gabor filters have elongated shape whereas steerable filters are more isotropic in size, i.e. log-Gabors have larger bandwidth in frequency (1.43 octave against 1 octave for steerable pyramids) and narrower bandwidth in orientation (37 degrees against 50 degrees for steerable). The elongated shape seems more appropriate for two reasons: they are adapted from biological data and they are also closer to the independent component of the natural images extracted by sparse coding or ICA techniques [141, 43]. Thus the log-Gabor functions should match better with edges of natural images (yielding a stronger statistical differential response between edge and noise features) and as an additional advantage they can appear more "natural looking" to human observers.

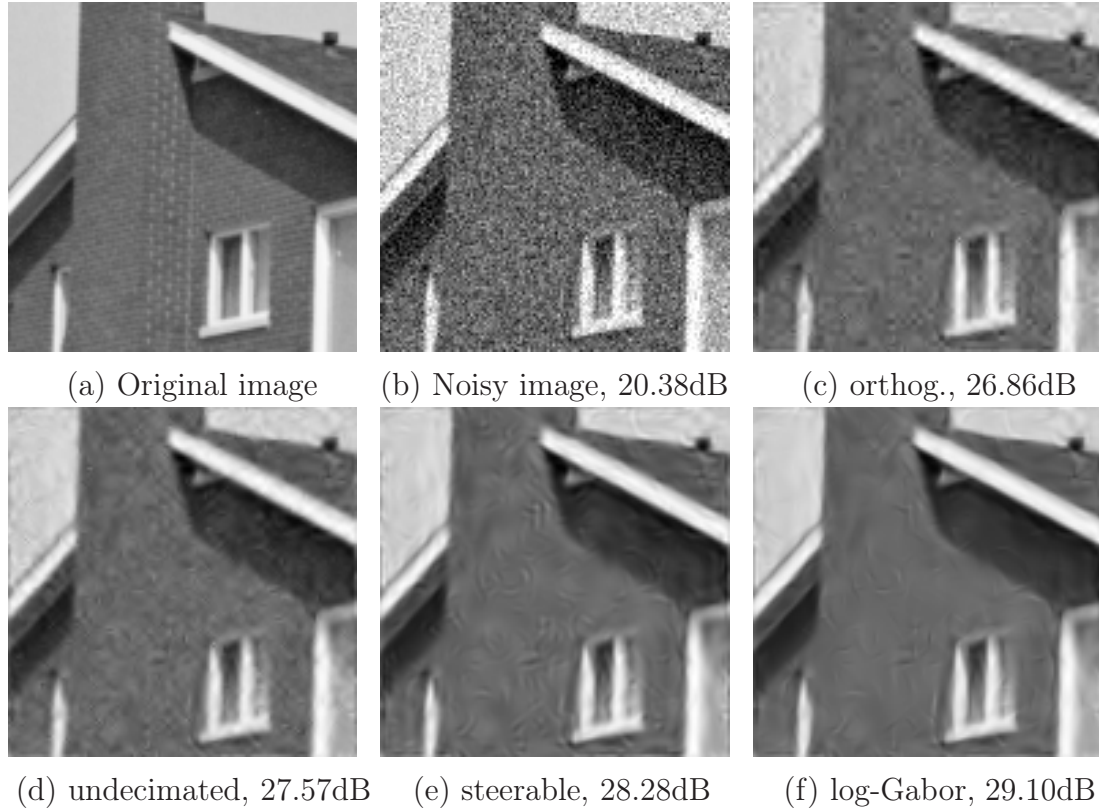


Figure 2.17: Denoising results on a 128×128 detail of the image "house".

2.4 Conclusions

We proposed an overcomplete multiresolution scheme mimicking the receptive field properties of simple cortical cells and optimized for achieving exact reconstruction through carefully designed filters. The proposed log-Gabor wavelet transform is optimized by taking into account mathematical, biological vision and image statistics considerations. From the mathematical point of view, Gabor functions are optimally joint-localized in frequency and space, which makes them optimal functions to characterize signals. The log-Gabor wavelet transform is moreover self-invertible which has been shown important for preventing the appearance of artifacts. Concerning biological vision criteria, the transform filters mimic closely the receptive field of V1 simple cortical cells. The proximity to biological vision can help simultaneously for choosing adequate transform parameters, for limiting the saliency of the artifacts and for further developing the biological models. And third, in relation with statistics of natural images, the transform filters have similar shape as the independent components learned from natural images, which support the proposed log-Gabor wavelets as an adequate scheme for matching natural image features.

The transform showed excellent results for segregating the signal from the noise by hard thresholding. The comparison with other methods (orthogonal wavelets, undecimated wavelets and steerable pyramids) showed an overall better performance of the proposed log-Gabor wavelet technique both in terms of the mathematical error and perceptual quality. The transform confirms then to be an adequate tool for representing features of natural images and to segregate them from incoherent noise.

Among further developments, a whole image denoising application could be set up by implementing a dedicated threshold determination method which took into account the relationships between neighboring coefficients as proposed e.g. in [21, 160]. Many other applications can also be derived since the ability of coding image features through a reduced set of large transform coefficients makes the transformation suitable for many applications such as image fusion or feature extraction. For image compression but also for general applications the redundancy induced by the overcompleteness could be recovered by deploying sparse coding algorithms (see Chapter 4).

Chapter 3

Edge extraction through energy models

"Toute la vie des sociétés dans lesquelles règnent les conditions modernes de production s'annonce comme un immense accumulation de spectacles. Tout ce qui était directement vécu s'est éloigné dans une représentation."

Guy Debord, *la société du spectacle*.

3.1 Introduction to edge extraction

First edge enhancement technics were developed early in the 1970. Prewitt [162] and Sobel [187], among others (see [155] for a review) proposed convolutions with small oriented kernels for edge enhancement. Later in the 1980 more complex models inspired from the visual system were based on Laplacian filters, i.e. on isotropic bandpass filters. Edges were then extracted as coincidence across scales of zero-crossing of the Laplacian responses [125, 80].

It was further demonstrated that oriented bandpass filters are more efficient than isotropic filters for edge extraction [16, 38]. In 1986 Canny [16] proposed a method based on criteria of good detection (i.e. low probability to miss existing edges), good localization and single detection (i.e. avoiding for an edge to be detected twice) to build

antisymmetric 2D filters for edge extraction. Canny's scheme implements additional mechanisms for adapting the filter shapes to the noise level of the image or to gather edges detected at different scales. It also avoids the appearance of gaps within continuous contours by using a hysteresis mechanism: the threshold of detection is dropped by a factor 2 to 3 if an edge was already detected in an adjacent pixel. Thanks also to the optimization of such important details, the Canny edge detector still constitutes nowadays a reference.

Energy models are based on the quadratic sum of two oriented bandpass filters in quadrature of phase [1, 133, 149]. Such non-linear combination of filters is directly inspired from biological models of complex cells of the primary visual cortex. It was shown particularly efficient for the simultaneous extraction of both *edges* (i.e. contrast changes in form of *steps* as in the example of the Fig 3.3.b), and *ridges* (i.e. dark or bright *lines* or "*roofs*", see Fig 3.3.a. The term *edge* is also used here to refer indistinctly to both the edges and ridges.) [175, 81, 102]. Later developments also deal with contour completion [75, 79, 112, 131, 221, 205]. Reviews on biologically inspired contour extraction methods can be found in [135, 109].

First energy models [133] proposed to extract edges similarly to the Canny method by *non local-maxima suppression*. This procedure consists in filtering through oriented kernels and in extracting only the coefficients which are local maxima across the direction perpendicular to the filter preferred orientation. This procedure is also called *competition* in [75, 131] or *lateral inhibition* in [221, 112]. A refinement incorporating inhibition between orientation filters was latter proposed [75]. This non-maxima suppression is itself biologically plausible since lateral interactions in V1 are also known to induce inhibiting effects between neighboring cells (see [94] and Section 1.4). A second known type of lateral interactions consisting in the *facilitation* between co-aligned and co-circular cells were also implemented in later energy models for improving the edge extraction [112, 221, 83, 75]. Those interactions are also called *enhancement*, *cooperation* or *excitatory horizontal connections*. The facilitation consists in convolving the oriented individual edges with a "*8-shape*" or "*association field*" inspired from psychophysics [53, 84] or on Gestalt psychology [201]. Although recent edge extraction energy models usually incorporate the same steps of oriented bandpass filtering, lateral inhibition and facilitation, different kind of implementation were proposed. Yen and Finkel proposed a facilitation based on the number of co-aligned edge segments [221]. Others simply deploy a convolution by the association field [75, 112]. Also, it is worth noting that some models deploy the facilitation before the inhibition [221, 201] while others implement first the inhibition [75, 131]. Another kind of interaction for edge extraction could be based on the phase congruency across scales as proposed in [102]. Those procedures of inhibition/facilitation where shown able to extract edges but also to segregate image features from spatially incoherent noise and to restore and enhance

image features [75, 131, 201].

Key-point features like crossing and angles were also shown important for improving the extraction. Their detection can itself be inspired from the V1 hyper-complex (or end-stopped) cells [83]. Such key-points are necessary e.g. for extracting anomalous contours due to occlusions [83, 129].

As a different biologically-inspired algorithm, Grigorescu et al. used a non-classical receptive field model incorporating inhibition from surround for segregating edges from texture [73]. Apart from energy models based on complex cells architecture, recent edge extraction technics like tensor voting [201] are based on the Gestalt psychology [99, 204]. Iverson and Zucker used logical operations to refine the linear filtering [90]. By its part the SUSAN operator is based on a similarity measurement inside a local window [186]. It can also be taken advantage of additional cues like texture [198] and color [126] to improve the edge extraction (see [98] for a review on color edge extraction).

In this chapter we propose in Section 3.2 an energy model adapted to log-Gabor wavelets. Results on edge/ridge extraction in comparison with the Canny method are presented in Section 3.3 and conclusions are given in Section 3.4.

3.2 Edge extraction using log-Gabor wavelets

We propose here an edge extraction method based on previous energy models and adapted to the log-Gabor wavelet scheme proposed in Chapter 2 (here the 6 orientations 5 scales scheme is employed. 6 orientations are used instead of 8 to limit the overcompleteness factor, nevertheless the 8 orientation scheme could also be used advantageously here). The proposed edge extractor incorporates three steps: first a hard-thresholding methodology for noise elimination (Section 3.2.1), second a lateral inhibition step for edge localization (Section 3.2.2) and third a lateral facilitation for evaluating the contour saliency and for eliminating residual noise or irrelevant edges (Sections 3.2.3 and 3.2.4). The inhibition/facilitation stage is summarized Fig. 3.1 and included within the whole visual cortex model in Fig. 3.2. Table 3.1 gathers the correspondences between the visual physiology and the image processing methods implemented for edge extraction and also for compression (Chapter 4).

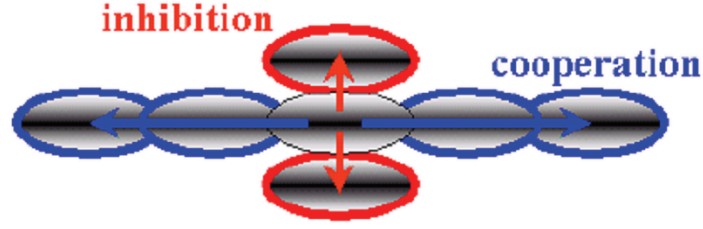


Figure 3.1: Inhibition and facilitation between Gabor wavelet coefficients. The inhibition zeroes coefficients which are not local-maxima in the direction perpendicular to the preferred orientation of the filter. It also occurs across orientations (see Section 3.2.2). The facilitation or cooperation takes place between coefficients co-aligned in the preferred orientation of the filter (Section 3.2.4) and also across scales between coefficients of the same or adjacent position and orientation (see Section 3.2.3).

3.2.1 Hard thresholding for noise elimination

The present contour extraction method is based on energy models [133, 82, 1]. Those models simulate complex-cells activities through the quadratic sum of simple cells in quadrature of phase according to the physiological knowledge [17, 82]. The complex-valued log-Gabor functions defined in Chapter 2 are employed here to model the simple cell activities, and their modulus model the complex cell activities (see also Table 3.1).

In the first step of thresholding, the complex cells whose activities do not reach a certain spike rectification threshold are considered as inactive. The Contrast Sensitivity Function (CSF) proposed in [176] is implemented here to model this thresholding. The CSF establishes the threshold of detection for each channel, i.e. the minimum amplitude for a coefficient to be visible for a human observer. All the non-perceptible coefficients are then zeroed out.

In presence of noise, the CSF is known to modify its response to filter down the highest frequencies (see [3] for a model of such behavior). This change in the CSF is modeled here by lowering the spike threshold depending on the noise level. The new threshold level is determined according to classical image processing methodologies for eliminating most of the noise: the noise level σ_c induced in each channel c is evaluated following the method proposed in [160] (if the noise variance in the source image not known it is evaluated as in [20]). The spike threshold is set up experimentally to $1.85\sigma_c$. This threshold allows to eliminate most of the noise apart from residual and relatively isolated noise features. Indeed, the threshold is fixed to a lower level than the optimal threshold so as to preserve a larger part of the signal while the residual noise is expected to be removed through the further processes of facilitation (Sections 3.2.3 and 3.2.4).

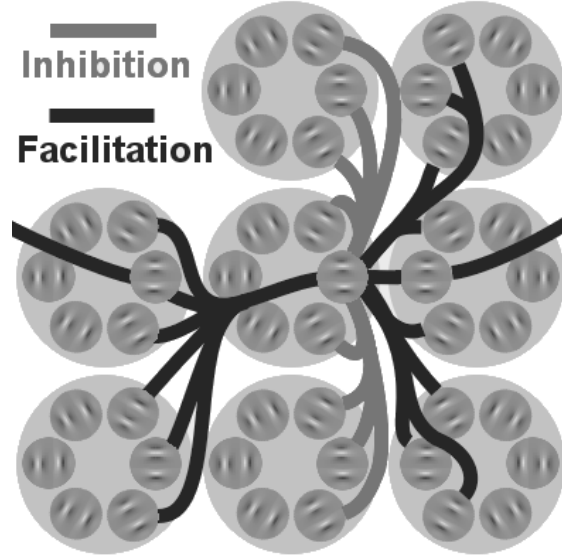


Figure 3.2: Schematic structure of the primary visual cortex as implemented in the present study. Simple cortical cells are modelled through log-Gabor functions (see Chapter 2). They are organized in pairs in quadrature of phase (dark-gray circles). For each position the set of different orientations compose a *pinwheel* (large light-gray circles). The retinotopic organization induces that adjacent spatial positions are arranged in adjacent cortical columns. Inhibition interactions occur towards the closest adjacent positions which are in the directions perpendicular to the cell preferred orientation and toward adjacent orientations (light-gray connections). Facilitation occurs towards co-aligned cells up to a large distance (dark-gray connections).

3.2.2 Oriented inhibition

The inhibition step also follows energy models [133, 82] which implement non-local maxima suppression between complex cells for extracting edges and lines. A very similar strategy is also deployed in the Canny operator [16] which mark edges at local maxima after the filtering through oriented kernels. As indicated by the light-gray connections in Fig. 3.2 the inhibition occurs toward the direction perpendicular to the edge, that is to the filter orientation. It inhibits (i.e. zeroes out) the closest adjacent orientations and positions which have lower activity (no inhibition across scales is implemented here).

It is to note that the shape of the filter is critical here for an accurately localized, single and noise-robust detection [16]. Fig. 3.3 illustrates that log-Gabor filters are adequate for extracting both edges and ridges by non-local maxima suppression: (1) both edges and ridges induce local-maxima in the modulus of the log-Gabor coefficients and (2) that the modulus monotonously decreases on both sides of edges and ridges without creating extra local-maxima (the modulus response is monomodal).

Table 3.1: Correspondences between visual cortex physiology and image processing operations defined in the different sections.

Visual cortex structures	Image processing	Section
Even-symmetric simple cell Odd-symmetric simple cell Pair of simple cells Pinwheel Retinotopic organization Complex cell	Real part of log-Gabor coef. Coefficient imaginary part Complex-valued log-Gabor coefficient Set of orientations for a position (x,y) arrangement of pinwheels Modulus of log-Gabor coef.	2.2
Spike threshold	CSF and denoising	3.2.1
Anisotropic inhibition	Edge extraction	3.2.2
Interscale facilitation	Interscale reinforcement	3.2.3
Spatial facilitation	Selection of salient chains	3.2.4
Set of spiking cells	Selected subdictionary	3.2.4
Gain control	Amplitude assignation	4.2.4
Hypercomplex cells Contour shape Contour representation	Chain heads Chain movements Chain coding	4.2.5

In practice the procedure is the following one: for each channel the neighbors are labeled from "1" for the neighbor in the direction closest to the perpendicular to "8" for the closest to the channel orientation (see also Fig. 3.4). Coefficients are inhibited (zeroed out) in two steps. First they are inhibited if any of their neighbors "1" and "2" have larger modulus. Remaining coefficients constitute then one- or two-pixel-thick chains. Coefficients constituting two-pixel-thick chains are extracted by convolution with 3×3 kernels for determining if some of the 4 closest neighbors are themselves neighbors one another (see the example shown in Fig. 3.4.d). One-pixel-thick chains being desirable, the inhibition goes on with the neighbors "3" but only on two-pixel-thick chains, then with neighbors "4" and so on with the neighbors "5" and "6", in each case only on the two-pixel-thick chains. At the end only one-pixel-thick chains remain. All those operations have low computational cost and consist purely in convolutions through 3×3 kernels.

If the inhibition is realized only inside each channel, edges and ridges still induce multiple responses in the channels of different orientation. Therefore the local maxima criteria should be applied also across orientations. Each coefficient is then inhibited by larger coefficients located not only in the same channel but also in the two channels of adjacent orientation in the same scale (see Fig. 3.5). The rules and the neighbors in the interchannel inhibition are the same as those of the intrachannel inhibition.

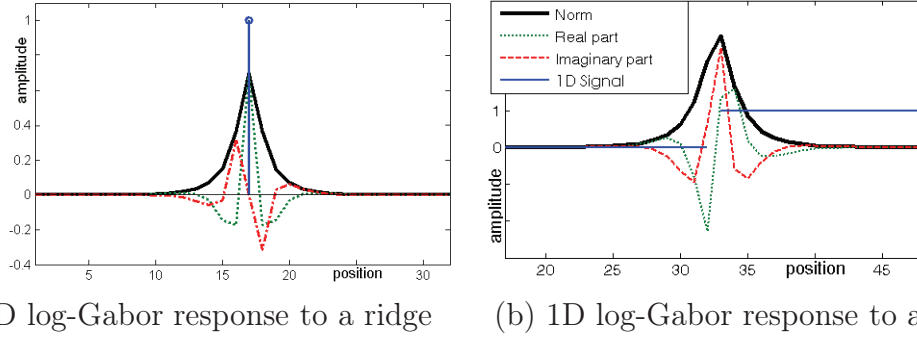


Figure 3.3: Edge extraction through non local-maxima suppression on complex-valued log-Gabor wavelets. **a.** Response of a 1D complex log-Gabor filter to an impulse (ridge): the modulus (black continuous curve) of the response monotonously decreases away from the impulse. It implies that the ridge is situated just on the local maximum of the response. On the contrary the real (dot) and imaginary (dash-dot) parts present various local-maxima and minima which make them less suitable for ridge localization. **b.** Same curves for a step edge. The same conclusions hold.

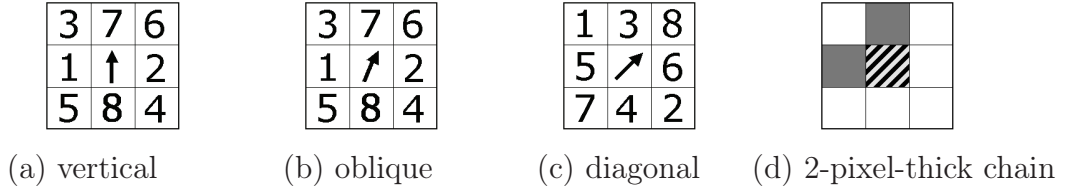


Figure 3.4: Implementation of the local inhibition. **a.-c.** For each channel orientation the neighboring coefficients are labeled from the most perpendicular to the most collinear to the preferred orientation of the filter. **d.** Configuration of two-pixel-thick chains where the central coefficient is candidate to be inhibited (see explanations in the text).

After inhibition is performed, most coefficients are set to zero and the remaining coefficients already show a strong similitude with the edges and ridges perceived by visual inspection (see Fig. 3.5 and Fig. 3.6.c). It is remarkable moreover that extracted contours appear continuous and mainly without gaps. Nevertheless some isolated coefficients still remain due to noise, irrelevant or less salient edges. Facilitation interactions between coefficients will allow to evaluate the saliency and reliability of edges and to remove residual noise.

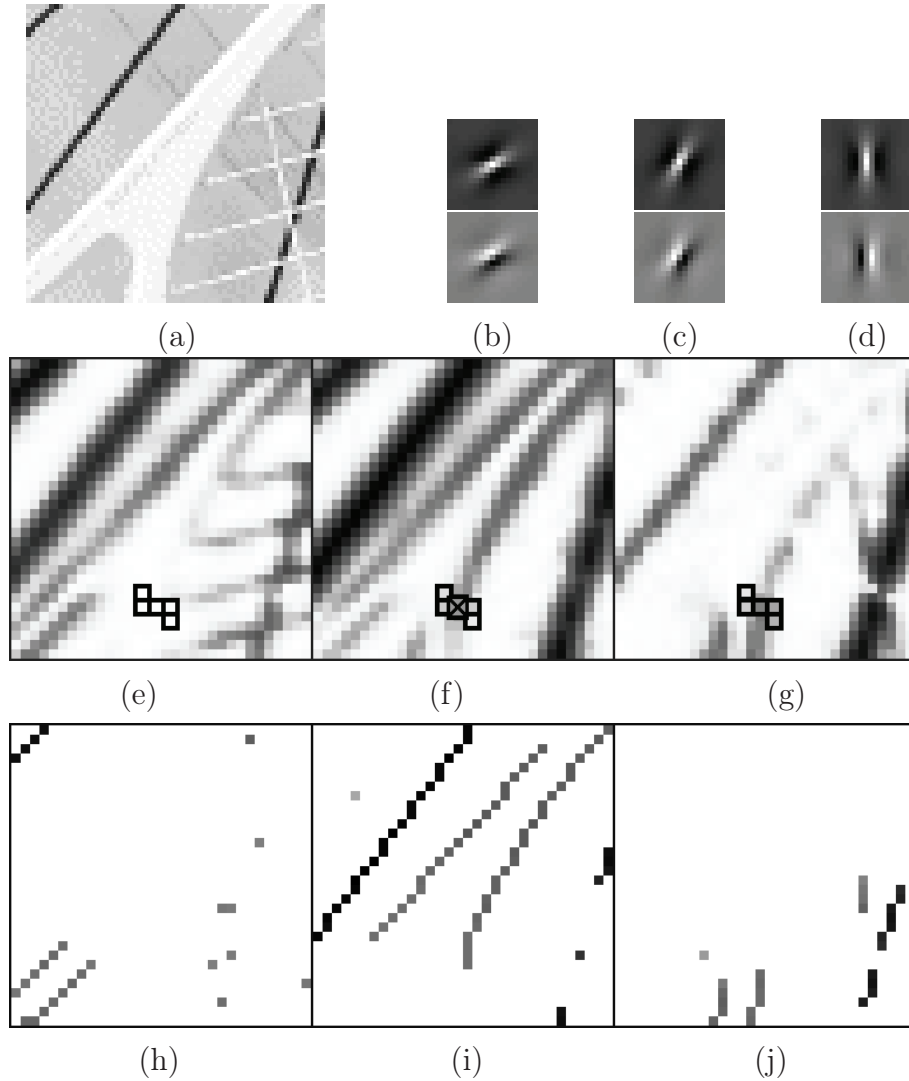


Figure 3.5: (e), (f) and (g) represent 3 third scale channels of adjacent orientation from the log-Gabor transform of the 64x64 tile of the image "bike" presented in inset (a). The real and imaginary components of the log-Gabor filters corresponding to the channels (e), (f) and (g) are shown respectively in (b), (c) and (d). In the competition any coefficient (e.g. the one marked with the cross-box symbol in (f)) is susceptible to be inhibited by its neighbors which are located in the same channel (f) or in adjacent channels (e) and (g). The inhibiting neighbors are situated in the directions perpendicular to the channel preferred orientation (thus in this case the inhibition occurs first in the horizontal and NW-diagonal directions). Those inhibiting neighbors are plotted within box-symbol. (h), (i) and (j) show the remaining coefficients in each channel after the competition. They confidently localize the contours appearing in the image (a) in a non-redundant way (i.e. no double detection occurs) and with almost no gap between extracted edge coefficients. Note that contours can skip across orientations.

3.2.3 Facilitation across scales

Facilitation interactions have been described in V1 as excitative connections between co-oriented, co-axial and aligned neighboring cells [94, 13]. Psychophysical studies and the Gestalt psychology determined that co-aligned or co-circular stimuli are more easily detected and more perceptually salient [84, 122]. Studies of natural image statistics also show that statistically edges tend to be co-aligned and co-circular [104, 69]. Experimentally we observe that log-Gabor coefficients arranged in chains of co-aligned coefficients or present across different scales correspond to reliable and salient edges. Moreover the probability that remaining noise features could be responsible for chains of coefficients is decreasing with the chain length. Thus a facilitation reinforcing co-circular cells conforms a noise segregation process. For all those reasons a facilitation across scale is set up to reinforce co-oriented cells across scales and a facilitation in space and orientation reinforce chains of co-aligned coefficients (Section 3.2.4).

The facilitation across scales consists here in favoring those coefficients located where there exist also non-inhibited coefficients at coarser scales. The *parent* coefficient (i.e. the one in the coarser scale) must be located in the same spatial location (tolerating a spatial deviation of one coefficient), in an adjacent orientation channel and be *compatible in phase* (i.e. it must have a difference in phase with the *child* coefficient lower than $2\pi/3$. See also [102] for an edge extraction method based on phase congruency across scales). It is remarkable (see Fig. 3.6.c) that many edges and ridges extracted are closely repeated across scales with coefficients linked by parent relationship. This regularity is of great importance and it could be further exploited e.g. for efficient coding.

3.2.4 Facilitation in space and across orientations

As proposed in Yen and Finkel's V1 model [221], we implement a saliency measurement linked with the *chain length* defined as the number of coefficients composing the chain. It is calculated for each coefficient and consists in counting the number of coefficients forward n_f and backward n_b along the chain. The *compatibility in phase* is also checked, that is two successive coefficients which do have a difference of phase superior to $2\pi/3$ are not considered to belong to the same chain. The adjacent coefficient must moreover be aligned in the direction of the channel tolerating a variation of $\arccos(0.6)$ (i.e. around $\pi/3.4$. Note that this choice is not casual but optimized for the rectangular grid). The present implementation makes possible for a chain to begin in one orientation channel and latter change of channel. The number of coefficients is counted in each direction to a maximum of l_{max} coefficients (with $l_{max} = 16$). The saliency is finally calculated as

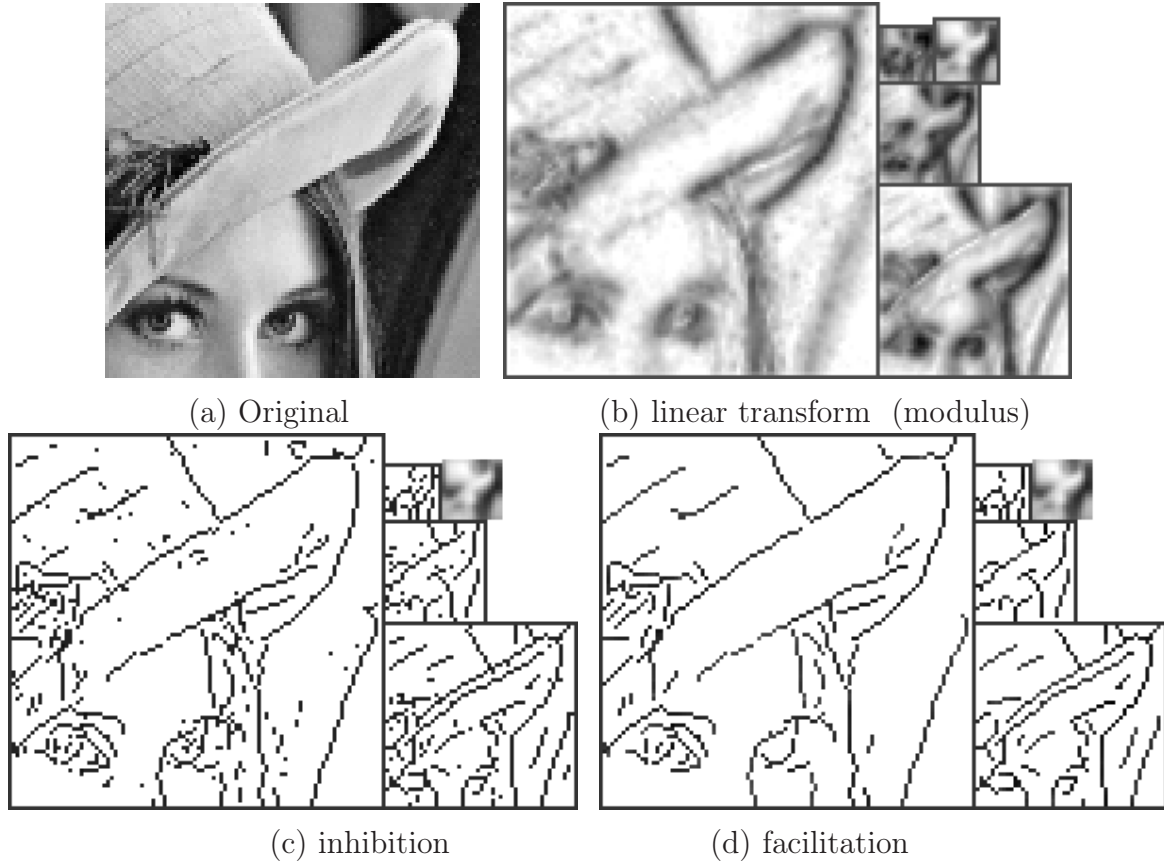


Figure 3.6: Successive steps for multiscale edge extraction. **a.** 96×96 detail of the "lena" image. **b.** The linear log-Gabor wavelet response is obtained from through the log-Gabor wavelet transform defined in Chapter 2. All the orientation channels are summed together so that one inset is shown for each scale. The first scale is not represented. The different scales have different sizes due to the downsampling applied. From the largest to the smallest the insets correspond respectively to the 2nd, 3rd, 4th, low-pass and 5th scales. **c.** Remaining coefficients after the inhibition step (Section 3.2.2). **d.** The facilitation steps (Sections 3.2.3 and 3.2.4) preserve only the coefficients arranged in sufficiently long chains or having parent coefficients within coarser scales.

$c = \min(n_f + n_b, l_{max})$, which permits to obtain constant responses along a same chain.

Since the log-Gabor wavelet transform and the inhibition step already provide chains arranged along contours with almost no gaps between coefficients, the facilitation in space is a very reliable measurement of the length of the contours. The measurement can then be used with the facilitation across scales to remove the short chains and isolated coefficients corresponding to noise or to less salient contours.

Finally the facilitation consists in retaining those edge coefficients which fulfill the

following two criteria (while the other coefficients are withdrawn to be considered as noise or less salient edges): First they must pass a certain length threshold depending of the scale and the presence of parent coefficients. Typically the chain length threshold is chosen as 16, 16, 8, 4, 2 respectively for the scales 1, 2, 3, 4, 5, half of these lengths if coefficients have a parent, and a fourth of these lengths are if they have a grandparent (i.e. the parent coefficient has itself a parent). Any coefficient which have a great-grandparent is selected. Second, the amplitude must overpass a spike threshold corresponding to twice the CSF threshold defined in Section 3.2.1. Each coefficient is selected with its chain neighbors which implies that chains are selected or rejected entirely. This avoids for a contour to be selected only by parts. It is equivalent to the hysteresis implemented in the Canny edge extractor [16]. The final selection can be seen in Fig. 3.6.d.

Both the lengths and CSF thresholds can be modified depending on the application since for compression (see Chapter 4) the thresholding must be severe while for image denoising (see Section 4.3.3) most of the edge information should be preserved which requires more permissive thresholds.

Edges of the first scale are less reliable because the orientation selectivity is necessarily lower at high frequencies since at Nyquist frequency both diagonals are indistinguishable. Another reason is that first scale filters are not totally in quadrature of phase, since there is a half-pixel shift in the imaginary part location (see Section 2.2.2). In the present implementation first scale edges are not considered. Alternatively, for image denoising (Section 4.3.3) or here for the reconstructions, edges selected in the second scale will also be used for the first scale.

3.3 Results on edge extraction

Examples of contours extracted after the inhibition/facilitation processes are shown in Fig. 3.6, Fig. 3.7 and Fig. 3.8. The different orientations are summed up so that edges belonging to each scale are represented together. As a comparison, Fig. 3.7.d and Fig. 3.8.d show the edges extracted by the Canny operator. Log-Gabor wavelet based edge extraction presents the following advantages: (1) it extracts both edges and ridges while Canny only extracts edges drawing generally two edges where there is one ridge (see in particular Fig 3.8.d). (2) As it will be shown in Chapter 4, it is possible to reconstruct the image through the edge information. This is a warranty of the nearly completeness of the extracted edges (see Fig. 3.7.c and Fig. 3.8.c for the reconstructions). Reconstruction quality will be discussed in the Chapter 4 both in

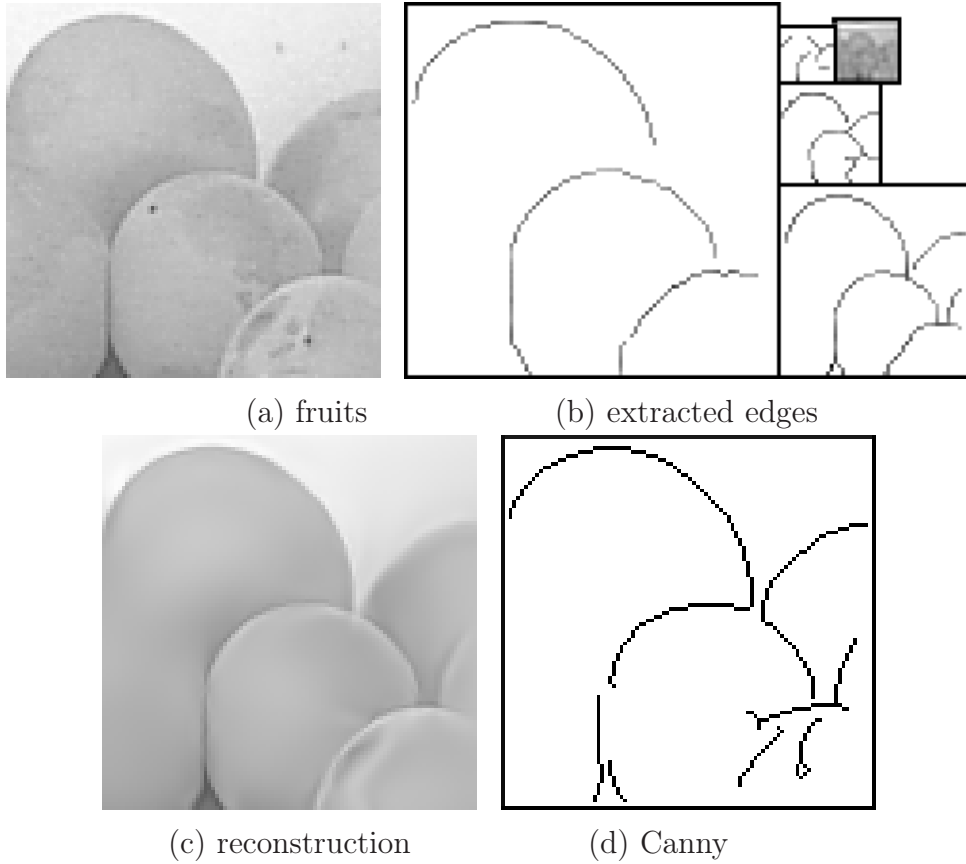


Figure 3.7: Edge extraction and reconstruction. **a.** A 96×96 pixels detail of the image "Fruits". **b.** Multiscale edges extracted by the proposed log-Gabor wavelet based energy model. **c.** Reconstruction from edges (see Chapter 4). **d.** Edges extracted by Canny method.

cases where few edges are selected (image compression, Sections 4.3.1 and 4.3.2) or when most of the edges are preserved (image denoising, Section 4.3.3). In spite of those advantages of the present scheme, the Canny operator still contains sophisticated mechanisms that were not implemented here yet, as e.g. the adaptative choice of the filter width or the multiscale integration of edges so that it could eventually in some cases offer more perceptually pleasant results.

In comparison with other energy models, the present implementation have the advantage to be implemented on a self-invertible wavelet scheme, what allows to extract edges at multiples scales and also to reconstruct from edges (see Chapter 4). As an important difference, it was not seen advantageous here to use the 8-shaped association field [135] for the facilitation mechanism, although experiments were done in this direction (see also Chapter 5). Instead the chain length saliency calculation is employed. The first reason is that edge coefficients obtained after the inhibition are organized in chains which rarely contain gaps. And because even if the 8-shape strategy could allow

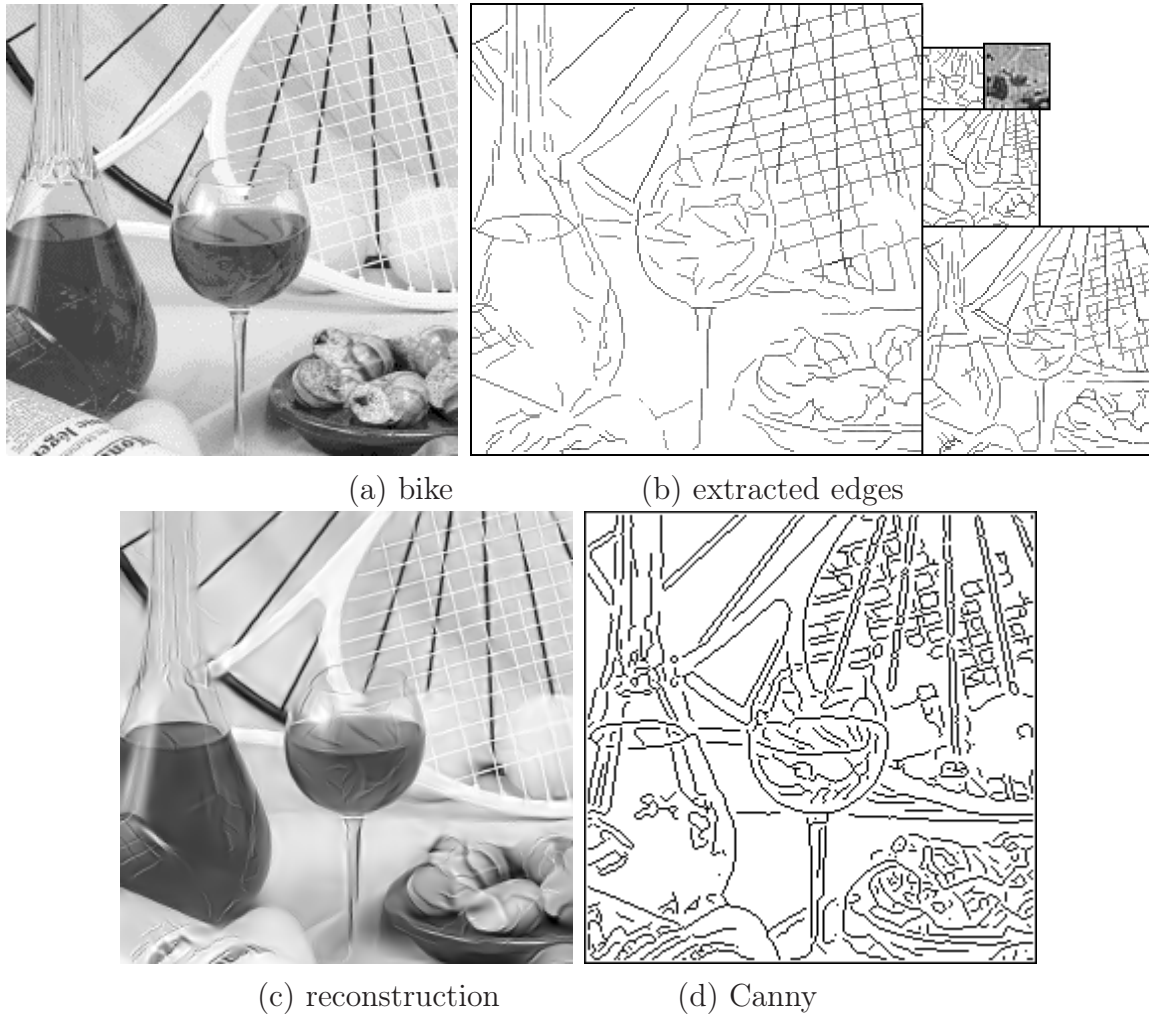


Figure 3.8: Edge extraction and reconstruction. **a.** A 224×224 pixel tile of the image "Bike". **b.** Multiscale edges extracted by the proposed log-Gabor wavelet based energy model. **c.** Reconstruction from edges (see Chapter 4). **d.** Edges extracted by Canny method.

to close the few existing gaps, it has been shown to induce many artifacts when applied to natural images, as e.g. the inadequate closure between non-related edge segments. We conclude then that the 8-shape facilitation should not be used as a general mechanism for edge extraction but rather be restricted to extract occluded or discontinuous contours. An implementation is proposed for such purpose in Chapter 5. As a physiological support to this hypothesis, it has been shown that occluded contours are not detected in the visual pathways before area V2 [212, 211]. This would suggest that inhibition/facilitation mechanisms taking place in V1 are not dedicated to reconstruct occluded or incomplete contours.

3.4 Conclusion

We proposed an edge extraction mechanism inspired from biological knowledge on V1 cortical cells in particular through (1) the complex cell model consisting in the quadratic sum between even and odd simple cells (corresponding to the real and imaginary parts of log-Gabor coefficients) and (2) the lateral interactions of inhibition and facilitation between neighboring complex cells. In the implementation it consists then in a self-invertible linear log-Gabor wavelet transform followed by non-linear steps of quadratic sum, hard thresholding for denoising and inhibition and facilitation between neighboring transform coefficients. Those steps are able to extract continuous chains of coefficients located on the edges and ridges of the image, achieving an efficient contour extraction. The resulting mechanism show advantages in comparison with the Canny method, since it simultaneously extract edges and ridges, and in comparison with other energy models thanks to the use of a wavelet framework allowing image reconstruction.

As a further improvement the coincidence across scales could be further exploited for a more robust edge extraction and for building an unified edge representation incorporating edges of any scales.

Chapter 4

Sparse approximation based on edge extraction

"L'américain type consacre, pour sa part, plus de 1500 heures par an à sa voiture : il y est assis, en marche ou à l'arrêt, il travaille pour la payer, pour acquitter l'essence, les pneus, les péages, l'assurance, les contraventions et les impôts. Il consacre donc quatres heures par jour à sa voiture, qu'il s'en serve, s'en occupe ou travaille pour elle. Et encore, ici ne sont pas prises en compte toutes ses activités orientées par le transport : le temps passé à l'hôpital, au tribunal ou au garage, le temps passé à gagner de l'argent pour voyager pendant les vacances, etc. A cet Américain, il faut donc 1500 heures pour faire 10000 kilomètres de route, 6 kilomètres lui prennent une heure."

Ivan Illich, *La convivialité*.

Several drawbacks of critically sampled wavelets can be solved by over-complete and sparse multiresolution transforms. Facing the difficulty to optimize such non-orthogonal and non-linear transforms, we implement in this chapter a sparse approximation scheme based on models of the primary visual cortex. The scheme consists in a linear log-Gabor wavelet transform followed by inhibition and facilitation steps between neighboring transform coefficients for extracting the edges and ridges of the image. The edge coefficients are shown sufficient for reliably reconstructing the images, while their coding permits efficient image compression. Additionally, the ability to segregate the edges from the noise yields promising image denoising results. Parts of this study were previously presented in [64, 63, 62, 60, 59, 56, 169]

for image compression and in [64, 61, 63, 169] for denoising applications.

4.1 Introduction

Recent works on multiresolution transforms showed the necessity of using overcomplete transformations to solve drawbacks of (bi-) orthogonal wavelets, namely their lack of shift invariance, the aliasing between subbands, their poor resolution in orientation and their insufficient match with image features [182, 189, 41, 96, 43]. Nevertheless the representations from linear overcomplete transforms are highly redundant and consequently inefficient for such tasks needing sparseness as e.g. image compression. Several *sparse coding* algorithms have been proposed to address this problem by approximating the images through a reduced set of decomposition functions chosen from an overcomplete set called *dictionary* [118, 22, 167, 153] (see reviews in [22, 147, 202]). Such methods are also referred perhaps more adequately as *sparse approximation* algorithms. Fast and optimal algorithms (e.g. [29]) have been described for particular cases. In the general case it is often not possible to build optimal algorithms. Two main classes of algorithms are then available: Matching Pursuit (MP) [118, 167, 151] which chooses one by one the highest coefficients in all the dictionary and Basis Pursuit (BP) [22] which solves the linear system minimizing a penalizing function corresponding to the sum of the amplitude of all coefficients. Both of these algorithms perform iteratively and globally through all the dictionary, they are computationally costly algorithms which only achieve approximations of the optimal solutions.

The objective to improve and optimize image transformations which are overcomplete, non-orthogonal and non-linear is extremely complicated by the huge degrees of freedom offered by the new conditions. Non-orthogonality allows the use of any shape of filters while non-linearity is the most general case and permits for instance the implementation of a wide variety of interactions between the transform coefficients. Entire new classes of algorithms are now potentially candidates for the multiresolution transform. We propose here to build a new sparse overcomplete transform designed not only through mathematical and classical image processing criteria but which would also be inspired from the known physiology of the Primary Visual Cortex (V1) of primates. The rationale behind the biological modeling is the plausibility that V1 could accomplish an efficient coding of the visual information: (1) V1 simple cells show a Gabor shape oriented receptive field [36] and they plausibly implement an oriented multiresolution Gabor-like transform (see Chapter 2). (2) V1 supposedly implements a sparse coding scheme [139]. (3) Non-linear interactions between V1 cells such as inhibitions between non-aligned neighboring cells and facilitation between co-aligned and collinear cells have

been described by physiological and psychophysical studies [94, 122, 84]. These interactions have been shown efficient for image processing in applications such as contour extraction and image restoration [131, 79, 83, 221, 208]. We make here the hypothesis that these lateral interactions deal not only with contour extraction or noise segregation but also allow to achieve a sparse representation of the visual information.

The sparse approximation method presented in the following is based on previous image processing works: first on the study of Mallat and Zhong [119] which proposed to code the images through chains of wavelet coefficients located on the image edges extracted through the Canny operator, and on Elder's work which showed the possibility of reconstructing the images from their edges [51]. By their part Desai et al. [39] used basic methods to extract edges and code them through chains for image compression at high compression rate. Those studies motivate the present sparse approximation strategy consisting in coding the image by uniquely the transform coefficients located on image edges and contours. Because such representation is not necessarily adequate for representing textures, recent studies proposed an additional separate coding of residual textures generally by means of sparse coding algorithms [148, 190, 214].

The method is also based on image denoising applications by wavelet thresholding or shrinkage [197]. Overcomplete transforms have been shown more efficient than critically sampled ones in the denoising tasks. An augmented resolution in orientations has also been shown important [189] as well as interactions between neighboring coefficients [160]. Moreover denoising has been described as closely related to wavelet image compression [20].

Therefore, we propose to build a biologically inspired sparse approximation method constituted by a linear overcomplete log-Gabor wavelet transform and a non-linear stage of sparse approximation based on the edge extraction proposed in Chapter 3. The sparse approximation step aims at selecting a reduced set of transform coefficients located on edges and ridges of the image and able to approximate the image. Therefore each edge segment must be coded by very few coefficients. This implies the decomposition functions are diverse and adequate for matching with edges of natural images, which advocates one additional time for the use of an overcomplete transform. It is to note moreover that the image-adaptive selection of a few coefficients and the elimination of the other coefficients is only possible because the transform is overcomplete which implies that the coefficients are not linearly independent and consequently that there exist important degrees of freedom in the distribution of the amplitude. It is notable that in opposition to classical sparse approximation algorithms like MP or BP, here the selection of the coefficients is achieved through local operations and either through multiple selections at each iteration or even in a non-iterative manner. In both cases the computational complexity is dramatically reduced.

The chapter is structured as follows, Section 4.2 describes the sparse log-Gabor transforms. Section 4.3 presents the results on image compression and denoising in comparison with the state of the art image processing algorithms. Conclusions are drawn in section 4.4.

4.2 Methods

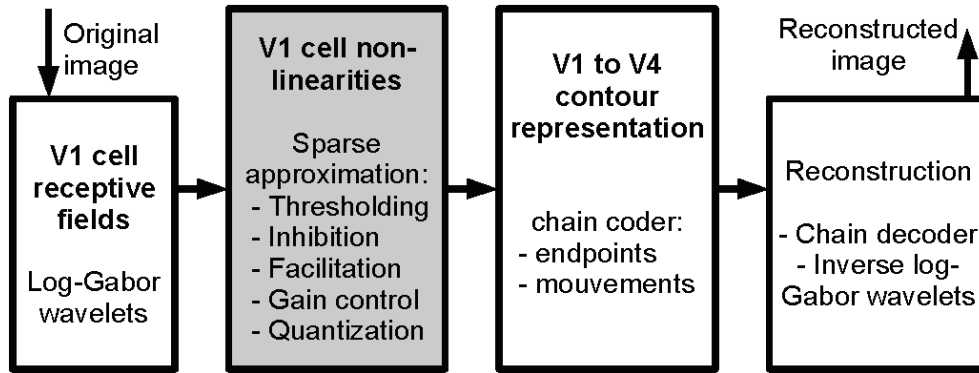


Figure 4.1: Scheme of the algorithm. The lossy parts, i.e. the operations inducing information losses, are depicted with gray color.

The present algorithm is composed of three parts (as summarized in Fig. 4.1): it begins with the linear and self-invertible log-Gabor wavelet transform defined in Chapter 2, decomposing the image in oriented band-pass channels. Afterwards it follows with the different non-linear operations described in Chapter 3: a noise elimination through thresholding, an inhibition of the non-edge coefficients by non-maxima suppression and a facilitation across scales and between co-aligned coefficients. Those non-linear operations aim at extracting the edges of the image and are biologically-inspired from lateral interactions between V1 cells. The present chapter describes the third part and proposes algorithms for building sparse representations based on the edge content of the images. Once the edge coefficients are extracted, there is still to assign them the amplitude which provides the closest possible reconstruction. This is achieved through two algorithms: the first one exposed in Section 4.2.3 proceeds by iterative edge selection while the second one (Section 4.2.4) operates a single-pass edge extraction. The first algorithm was designed at an earlier stage of the development of the present thesis. Although important improvements were incorporated in the second algorithm like the predictive chain coding strategy, this first algorithm is here reproduced since it provides better results in terms of mathematical error. The second one has the advantage to

extract continuous contours which can be further decorrelated by a predictive coding through chains of coefficients as described in Section 4.2.5. First of all, in Section 4.2.1, some examples are presented to illustrate the concept of sparse approximation under overcomplete dictionary, after what in Section 4.2.2 the problem of amplitude assignation will be formulated in the matrix notation.

4.2.1 Non-uniqueness under overcomplete dictionaries

As a first example illustrating the sparse approximation objectives let's consider an image \mathbf{x} composed of a single log-Gabor function (Fig. 4.2.a) which can be coded by a pyramid \mathbf{h}^* containing a single non-zero coefficient (Fig. 4.2.b). Applying the linear log-Gabor wavelet transform \mathbf{W} (the 4 scale 4 orientation with high-pass filter scheme is used here, see Fig. 2.1) on \mathbf{x} does not provide the pyramid \mathbf{h}^* but a pyramid \mathbf{h} (drawn in Fig. 4.2.c). This pyramid \mathbf{h} contains a large number of coefficients spread out in the transform domain in the spatial and frequency neighborhood of the original Gabor function. It is straightforward to show that both pyramids \mathbf{h}^* and \mathbf{h} can be used for exact reconstruction of \mathbf{x} by the inverse transform \mathbf{W}^T (see Chapter 2 and Section 2.2.9 for the exact reconstruction properties of \mathbf{W}). Nevertheless \mathbf{h}^* is a much sparser representation: after quantization, an entropy calculation gives 19 bits for \mathbf{h}^* and 1697 bits for \mathbf{h} . Hence, because of the overcompleteness, to each image correspond different pyramids of exact reconstruction which have very different sparseness. The goal of the sparse approximation algorithm would then be to obtain the less redundant pyramid, that is the \mathbf{h}^* pyramid preferentially to the \mathbf{h} one.

Coefficients of \mathbf{h} represented Fig. 4.2.c can be considered as the projection of the unique non-zero coefficient of \mathbf{h}^* (Fig. 4.2.b) over the whole dictionary of log-Gabor functions, since from Eq. 2.7 and Eq. 2.8, we have:

$$\mathbf{W}\mathbf{x} = (\langle \mathbf{w}_k, \mathbf{w}_j \rangle)_{k \in \{1, \dots, M\}} \quad (4.1)$$

where $(\mathbf{w}_k)_{k \in \{1, \dots, M\}}$ is the log-Gabor dictionary and \mathbf{w}_j is the single non-zero log-Gabor function of \mathbf{h}^* (see also Table 4.1 for the notation). Thus from Fig. 4.2.c it can be seen that the transform coefficients are not orthogonal to their spatial and frequency neighbors. Moreover the magnitude of their correlation $\langle \mathbf{w}_k, \mathbf{w}_j \rangle$ is directly appreciable through the gray-scale level in Fig. 4.2, showing its progressive decrease away from the original non-zero \mathbf{h}^* coefficient.

Note also that the linear transform does not optimally localize image features: the single Gabor function is detected in a large set of coefficients at different scales, orientations and positions. From \mathbf{h} pyramid it is even difficult to say how many Gabor

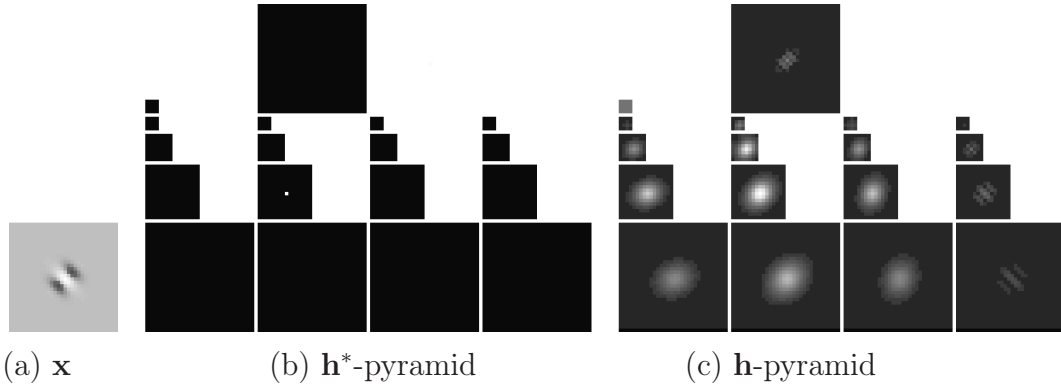


Figure 4.2: Comparison between two exact reconstruction pyramids **a.** The 32×32 pixel image \mathbf{x} is composed of an unique Gabor function that can be coded by the pyramid \mathbf{h}^* shown in (b). **b.** The \mathbf{h}^* -pyramid contains a single non-zero coefficient. Here a 4 scale, 4 orientation scheme with low-pass and isotropic high-pass is used (see the filters in Fig. 2.1). The low-pass and high-pass channels are located on the first row. The bottom row represents the first scale, i.e. the highest frequency channels. The different orientations are arranged in columns. Each band-pass coefficient contains a real and an imaginary part, although only the modulus is shown here. **c.** \mathbf{h} pyramid is the linear transform of \mathbf{x} (explanations are given in the text).

functions are present in the image (a). From the \mathbf{h}^* pyramid, the answer is clearly one. Thus only a non-linear sparse representation is potentially able to describe image features at the optimum of localization theoretically offered by the Gabor framework [68], that is through single Gabor functions, what a linear representation is unable.

As a second illustration let's consider a pyramid \mathbf{h}^* containing a horizontal chain of coefficients (real and positive) in the third scale, fourth orientation channel (see Fig. 4.3.a, here a 4 scale and 6 orientation scheme is used, see Fig. 2.2). Its reconstruction is the image $\mathbf{x} = \mathbf{W}^T \mathbf{h}^*$ (drawn in Fig. 4.3.b). And the linear transform of \mathbf{x} is $\mathbf{h} = \mathbf{W} \mathbf{x} = \mathbf{W} \mathbf{W}^T \mathbf{h}^*$ shown Fig. 4.3.c. It is clear from Section 2.2.9 that \mathbf{h} and \mathbf{h}^* have the same reconstruction by \mathbf{W}^T . Nevertheless the linear transform \mathbf{h} is a much less sparse representation than \mathbf{h}^* . $\mathbf{W} \mathbf{x}$ represents the edge present in \mathbf{x} by an important redundant number of Gabor coefficients spread in a spatial, scale and orientation neighborhood of the optimal coefficients (i.e. \mathbf{h}^* coefficients).

4.2.2 The problem of amplitude assignation

Chapter 3 allows to select a sub-dictionary of functions \mathcal{D}^* from which we propose here to approximate the image. There is still to assign the correct amplitude to each coef-

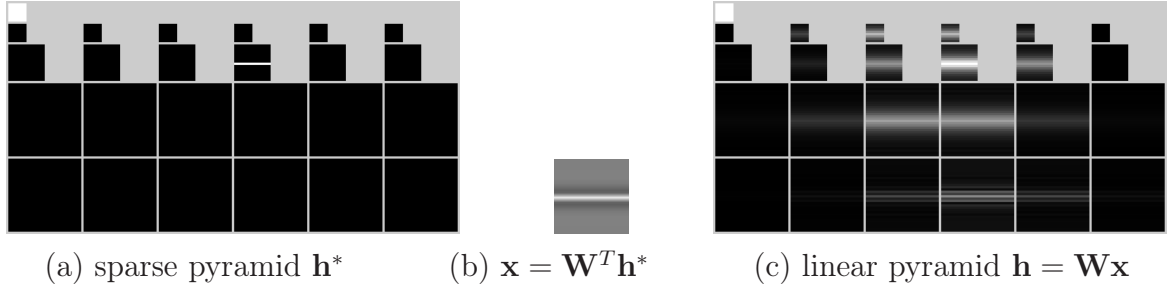


Figure 4.3: Example of the behavior of the linear Gabor wavelets. **a.** The \mathbf{h}^* pyramid is a sparse representation of the image \mathbf{x} with only one line of coefficients in the 3rd scale. Here a 4 scale and 6 orientation scheme inspired from Fig.2.2 is used. **b.** The image \mathbf{x} is constructed from the pyramid \mathbf{h}^* as $\mathbf{x} = \mathbf{W}^T \mathbf{h}^*$. **c.** The linear transform of \mathbf{x} is $\mathbf{h} = \mathbf{W} \mathbf{x}$. The amplitude of \mathbf{h} is spread in a neighborhood of the coefficients of \mathbf{h}^* . It is remarkable that the largest coefficient of \mathbf{h} is 6.79 times lower than the non-zero coefficients of \mathbf{h}^* .

ficient of this sub-dictionary. The problem of amplitude assignation can be formalized in the matrix notation as follows. The source image is \mathbf{x} , and the sub-dictionary \mathcal{D}^* has been chosen through the edge extraction energy model from the overcomplete dictionary \mathcal{D} . \mathbf{h} is the log-Gabor wavelet transform of \mathbf{x} , defined by $\mathbf{h} = \mathbf{W} \mathbf{x}$. $\bar{\mathbf{h}} = \mathbf{W}^* \mathbf{x}$ corresponds to the linear pyramid \mathbf{h} where only the coefficients of \mathcal{D}^* remain. It is to find a subpyramid \mathbf{h}^* and the reconstruction $\mathbf{x}^* = \mathbf{W}^{*T} \mathbf{h}^*$ which minimizes the squared error $SE = \|\mathbf{x}^* - \mathbf{x}\|^2$. The whole notation is summarized in Table 4.1. We have then:

$$SE = \left\langle \sum_{i \in I^*} h_i^* \mathbf{w}_i - \mathbf{x}, \sum_{i \in I^*} h_i^* \mathbf{w}_i - \mathbf{x} \right\rangle \quad (4.2)$$

$$\frac{d}{dh_i} SE = 2 \left\langle \mathbf{w}_i, \sum_{j \in I^*} h_j^* \mathbf{w}_j - \mathbf{x} \right\rangle \quad (4.3)$$

SE is minimum if $\frac{d}{dh_i} SE = 0, \forall i \in I^*$. Then:

$$\left\langle \mathbf{w}_i, \sum_{j \in I^*} h_j^* \mathbf{w}_j \right\rangle = \left\langle \mathbf{w}_i, \mathbf{x} \right\rangle, \forall i \in I^* \quad (4.4)$$

We have then:

$$\mathbf{W}^* \mathbf{W}^{*T} \mathbf{h}^* = \mathbf{W}^* \mathbf{x} = \bar{\mathbf{h}} \quad (4.5)$$

$\mathbf{W}^* \mathbf{W}^{*T}$ has a direct interpretation, since it is the matrix of correlations between the vectors \mathbf{w}_i of the sub-dictionary \mathcal{D}^* .

$$(\mathbf{W}^* \mathbf{W}^{*T})_{i,j} = \left\langle \mathbf{w}_i, \mathbf{w}_j \right\rangle \quad (4.6)$$

Note that $\mathbf{W}^* \mathbf{W}^{*T}$ is invertible only if the vectors of \mathcal{D}^* are linearly independents, which is generally not the case. $\mathbf{W}^* \mathbf{W}^{*T}$ being generally not invertible, the optimal \mathbf{h}^* could

Table 4.1: Sparse approximation variables

Name	Symbol	Description
Original image	\mathbf{x}	$\mathbf{x} \in \mathbb{R}^N$
Overcomplete Dictionary	\mathcal{D}	$\text{card}(\mathcal{D}) = M$
Dictionary functions	\mathbf{w}_i	$\mathcal{D} = (\mathbf{w}_i)_{i \in \{1, \dots, M\}}$
Dictionary indices	I	$I = \{1, \dots, M\}$
Multiresolution transform	\mathbf{W}	$\mathbf{W} \in \mathbb{R}^{N \times M}$
Inverse transform	\mathbf{W}^T	$\mathbf{W}^T \in \mathbb{R}^{M \times N}$
Transform pyramid	\mathbf{h}	$\mathbf{h} = \mathbf{W}\mathbf{x} = (\langle \mathbf{w}_i, \mathbf{x} \rangle)_{i \in I}$
Reconstruction	$\mathbf{x}^r = \mathbf{x}$	$\mathbf{x}^r = \mathbf{W}^T \mathbf{h}$
Selected subdictionary	\mathcal{D}^*	$\mathcal{D}^* \subset \mathcal{D}, \text{card}(\mathcal{D}^*) = M^*$
Subdictionary indices	I^*	$\text{card}(I^*) = M^*$
Transform on subdictionary	\mathbf{W}^*	$\mathbf{W}^* \in \mathbb{R}^{N \times M^*}$
Inverse on subdictionary	\mathbf{W}^{*T}	$\mathbf{W}^{*T} \in \mathbb{R}^{M^* \times N}$
Subpyramid on \mathcal{D}^*	$\bar{\mathbf{h}}$	$\bar{\mathbf{h}} = \mathbf{W}^* \mathbf{x} = (\langle \mathbf{w}_i, \mathbf{x} \rangle)_{i \in I^*}$
Optimal subpyramid on \mathcal{D}^*	\mathbf{h}^*	$SE = \ \mathbf{W}^{*T} \mathbf{h}^* - \mathbf{x}\ ^2$ minimal on \mathcal{D}^*
Reconstruction on the selection \mathbf{h}^*	\mathbf{x}^*	$\mathbf{x}^* = \mathbf{W}^{*T} \mathbf{h}^*$

be found by applying the pseudo-inverse of $\mathbf{W}^* \mathbf{W}^{*T}$ on $\bar{\mathbf{h}}$. Nevertheless, the results obtained by this method were not shown satisfactory for the computational cost, the apparition of artifactual noise and the presence of very small eigenvalues of $\mathbf{W}^* \mathbf{W}^{*T}$. Other solutions are then proposed in the following.

4.2.3 Iterative sparse approximation algorithm

The first proposed algorithm is iterative and proceed through the following steps to define at each iteration n a new pyramid \mathbf{h}_n with exact reconstruction.

Step 1. Coarse edge extraction by linear Gabor wavelets: A first coarse extraction of the edges is obtained by the linear Gabor wavelets presented in Chapter 2 ($\mathbf{h}_0 = \mathbf{W}\mathbf{x}$).

Step 2. Edge selection and coefficient inhibition: In the first iterations the image is approximated with only few coefficients situated on the edges considered as the most reliable, and at each further iteration a new set of coefficients is additionally employed. The reliability for a coefficient to belong to an edge is given by \mathbf{e}_n defined

as:

$$\mathbf{e}_n = \left| \sum_{k=0}^{n-1} \mathbf{c}_k \cdot \mathbf{h}_k \right| \quad (4.7)$$

where \cdot is the multiplication coefficient by coefficient, $|\cdot|$ is the absolute value and \mathbf{c}_k is the contour saliency calculated as the chain length in Section 3.2.4. In this algorithm the facilitation based in the "parent" coefficient is not taken into account although it could constitute a further improvement. Eq. 4.7 implies that the edge reliability of any coefficient increases according to three parameters: its contour saliency $(\mathbf{c}_k)_{k \in \{0, \dots, n-1\}}$ calculated in the successive iterations, its amplitude in all the previous pyramids $(\mathbf{h}_k)_{k \in \{0, \dots, n-1\}}$, and the number of iterations n already completed (then the reliability of coefficients tends to increase through iterations).

\mathbf{h}_{n-1} coefficients which \mathbf{e}_n parameter surpasses a given threshold τ are considered as *reliable*. τ is chosen 10 times higher than the highest \mathbf{h}_0 coefficient ($\tau = 10 \max|\mathbf{h}_0|$).

Two new pyramids \mathbf{a}_n and \mathbf{r}_n are build such that each \mathbf{h}_{n-1} coefficient appears only one time either in \mathbf{a}_n or in \mathbf{r}_n (while the other \mathbf{a}_n and \mathbf{r}_n coefficients are zeros). Then:

$$\mathbf{h}_{n-1} = \mathbf{a}_n + \mathbf{r}_n \quad (4.8)$$

$\mathbf{a}_n \in \mathbb{R}^M$ is the *approximation pyramid* which gathers the reliable \mathbf{h}_{n-1} coefficients. \mathbf{a}_n is then a selection of coefficients supposedly best adapted for reconstructing the image. $\mathbf{r}_n \in \mathbb{R}^M$ is the *residual pyramid* composed by the rest of the coefficients, i.e. the coefficients susceptible to be inhibited.

Step 3. Exact reconstruction by projection of the residual: The new \mathbf{h}_n is defined as follows (it is important to note that for many $\mathbf{z} \in \mathbb{R}^M$, $\mathbf{W}\mathbf{W}^T \mathbf{z} \neq \mathbf{z}$, although, as pointed out in Section 2.2.9, $\mathbf{W}^T \mathbf{W} \mathbf{y} = \mathbf{y}$, $\forall \mathbf{y} \in \mathbb{R}^N$):

$$\mathbf{h}_n = \mathbf{a}_n + \mathbf{W}\mathbf{W}^T \mathbf{r}_n \quad (4.9)$$

Note first that \mathbf{h}_n reconstructs the image \mathbf{x} through \mathbf{W}^T (the following successive equalities, based on the Eq. 4.9, Eq. 4.8 and the Section 2.2.9 statements constitute an iterative proof):

$$\begin{aligned} \mathbf{W}^T \mathbf{h}_n & \stackrel{\text{by Eq. 4.9}}{=} \mathbf{W}^T \mathbf{a}_n + \mathbf{W}^T \mathbf{W} \mathbf{W}^T \mathbf{r}_n \stackrel{(\text{Section 2.2.9})}{=} \\ & \stackrel{\text{by Eq. 4.8}}{=} \mathbf{W}^T \mathbf{a}_n + \mathbf{W}^T \mathbf{r}_n \stackrel{\text{iteratively}}{=} \mathbf{W}^T \mathbf{h}_{n-1} \stackrel{(\text{Section 2.2.9})}{=} \mathbf{W}^T \mathbf{h}_0 = \mathbf{x} \end{aligned} \quad (4.10)$$

In Eq. 4.9, $\mathbf{W}\mathbf{W}^T\mathbf{r}_n$ is the projection of the residual \mathbf{r}_n on the whole transform space. Thus part of the amplitude of \mathbf{r}_n coefficients (but not all) is projected over the non-zero coefficients of \mathbf{a}_n . Through that mechanism, at each iteration, some amplitude transfers from \mathbf{r}_n coefficients to non-zero \mathbf{a}_n coefficients, hence the amplitude tends to concentrate on the non-zero \mathbf{a}_n coefficients, i.e. on the reliable edge coefficients.

Step 4. Loop: Next iteration begins from step 2.

Difference with Matching Pursuit in the amplitude assignation

MP selects one unique coefficient h_k at each iteration and assigns it an amplitude defined by:

$$h_k = \langle \mathbf{x}, \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|} \rangle \quad (4.11)$$

where the $\frac{\mathbf{w}_k}{\|\mathbf{w}_k\|}$ are the normalized dictionary functions, see also Section 2.2.6 for the matrix notation. In the present case of multiple simultaneous selections, the iterative approximation could be defined similarly as in MP by:

$$\mathbf{h} = (\langle \mathbf{x}, \delta_{\mathcal{D}^*} \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|} \rangle)_{k \in I} \quad (4.12)$$

where \mathcal{D}^* is the set of selected decomposition functions, $\delta_{\mathcal{D}^*}$ value is 1 if $k \in \mathcal{D}^*$ and 0 otherwise. But because the decomposition functions overlap, this amplitude assignation strategy would artificially enhance the edges in the image domain as illustrated Fig. 4.4.

Alternatively, and making the hypothesis that an edge can be coded by its central chain (as in Fig. 4.3.a-b), the residual will be projected on the \mathbf{a}_n coefficients. This is accomplished by the operation $\mathbf{W}\mathbf{W}^T\mathbf{r}_n$ which project \mathbf{r}_n on the whole transform space.

$$\mathbf{h}_n = \mathbf{a}_n + \mathbf{W}\mathbf{W}^T\mathbf{r}_n \quad (4.13)$$

Through this operation, part (but not all) of the residual amplitude transfers to \mathbf{a}_n coefficients. The operation requires to be iterated to be able to transfer most of the amplitude (see in Fig. 4.5 how the amplitude concentrates through iterations). There is a second reason for iterating the algorithm: it permits to refine the approximation \mathbf{a}_n through iterations, selecting first the most salient chains and adding less salient chains at later iterations. This allows to process less salient edges without to be influenced by residual coefficients coming from edges of higher amplitude, which could yield a displacement of the local maximum and then a miss in their localization.

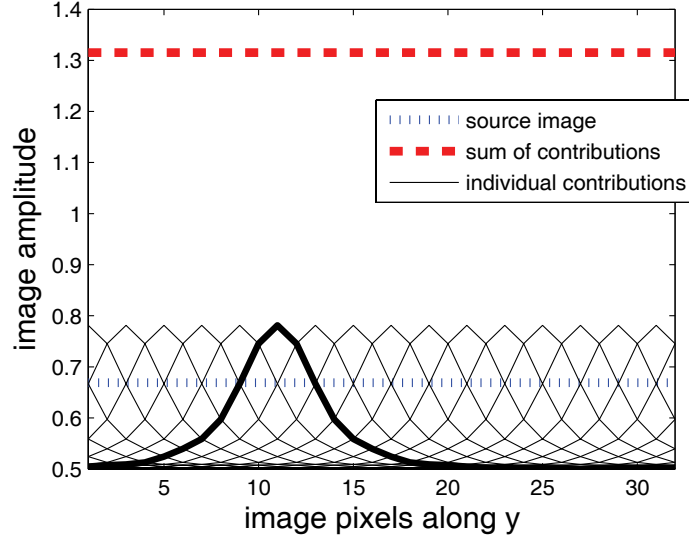


Figure 4.4: Amplitude assignation in a MP-like strategy. The central edge line of \mathbf{x} (Fig. 4.3.b) is drawn here as the dotted line. If the amplitude of the selected coefficients were assigned as in MP following Eq. 4.12, each of them would reconstruct a Gabor function (in continuous line, in the 1D slide along the y dimension, Gabor functions have a Gaussian shape) with a maximum of amplitude higher than the edge to code by a factor of $\sqrt{2}$. Moreover the sum of contributions of all selected coefficients (in dashed) would exceed the edge by a factor 5, yielding an inappropriate enhancement in the reconstruction. Note that the present example is representative of an usual situation in natural images where edges form long and continuous contours of constant contrast. The MP amplitude assignation seems then to overrepresent at least by a factor $\sqrt{2}$ the edge segments embedded in continuous contours (Note nevertheless that in classical MP the factor 5 is avoided thanks to the iterative selection).

Remarks on the algorithm

In contrast to other sparse approximation methods like Matching Pursuit [22, 147, 202], this algorithm is able to select many coefficients at the same iteration. Thus the number of iterations is reduced and it does not depend on the image size (for the following results 300 iterations are used). No proof of convergence has been established for the present implementation. Note nevertheless that experimentally the algorithm always converges in a very regular way. Moreover a proof of convergence for a variation of the algorithm is also proposed in the following section (see Section 4.2.4).

The algorithm is entirely implementable through a fixed number of local operations (\mathbf{W} and \mathbf{W}^T can also be implemented through local operations) being the complexity $O(N)$, where N is the number of pixels of the image.

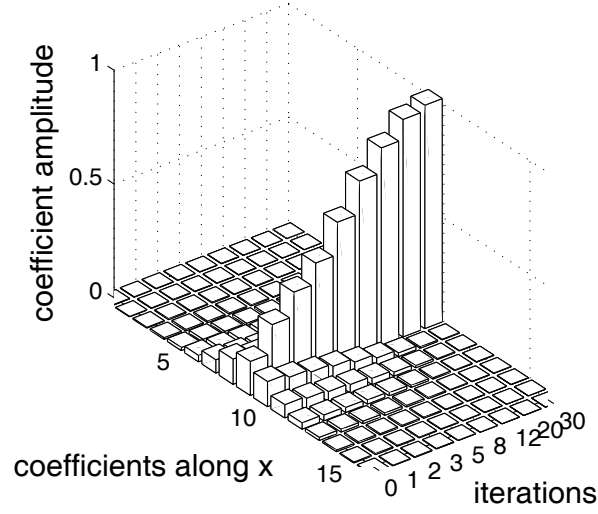


Figure 4.5: The dynamical evolution throughout iterations of the amplitude of a vertical slide of $\mathbf{h}_{(5,3)}$ (from Fig. 4.3.c) is shown. While the residual is backprojected by the operation Eq. 4.9, the selected coefficients grow to the maximum (i.e. by a factor of 6.7) in 30 iterations, whereas close coefficients in the same channel and in neighboring channels (not shown) vanish to zero. The edge is then correctly extracted.

As an example illustrating the abilities of the present method, the algorithm is applied in a basic image containing three Gabor functions shown in Fig. 4.6. It is able to extract progressively through iterations the corresponding three transform coefficients and to zero all the other coefficients.

Quantization and entropy evaluation

A compression application based on the current scheme is described in the following. It contains three stages: the log-Gabor wavelet transform, a quantization and an entropy calculation. The transform is of exact reconstruction thus the quantization is the only lossy part of the compression. The quantization stage is based on a *Contrast Sensitivity Function (C.S.F.)* [120], which gives the sensitivity of the overall visual system at each frequency. The CSF proposed by Rust [176] has been implemented with a low-pass correction for avoiding the low-pass frequency decay of the CSF [50]. Each channel is normalized by its mean contrast sensitivity level. All coefficients can finally be quantized by the same non-linear steps following the method described in [222]. The step widths follow a non-linear progression with an exponential factor of 1.25.

For evaluating the compression rate, an *entropy* calculation is performed in *bit per*

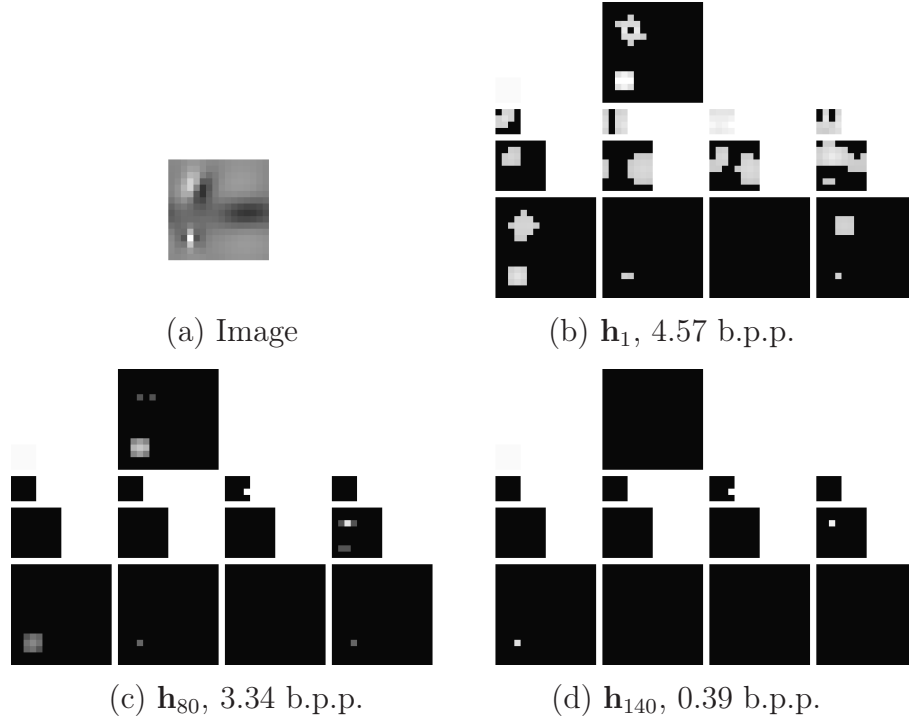


Figure 4.6: Results of the algorithm for a basic example. **a.** The original image (16×16 pixels) is composed of 3 Gabor functions: one in the first scale (phase 0), one in the second scale (phase $-\frac{\pi}{2}$) and one in the 3rd scale (phase π). Here the 4 scales, 4 orientations and high-pass filter scheme is used (see Fig. 2.1). The following insets shows the quantized pyramids. **b.** The linear transform (before applying the first iteration) contains 197 non-zeros. Entropy is evaluated as 4.57 b.p.p., the PSNR of the reconstruction is 33.49dB. **c.** After 80 iterations, only 45 coefficients are kept, most of them close to zero. One 2nd scale and one 3rd scale coefficient appear enhanced. Entropy is 3.35 b.p.p., PSNR=31.78dB. **d.** After 140 iterations, the algorithm has been able to code the image by only 19 coefficients (including the 16 coefficients of the low-pass channel): it has localized the 3 original Gabor functions. Entropy is 0.39 b.p.p., PSNR=35.97dB.

pixel (bpp) as:

$$H = -\frac{M}{N} \sum_q p(q) \log_2 p(q) \quad (4.14)$$

where q is the quantization level, and $p(q)$ its probability, N is the number of pixels of the image and M the number of coefficients of the transform domain. The factor $\frac{M}{N} > 1$ is due to the transform domain expansion. (Note because every scale has different statistics, the entropy of each scale is separately calculated before to be summed up. For the low-pass filter a basic DPCM is applied). Entropy is a measurement of the amount of information contained in the source to be coded. The difference of entropy between the linear and the sparse pyramids can then be considered as a direct evalua-

tion of the redundancy reduction (with the condition that reconstruction qualities are similar). Moreover the entropy gives the theoretical limit (in bit per pixels, b.p.p.) of the compression rate and it is a good evaluation although somewhat optimistic, of the file size that can be produced by an efficient coder. Note nevertheless that contextual coding methods like the one used in JPEG-2000 or chain coding ([168] and Section 4.2.5) are able to produce a file size smaller than this entropy value.

The quality of the reconstructed image is mathematically evaluated using the *root mean square error* (RMSE) and the *Peak Signal to Noise Ratio* (PSNR), defined as $-20\log_{10}(RMSE)$. However it is well known that such mathematical error calculation does not correlate well with the perceptual quality evaluated by human observers. Therefore direct observation will provide a better insight about the performance of the scheme.

4.2.4 Non-iterative sparse approximation

Instead of selecting iteratively the most reliable edges, it is also possible to select in one-pass through the inhibition and facilitation steps (see Chapter 3) the whole set of coefficients. We desire to reconstruct an approximation of the image from this set called *subdictionary*. As previously exposed in Section 4.2.2, to realize the sparse approximation, there is still to assign an amplitude to each of those selected coefficients. As most coefficients of the log-Gabor wavelet transform have been zeroed out, the transform domain lost part of its energy, and the amplitude of selected coefficients has to be enhanced to compensate this lost.

This compensation could be realized through a fixed gain factor. But for a better reconstruction, we adopt a strategy close to Matching Pursuit [118] which plausibility as a biological model has been studied in [153]. MP selects at each iteration the largest coefficient which is added to the approximation while its projection is subtracted from the *residual* (i.e. the transform of the approximation error). This projection, which depends on the correlation between dictionary functions, can be interpreted as a lateral interaction [153]. Here as a difference with MP, the residual r_0 defined as the set of non-selected coefficients is projected on the subspace \mathcal{V} spanned by the selected subsdictionary. We do not know the projection operator P^* that realize this operation. Thus the projector $P = \mathbf{W}\mathbf{W}^T$ that projects the residual on the whole transform space is iteratively used instead¹. Let's be D the diagonal matrix defined on the dictionary space and which eigenvalues are 1 on the selected subsdictionary and 0 on the residual

¹It is direct that P is linear and $P^2 = P$, hence P is a projector.

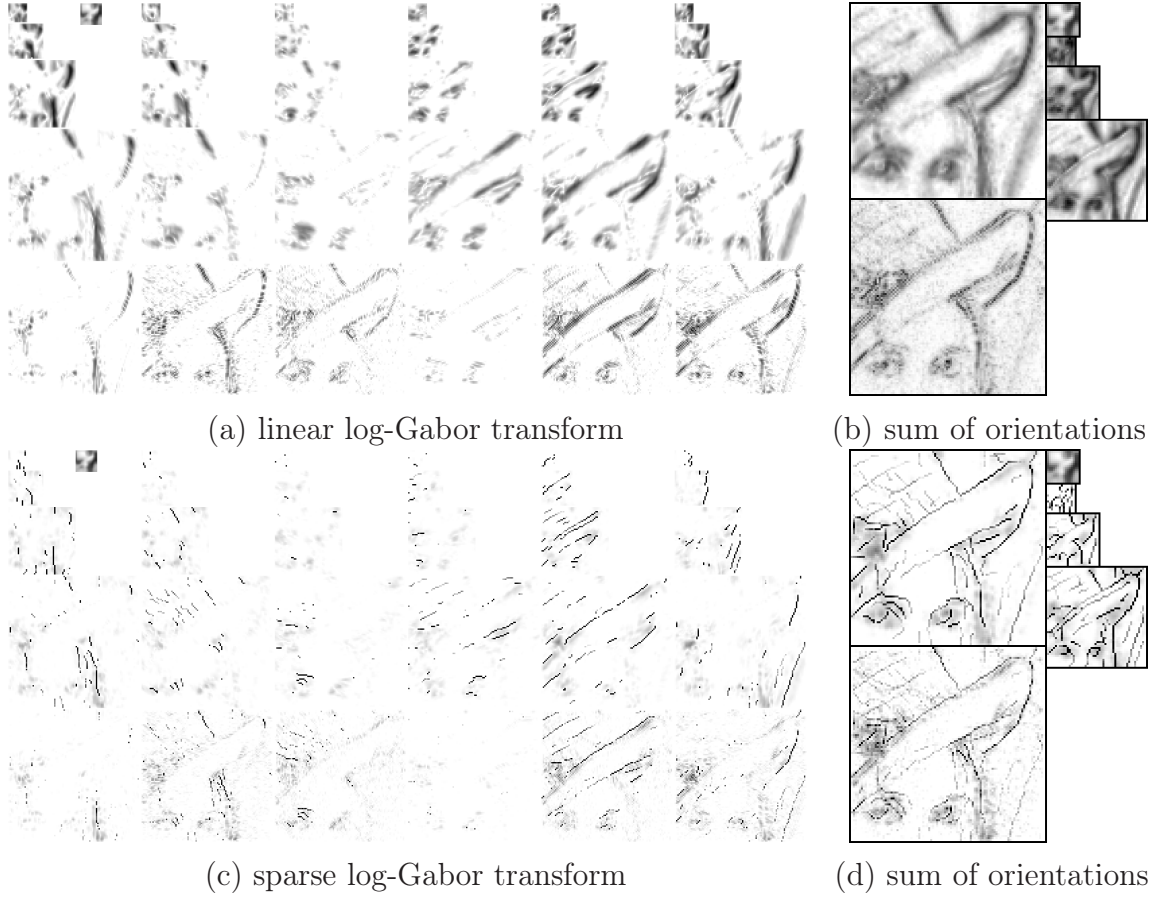


Figure 4.7: Two pyramids providing exact reconstruction for the 96×96 tile of the image 'Lena' presented in Fig. 4.19.a. **a.** In the linear transform each edge is described through a large number of coefficients spread in a space/scale/orientation neighborhood. **b.** The same pyramid is shown with orientations summed together for a better visualization (3th, 4th, 5th and low-pass are drawn on the right side). **c.** After applying the non-iterative sparse approximation algorithm, the signal energy is concentrated in much fewer coefficients which form continuous chains located along the object contours. **d.** Sum of orientations for the sparse pyramid.

(D is clearly a projection). \mathbf{x} is the source image. We set:

$$\begin{cases} \mathbf{a}_0 = \mathbf{D}\mathbf{W}\mathbf{x} \\ \mathbf{r}_0 = (1 - \mathbf{D})\mathbf{W}\mathbf{x} \\ \mathbf{a}_k = \mathbf{a}_{k-1} + \mathbf{D}\mathbf{P}\mathbf{r}_{k-1} \\ \mathbf{r}_k = (1 - \mathbf{D})\mathbf{P}\mathbf{r}_{k-1} \end{cases} \quad (4.15)$$

By the self invertible property we have $\mathbf{W}^T\mathbf{P} = \mathbf{W}^T\mathbf{W}\mathbf{W}^T = \mathbf{W}^T$ and it comes:

$$\mathbf{W}^T(\mathbf{a}_k + \mathbf{r}_k) = \mathbf{W}^T(\mathbf{a}_{k-1} + \mathbf{P}\mathbf{r}_{k-1}) = \mathbf{W}^T(\mathbf{a}_{k-1} + \mathbf{r}_{k-1}) \quad (4.16)$$

Iteratively and using again the self-invertible property and Eq. 4.15 we have finally:

$$\mathbf{W}^T(\mathbf{a}_k + \mathbf{r}_k) = \mathbf{W}^T(\mathbf{a}_0 + \mathbf{r}_0) = \mathbf{W}^T \mathbf{W} \mathbf{x} = \mathbf{x} \quad (4.17)$$

Hence, $\mathbf{W}^T(\mathbf{a}_k + \mathbf{r}_k)$ reconstructs exactly the source image.

It is also straightforward to show that \mathbf{a}_k and \mathbf{r}_k converge: let's \mathbf{Q} be defined as $\mathbf{Q} = (1 - \mathbf{D})\mathbf{P}$. We have now,

$$\begin{cases} \mathbf{a}_k = \mathbf{a}_0 + \mathbf{D}\mathbf{P} \sum_{q=1}^k \mathbf{r}_{q-1} = \mathbf{a}_0 + \mathbf{D}\mathbf{P}(\sum_{q=1}^k \mathbf{Q}^q) \mathbf{r}_0 \\ \mathbf{r}_k = \mathbf{Q}^k \mathbf{r}_0 \end{cases} \quad (4.18)$$

\mathbf{P} and \mathbf{D} being projections, $\|\mathbf{Q}\mathbf{e}\| \leq \|\mathbf{e}\|$ for any vector \mathbf{e} (where $\|\cdot\|$ is the quadratic norm). Moreover any vector \mathbf{e}' which verifies $\|\mathbf{Q}\mathbf{e}'\| = \|\mathbf{e}'\|$, is eigenvector of \mathbf{P} (with eigenvalue 1) and of \mathbf{D} (with eigenvalue 0), then of \mathbf{Q} (with eigenvalue 1). Thus (1) $\mathbf{D}\mathbf{P}\mathbf{Q}^q \mathbf{e}' = 0$; and (2) the eigenvalues of \mathbf{Q} different than 1 are strictly smaller than 1. Hence for any \mathbf{r}_0 , $\mathbf{D}\mathbf{P}(\sum_{q=1}^k \mathbf{Q}^q) \mathbf{r}_0$ and \mathbf{a}_k converge. And from (2) we have the \mathbf{r}_k convergence. The convergence is moreover exponential with a factor corresponding to the highest eigenvalue of \mathbf{Q} which is strictly smaller than 1.

In practice we observe that the algorithm converges with regularity and that in 40 iterations the responses becomes stable. If the dictionary has been adequately selected, most of the residual coefficients dramatically decrease their amplitude and the selected coefficients encode almost all the image information (see example of reconstruction Fig. 3.7 and 3.8). But because some edges and ridges can lack in the dictionary, in particular around corners, crossing and textures, a second pass can also be advantageously deployed for selecting new edge coefficients.

Concerning the overall computational complexity, all the hard thresholding, inhibition and facilitation steps are computed by local operations consisting in convolutions by small kernels (mainly 3×3). The linear and inverse log-Gabor wavelet transforms \mathbf{W} and \mathbf{W}^T are computed in the Fourier domain in $O(N \log N)$ operations but could also be implemented as convolution in the space domain, which is moreover a more biologically plausible implementation. In such a case the algorithm would consist in a fixed number of local operations. As in the former implementation the computational complexity would then be as low as $O(N)$.

4.2.5 Chain coding

The former processes composing the sparse approximation allowed to represent the visual information through continuous and quite long chains of coefficients representing

contour segments (see Fig 3.6.d). The next step in the integration of the visual information would be to build an efficient representation of such chains. It is also plausible that such descriptions of integrated contours could take place in higher visual areas like V2 and V4 which are supposed to provide increasingly complex descriptions of visual shapes. A psychological hypothesis, synthesized in the Attneave's cat, affirms that curvatures are more perceptually salient than straight features, specially those abrupt changes like corners or contour endings. Many psychophysical and physiological experiments from the pioneering works from Hubel and Wiesel 50 years ago [87] have confirmed the existence of hypercomplex or end-stopped cells which are supposed to preferentially respond to ridge endings, abrupt corners and other types of junctions and crossings. Heitger et al. [82] and Dobbins et al. [42] provided end-stopped models for image processing. Recent advances have discovered cells in V4 area highly specialized on complex contour representations [144]. These cells would respond to curvature degree (concavity) and to angles between aggregated curved segments. Thus V1 hypercomplex cells are candidate to represent endings and curvature changes in contours. And V4 neurons could integrate whole curved segments and angles linking adjacent segments.

As a first implementation we choose to represent contours by their endpoints, called chain heads, simulating hypercomplex cells and the contour shape through elementary displacements called movements. This shape representation through successive movements is not biologically inspired but it corresponds to a relatively simple and classical image processing method called *chain coding*. In future implementations a full biological model representing contours through shape parameters such as curvatures and angles could advantageously be set up.

The contour representation aims at further integrating the visual information and at providing a description more easily exploitable by the highest visual areas dealing in particular with object recognition. This objective is known to be directly linked with the redundancy reduction and the decorrelation of the visual information [3]. The sparse approximation already induces an important decorrelation of the visual information and the contour representation aims at removing even higher order correlations. The chain coder is then designed for redundancy reduction, and consequently for image compression.

For improving the compression, we first take advantage of the overlapping of log-Gabor functions to periodically rule out coefficients along chains. In spite of introducing hollows, if these are sufficiently narrow they can be covered enough by the elongated log-Gabor filters so that they are not perceptible. This is the case, for instance, when two over three coefficients are zeroed (as it will be shown Section 4.3.2 and Fig. 4.12, 4.13). This strategy will be exclusively adopted for image compression tasks since it requires a much sparser representation.

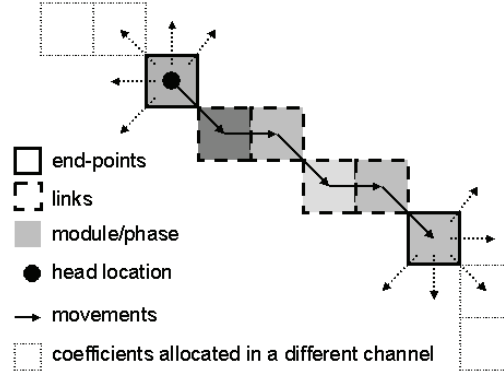


Figure 4.8: Scheme proposed for contour representation.

Chain coding has been many times revisited for efficient representation of contours, whose main precursor was Freeman [67]. He proposed to gather contours in clusters (chains) linking the non-zero adjacent pixels by elementary movements. Here a chain coder has been specially adapted from [168] to log-Gabor channels features². Chains are described by three data: *head coordinates* which are the starting point of chains, *movements* which are the displacement directions to scan chains and *amplitudes* which are the values of log-Gabor coefficients. Fig. 4.8 depicts a scheme of the proposed contour representation.

- **Head Coordinates:** The term *head* is often referred to as the beginning of chains. Two coordinates are then required, vertical and horizontal. They are coded considering the distance between the current head and the previous coded head. The benefit comes from the idea of avoiding to code always the absolute location within channels. Prefix codes compress efficiently such relative distances according to their probabilities. Since channels are scanned by rows, short vertical differences are more probable than long ones, whereas horizontal differences are almost equiprobable.
- **Movements:** Depending on each channel, only those movements not implicated in the lateral inhibition are possible. Two or three movements are then possible. The Fig. 4.9 shows the code-words associated to the three schemes of movement employed. One additional symbol codes the end chain points.
- **Amplitudes:** The modulus of the coefficients is quantified using steps depending on the Contrast Sensitivity Function (CSF) [176] as in Section 4.3.1, while the phase is quantized in 8 values ($-\frac{3\pi}{4}$, $-\frac{\pi}{2}$, $-\frac{\pi}{4}$, 0 , $\frac{\pi}{4}$, $\frac{\pi}{2}$, $\frac{3\pi}{4}$, π). Data to code is the difference between the coefficient value of a log-Gabor coefficient and the previous

²The present chain coding method and its implementations have been developed in collaboration with Rafael Redondo.

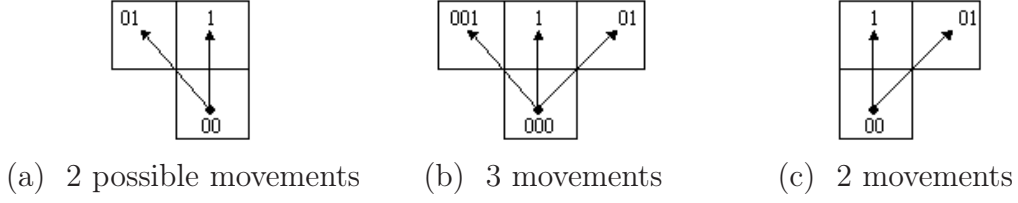


Figure 4.9: Coding schemes applied to movements along chains. Note that they must be oriented according to the log-Gabor channel. Code-word at the bottom of figures means the end of the chain.

one along the chain (prediction error). Moreover head amplitudes, which are used as offsets, can also be predicted, although their correlation is not so high. Two predictive codings (module/phase) for head's amplitudes and two for link (i.e. non-head chain coefficient)'s amplitudes are then encoded by arithmetic coding.

Additionally the residual low-pass channel \mathbf{h}_{nl} is coded by a simple neighboring and causal predictor followed by an arithmetic coding stage. An outstanding report about the here mentioned codings can be found in [86].

As previously mentioned, a much sparser representation can be accomplished if periodic hollows are inserted along chains. However this causes an obvious trouble to track the chains. In order to solve this problem two types of pyramids are used, one pyramid which preserves continuous chains to code heads and movements and a second pyramid which contains the leaky chains to code amplitudes.

Furthermore, revisiting the concept of contour and how is treated by the sparse approximation, natural contours usually present complex shapes unable to be covered by a single channel, they spread across different orientation channels and even across scales. Therefore the channels are not coded independently but an interchannel chain coding step concatenates adjoining chains by their end(starting)-points jumping from one oriented channel to another orientation not necessarily contiguous. See how the contour of the Lena's hat after sparse approximation (Fig. 4.7 and Fig. 3.6.d) can be almost completely outlined through orientations. Further implementations should also envisage interchannel coding between scales taking into account the strong predictability of contours across scales.

4.3 Applications

4.3.1 Image compression through iterative sparse approximation

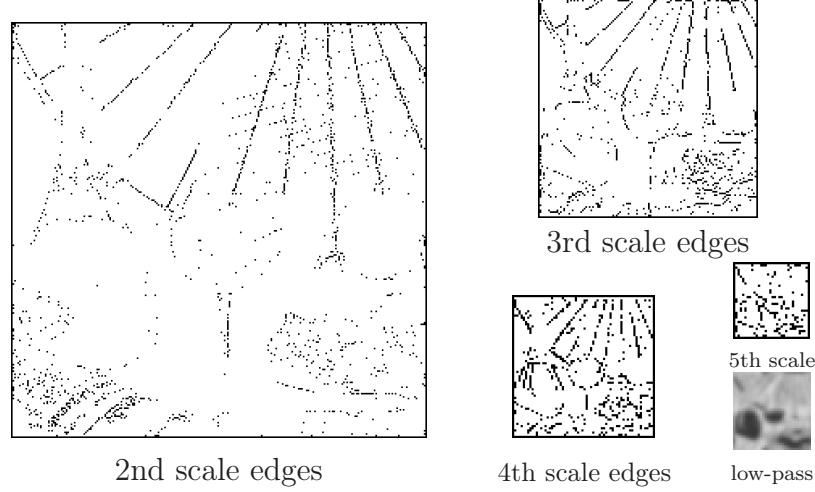


Figure 4.10: Edge coefficients and low-pass channel used to reconstruct the image at 0.57 b.p.p. (no 1st scale edge is used).

Table 4.2: Compression results for iterative sparse coding (see also Fig. 4.11). For sparse coding, the entropy is given as compression rate.

b.p.p.	scheme	$PSNR$	Comments
2.08	JPEG	31.15	good quality
	JPEG-2000	31.55	very good q.
	sparse coding	31.50	good quality
0.57	JPEG	24.50	ringing, blocks
	JPEG-2000	26.78	ringing, aliasing
	sparse coding	24.57	blurring

Results are shown for a 256×256 tile of the ISO image "bike". For both low and high compression rate i.e. entropy values of 2.08 and 0.57 b.p.p., the PSNR for sparse edge coding is slightly better than the one given by JPEG, although worse than the JPEG-2000 one (see Table 4.2). Considering moreover that many improvements are possible for the present scheme, those results show that such sparse edge coding strategy could be competitive with the orthogonal coding methods even in terms of mathematical error.

An observation of the reconstructed images through the three compression methods (Fig. 4.11) shows that sparse approximation using log-Gabor dictionary reduces importantly the high frequency artifacts usually introduced by the orthogonal methods (ringing, blocking or aliasing). Due to the smooth shape of the decomposition functions

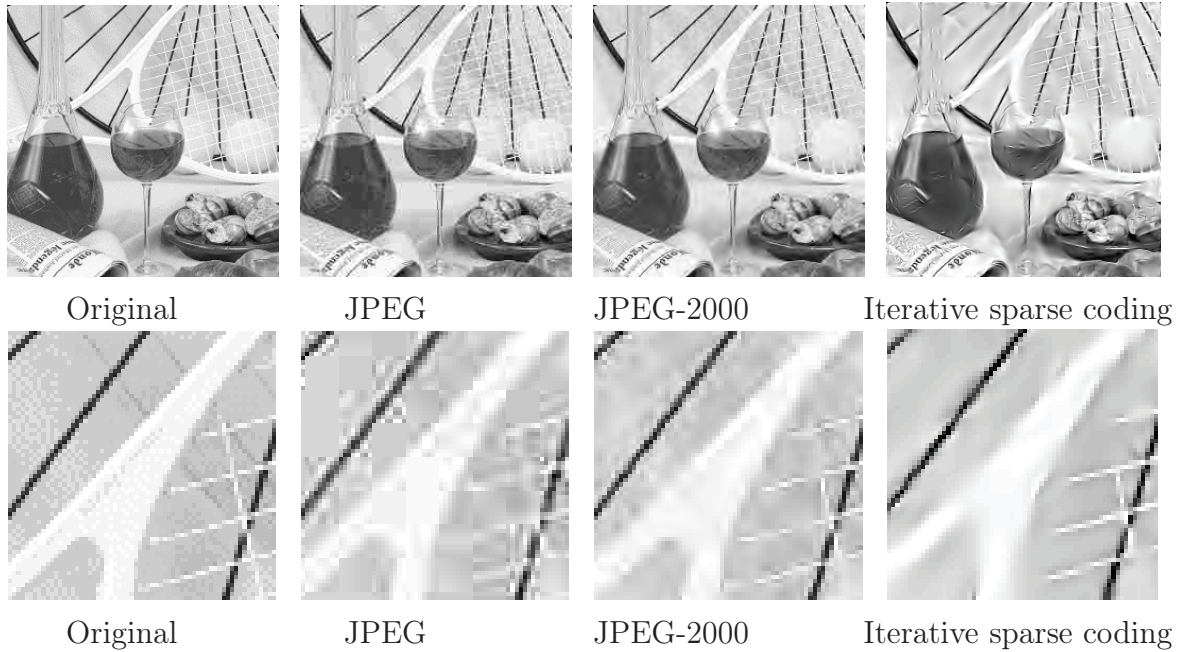


Figure 4.11: Compression results for iterative sparse coding at 0.57 b.p.p. on a 256×256 tile of the image "bike" in comparison with JPEG and JPEG-2000. Edges used for sparse coding are shown in Fig. 4.10. The second line shows the results for a 64×64 zoom tile of the same image.

and the lack of aliasing, the major degradation introduced is a smooth disappearance of edges or blurring. Thus the results can appear visually more pleasant to the human eye than JPEG or JPEG-2000 compressed images, even with higher or similar mathematical error.

4.3.2 Image compression through non-iterative sparse approximation and chain coding

Table 4.3: Compression results for non-iterative selection and chain coding in terms of PSNR for Lena, Boats and Barbara.

Image	bit rate	JPEG	JPEG2000	model
Lena	0.93	22.94	26.09	22.38
Boats	0.30	<21.08	24.48	22.67
	0.55	24.09	27.21	24.06
Barb	0.64	24.62	28.68	24.50
	1.16	30.38	34.10	25.14

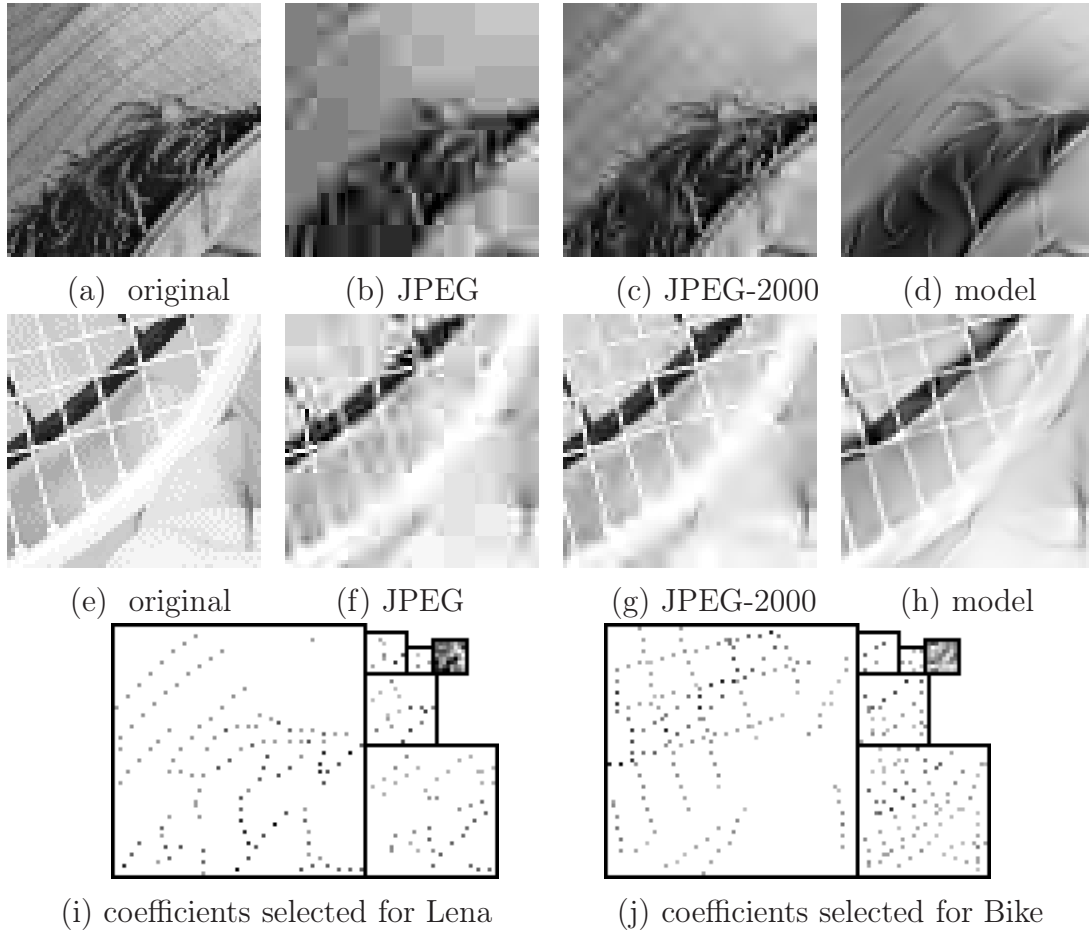


Figure 4.12: Non-iterative sparse approximation applied to the compression of "lena" at 0.93 bpp and Bike at 1.14bpp. **a. and e.** 64×64 tiles of Lena and Bike. Lena is compressed using **(b.)** JPEG, **(c.)** JPEG-2000, **(d.)** the porposed model. bike compressed using **(e.)** JPEG, **(f.)** JPEG-2000, **(g.)** proposed model. **i. and j.** Coefficients selected through the sparse approximation steps respectively for Lena and Bike (2 over 3 coefficients have been zeroed along chains).

The whole process of non-iterative sparse approximation composed by the inhibition, the facilitation and the amplitude assignation, and finally the chain coding are applied to several test images as summarized in Fig. 4.12, 4.13, 4.14, 4.15, 4.16 and 4.17 and in Table. 4.3. Such experiments also aim at evaluating the abilities of the present model of the priamry visual areas to reduce the redundancy of the visual information. Redundancy reduction can be measured through image compression applications which offer a direct evaluation of the methods in terms compression rate (in bpp, bit per pixel), mathematical error and perceptual quality (i.e. visual inspection). JPEG and JPEG-2000 are respectively the former and the actual golden standards in terms of image compression, they are then the principal methods to compare the model with. Additionally, a comparison with MP is included in Fig. 4.13 and Fig. 4.14.

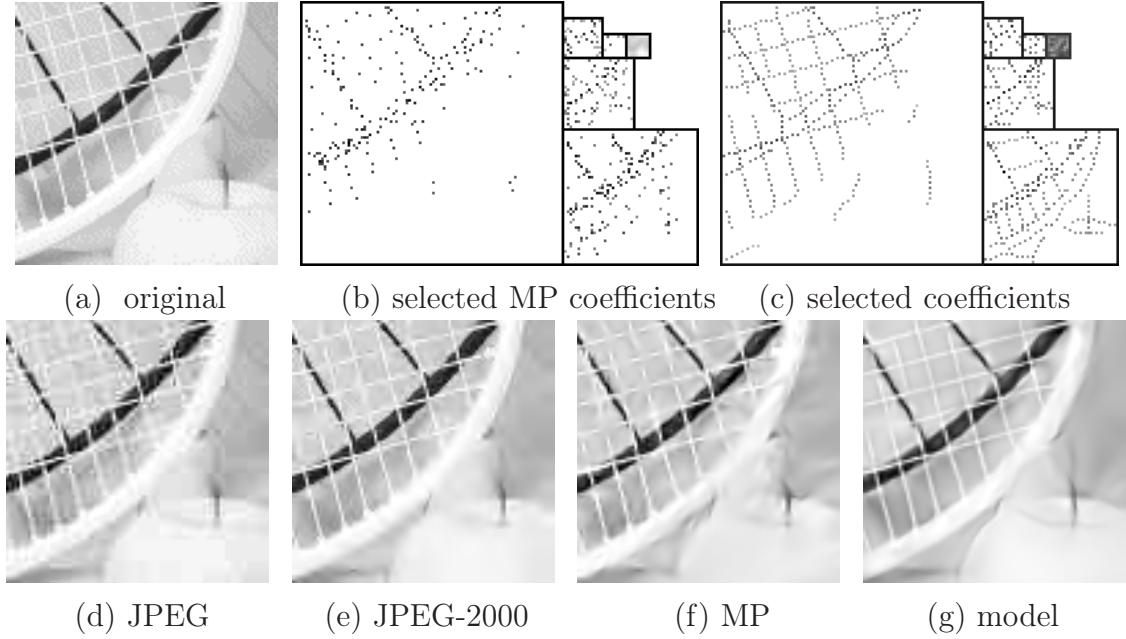


Figure 4.13: Compression results at 1.03 bpp. **a.** 96×96 tile of the "Bike" image. **b.** Coefficients selected through the MP algorithm. **c.** Coefficients selected through the sparse approximation steps (one over two coefficients are zeroed along chains). **d.** Compression with JPEG, PSNR=25.73dB. **e.** Compression with JPEG-2000, PSNR=29.61dB. **f.** Reconstruction by the MP algorithm, PSNR=25.03 dB. **g.** Compression using sparse approximation on log-Gabor wavelets and chain coding, PSNR=26.05dB.

The sparse approximation applied to a tile of "Lena" shown in Fig. 4.12.a induces the selection of a subdictionary shown in Fig. 4.12.i. The chain coding zeroing two over three coefficients along chains compresses the image at 0.93bpp and the reconstruction is shown in Fig. 4.12.d. The comparison at the same bit rate with both JPEG and JPEG-2000 compressed images are shown in (Fig. 4.12.b-c). Similar results for Bike at 1.14bpp are shown Fig. 4.12.e-j. Other results at 0.30, 0.55, 1.03 and 1.90 bpp for the image "bike" are shown in Fig. 4.13 and 4.15, where an additional comparison with MP is proposed.

As shown in Fig. 4.14 the compression standards provide better results in term of PSNR³ at bit rates higher than 1 bpp for the image "bike". In contrast at bit rates lower than 1 bpp the model provides better PSNR than JPEG, and at bit rates lower than 0.3 bpp better than JPEG-2000.

Nevertheless it is well known that mathematical errors are not a reliable estimation of the perceptual quality. Since images are almost exclusively used by humans, it is im-

³ The Peak-Signal-to-Noise-Ratio is measured in dB as $PSNR = -20 \log_{10}(RMSE)$ where $RMSE$ is the Root Mean Square Error between the original and the reconstructed image.

Table 4.4: Gain in terms of quantity of information obtained through the use of chain coding (first two experiences zero 2 over 3 coefficients along chains. The other ones zero 1 over 2. The mean is given for this last case).

Input image	Final bit rate	Coefficients entropy	Chain code entropy	Gain %
Lena	0.93	0.66	0.58	12.12%
Bike	1.14	0.90	0.70	22.22%
Bike	1.94	1.89	1.36	28.11%
	1.03	1.066	0.684	35.87%
	0.547	0.538	0.362	32.73%
	0.299	0.295	0.184	37.50%
Boats	0.55	0.538	0.362	32.71%
	0.30	0.25	0.143	42.80%
Barbara	1.16	1.125	0.81	27.97%
	0.64	0.582	0.412	29.21%
Mean				33.36%

portant to evaluate the perceptual quality of reconstructed images by visual inspection. As the proposed scheme is supposed to model the primary visual areas, it is hoped that the distortions introduced by the model present similarities with the ones produced by the visual system. Then one expectation could be that the distortions introduced by the model would appear less perceptible.

This expectation is fulfilled in a large part since a first remarkable property of the model is the lack of high-frequency artifacts. In contrast to JPEG or JPEG-2000, no ringing, aliasing nor blocking effects appear. Moreover the compressed images look natural and contours appear preserved even when the mathematical error is significantly higher. A requirement of the lossy compression algorithms is the ability to introduce errors in a low perceptible manner. In that sense the model behaves in a very satisfactory way.

Compared with MP, the model provides a more structured arrangement of the selected coefficients (compare Fig. 4.13.b with Fig. 4.13.c), which induces more continuity of the contours in the reconstruction. The model also preserves better the smooth regions thanks to the elimination of isolated coefficients.

Reconstruction quality appears worst in junctions, crossings and corners of the different scales (see also Fig. 4.16 for an image containing many of such features). This

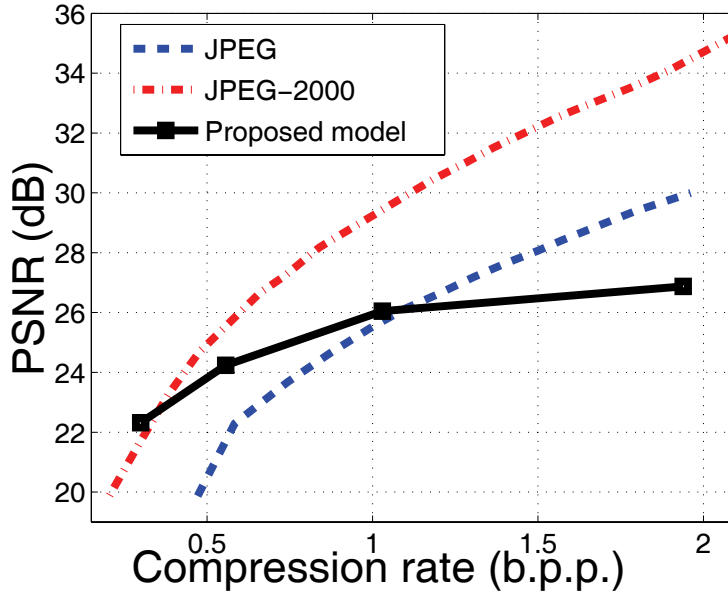


Figure 4.14: Evolution of the PSNR for different compression rates for the image Bike. The proposed method can offer a reconstruction competitive with the compression standard in terms of PSNR at very high compression rates.

can be explained by the good adequacy of log-Gabor functions for matching edges and ridges and their worst match with junction and crossing features. One can argue that the present sparse approximation should be completed by the implementation of junctions/crossing detectors as other models do [82]. This lies nevertheless out of the scope of the present study.

The second problem concerns textures which are generally not well treated by edge extraction methods. One of the worst possible cases are the pure sinusoidal patterns which in some conditions do not even induce local-maxima in the modulus of complex log-Gabor functions. Nevertheless in the majority of cases, textures can be considered as sums of edges. For example in Fig. 4.12 the bristles of Lena's hat constitute a texture and at least the most salient bristles are reproduced. In the same manner the texture constituted by the hat striation is not reproduced integrally but the most salient striations are preserved (note moreover that the striations also tend to disappear in the JPEG and JPEG-2000 compressed images). For further improving the reconstruction quality, and to extract more edges, a few additional passes of sparse approximation can be deployed. For example, a second pass allows the extraction of a significant part of the textures in Barbara's scarf and in its chair as shown in Fig. 4.17. Nevertheless the method does not allow to capture so much sparse representations for textures than it does with contours, the compression quality at the same rate is then significantly lower. As future improvements, it could then be advantageous to deal with textures

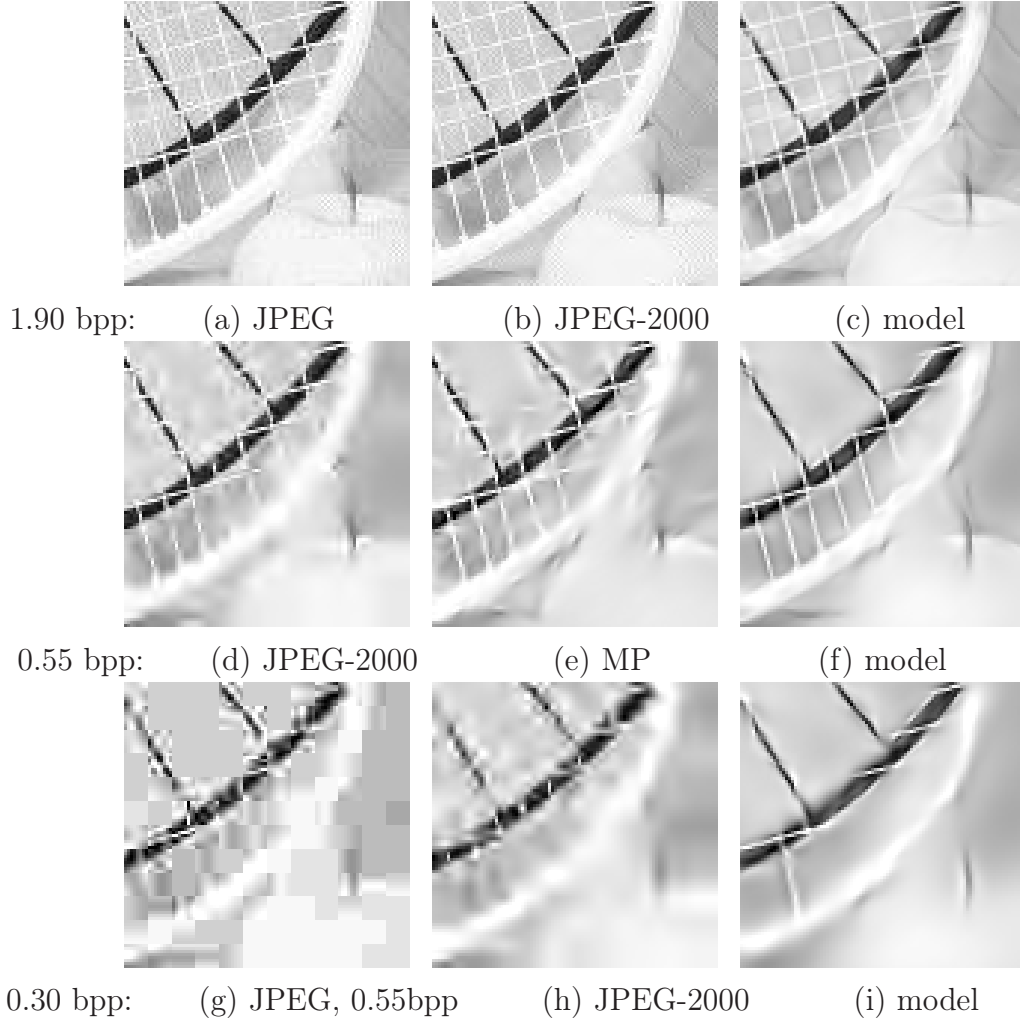


Figure 4.15: Compression results for the image Bike at 1.90 bpp (insets **(a)**, **(b)**, **(c)**), 0.55 bpp (insets **(d)**, **(e)**, **(f)**, **(g)**) and 1.90 bpp (insets **(h)**, **(i)**). The method used (JPEG, JPEG-2000, MP or the present model) is indicated under each inset.

through a separate dedicated mechanism exploiting the texture statistical regularities as those proposed e.g. in [159, 5, 190], or more simply using a standard wavelet coder as proposed in [148, 214]. Such improvements stay nevertheless out of the scope the present study.

The reduction of information quantity between the sparse and the chain representation can be evaluated as around 33% through classical entropy calculations (see Table 4.4). As the chain coder does not introduce information losses (the reconstruction is the same), the information quantity reduction is uniquely due to a redundancy reduction. Thus chain coding offers a significant redundancy reduction for free. This shows the importance of applying an additional transform for grouping selected coefficients in further decorrelated clusters like chains. It is then an important advantage upon MP

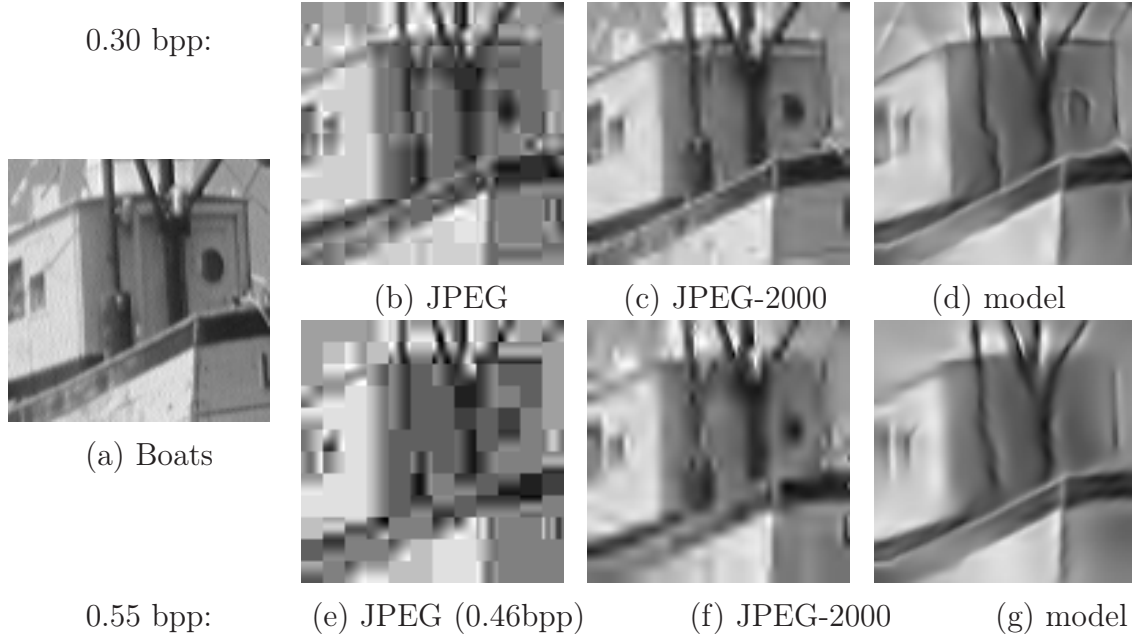


Figure 4.16: Compression results of "Boats" at 0.55 bpp (upper row) and 0.30 bpp (bottom row). **a.** This 96×96 tile of "boats" image contains many junctions and corners, which are difficult features to be captured by the model. **b.** and **e.** Compression with JPEG. **c.** and **f.** Compression with JPEG-2000. **d.** and **g.** Compression using sparse and chain coding.

which induces a representation less structured then harder to further decorrelate.

4.3.3 Non-iterative sparse approximation applied to denoising

Denoising results are presented in Fig. 4.18, 4.19, 4.21 and 4.20 in comparison with the standard method by wavelet shrinkage [28] (orthogonal and undecimated wavelets 'Db4' are used) and the GSM model using steerable pyramids [160]. For all methods the noise level is supposed to be known and the implementation proposed in [160] is used both for the GSM and the wavelet shrinkage methods. In denoising the quality of reconstruction is important, then no edges should be missed in the sparse approximation. Consequently the sparse approximation steps are deployed two additional times on the reconstruction error, so as to extract the residual edges not detected in the first passes.

It is first to note that the method is able to extract and reconstruct almost all the image features. E.g. the reconstruction of the image boats (Fig. 4.18.e) incorporates almost all the original image features. Nevertheless some few edges are lost, in particular close to intricate junctions; as e.g. in the junctions between the twin wires with another

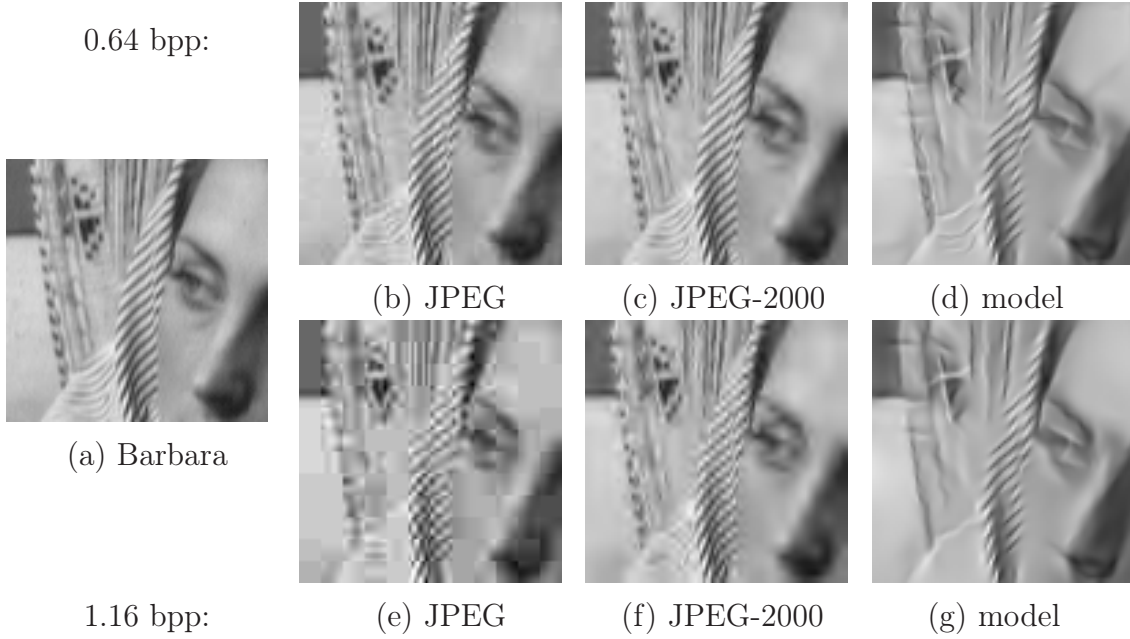


Figure 4.17: Compression results of "barbara" at 1.16 bpp (upper row) and 0.64 bpp (bottom row). **a.** 96×96 tile of "Barbara" image. This image contains textures which are also more difficult features to be encoded by the model. **b.** **and e.** Compression with JPEG. **c.** **and f.** Compression with JPEG-2000. **d.** **and g.** Compression using sparse approximation and chain coding. (Two passes of subdictionary selection are deployed.)

wire in the center/bottom part of the image (see also Lena's right eye and the upper part of the hat border in Fig. 4.19). Thus, at very low noise level the method can not compete with other denoising methods due to the imperfect reconstruction.

In contrast, long edges and lines are particularly preserved. E.g. the boat wires in Fig. 4.18 are particularly well preserved while they tend to disappear and are smoothed in the other methods. This contour preservation remains at high noise levels, where it allows for the image Lena a significant gain over the other methods (the difference is around 0.6 dB with GSM, see Fig. 4.20. Fig. 4.21.g-k lets show that contours are preserved sharper than in the other methods also at very high noise level.

Moreover as in compression, an important quality of the model is to yield reconstructions without high frequency artifacts which allows in particular the preservation of smooth gradients of luminance (see e.g. Lena's skin in Fig 4.19).

For explaining the results, it is worth noting that an important difference between methods reside in the thresholding. Wavelet shrinkage only considers the amplitude of the coefficient, retaining the highest ones as signal and eliminating the smallest coeffi-

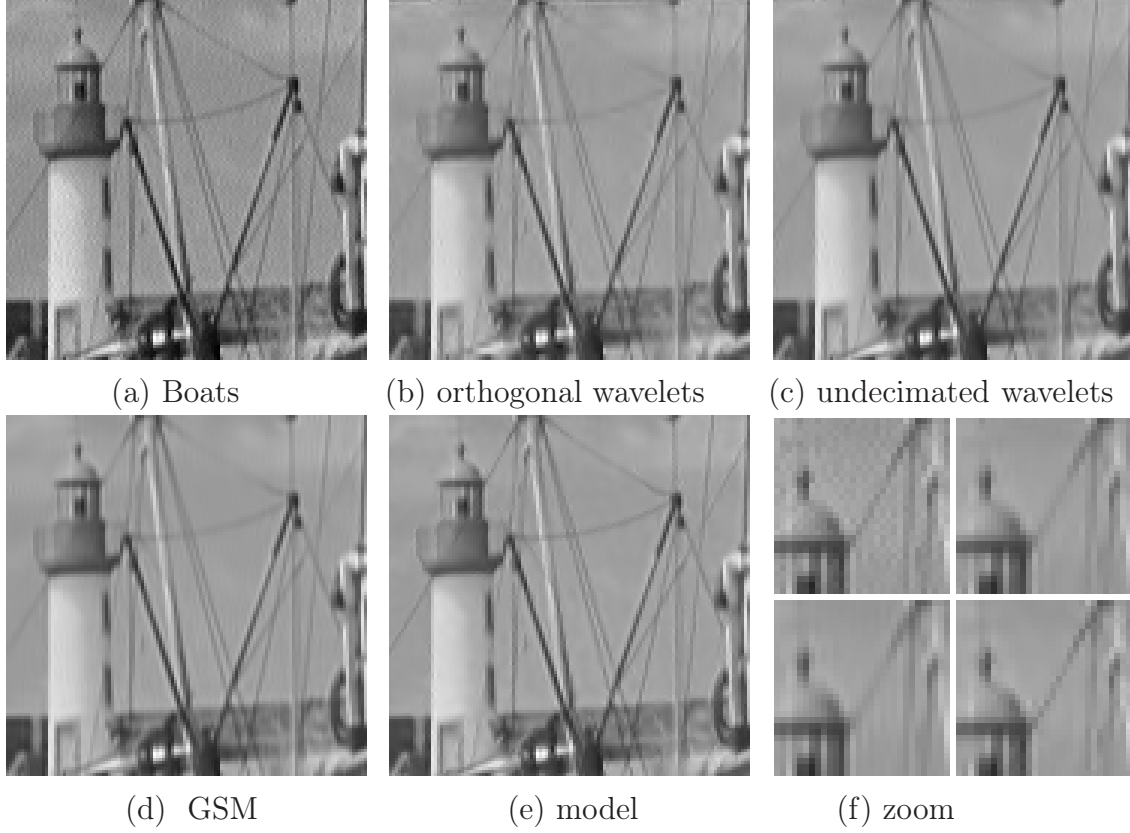


Figure 4.18: Denoising results for a low noise level. **a.** The original image "Boats" already contains a low level of noise. The image is denoised using **(b)** orthogonal wavelet, **(c)** undecimated wavelet, **(d)** the GSM model and **(e)** the present model. The proposed method preserves particularly along lines and edges, as e.g. the wires at the top and at the left of the image. Different details appear smoothed by the other methods while they remain sharp in inset (e). **(f)** Zoom for original, undecimated, GSM and the present model respectively (from left to right and top to bottom).

cients as noise. GSM model considers the 3×3 neighborhood and the parent coefficient in the thresholding decision. In contrast the proposed model takes into account larger neighborhoods by considering that contours are arranged in long chains of co-aligned edges while noise is spatially incoherent.

4.4 Conclusions

We proposed two sparse approximation methods inspired from biological knowledge

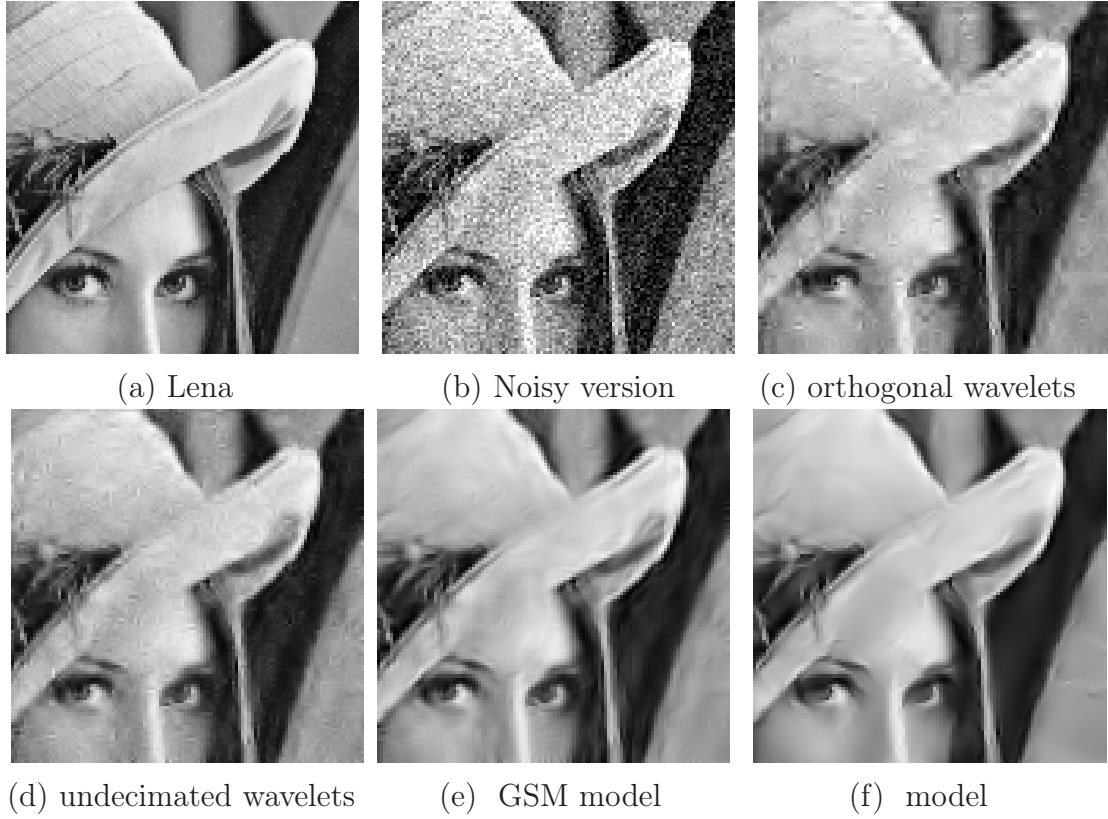


Figure 4.19: Denoising results with Lena image at medium noise level. **a.** 112×112 detail of the image "Lena" corrupted by Gaussian noise for a PSNR of 20.22 dB in **(b)**. **c.** After denoising using orthogonal wavelets a high level of artifacts appears. The level of artifacts is reduced thanks to the use of undecimated wavelets **(d)**, GSM model **(e)** or the proposed model **(f)**. The model preserves particularly smooth gradients of luminance. Nevertheless it shows difficulties in capturing some intricate features in particular close to junctions (see e.g. the right eye) and to adjacent parallel lines (e.g. upper end of the hat border).

on V1 cortical cells and optimized for image processing tasks. They build on the self-invertible linear log-Gabor transform propose in Chapter 2 and in the non-linear steps of inhibition and facilitation between neighboring transform coefficients described in Chapter 3. Those steps are able to extract continuous chains of coefficients located on the edges and ridges of the image, achieving an efficient contour extraction. The sparse approximation schemes select either iteratively or non-iteratively uniquely those contour coefficients and assign them an amplitude for building an approximation of the image.

Such sparse approximations gather a number of interesting properties which can be summarized as follows. First, they are *non-linear*, invertible and allow an *exact reconstruction* (if the residual is included). Second, the inverse transform is just performed

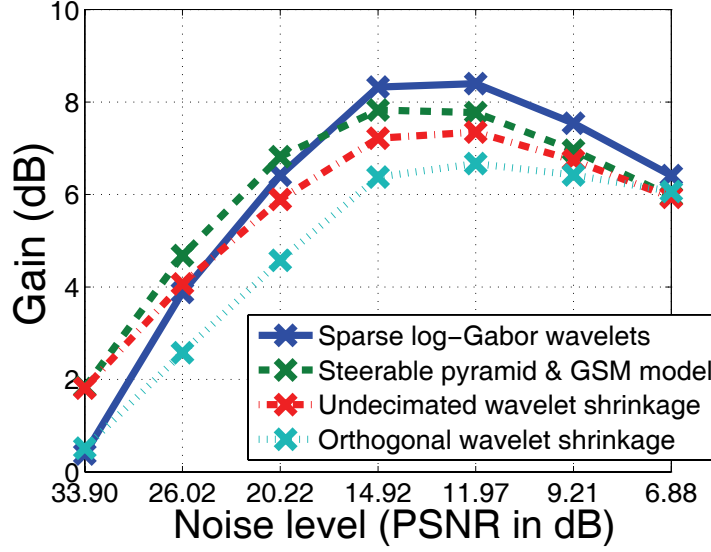


Figure 4.20: Denoising results at different noise levels. (a) The evolution of the denoising results is plotted for Lena (see also Fig. 4.19) as the gain (i.e. the improvement in terms of PSNR), as a function of the noise level (also measured as PSNR). For low level of noise (PSNR of 33.90 and 26.02 dB), the model yields poorer gain than with other methods. This could be due to the approximated reconstruction offered by the sparse approximation method. For high noise level (PSNR lower than 15 dB), the model offers a higher gain than the other methods.

by the linear log-Gabor wavelet inverse transform, what makes it very fast. Third, the *information contents* and the *redundancy* are highly reduced (i.e. efficient compression is possible). Fourth, *visual events (edges) are enhanced, sharper and thinner localized* in the transform domain (in opposition to the linear log-Gabor wavelet transform, the sparse log-Gabor representation allow to achieve the theoretical optimal limit of localization, being able to represent features through single Gabor coefficients. Finally the computational complexity $O(N \cdot \log(N))$ is reasonably low and can be even reduced to $O(N)$ if log-Gabor wavelets are implemented in the space domain through filters of limited support.

The iterative selection seems to yield better results in terms of mathematical errors, but non-iterative selection provides particularly good results in terms of perceptual quality. As an additional advantage of the non-iterative method, the coefficients can be predictively encoded by a chain coding procedure. For both methods the schemes appear potentially competitive with standard image compression algorithms, particularly at high compression rates. By its part, the chain coding is able to code predictively the edge coefficients arranged in continuous contours. It offers then an efficient representation of contours. It has been shown to significantly decrease the entropy of



Figure 4.21: Denoising results with Lena image at low and high noise level. **a.** Original tile "Lena". **b.** The same tile corrupted by Gaussian noise for a PSNR of 33.90 dB. Denoising using **(c)** orthogonal wavelets, **(d)** undecimated wavelets, **(e)** GSM model, **(f)** the proposed model. Although the proposed method yields the worst PSNR of all four methods, the perceptual quality is outstanding first for the elimination of high frequency artifacts and second for the preservation of smooth gradients. It is again remarkable that the sparse approximation reconstructs almost all image features. Bottom insets show denoising results using **(g)** orthogonal wavelets, **(h)** undecimated wavelets, **(i)** the GSM model, and **(j)** the present model on a noisy version of "Lena" **(k)** including a high level of noise (the PSNR is 11.97 dB).

the representation and at the same time it offers an encoding method for writing the compression file. The low redundancy of such chain representation could also make it adequate to build descriptors for object recognition.

In parallel the ability for extracting contours can be exploited for signal to noise segregation. Indeed the method shows promising results for image denoising since it preserves particularly long lines and contours and reduces the appearance of artifacts. Best results are obtained at high noise levels, the worst results at low noise level can be explained in part by the approximated reconstruction.

This confirms the potential of overcomplete transforms and sparse approximation algorithms for image processing and in particular for compression applications. The present study confirms that overcomplete transforms can offer important advantages in terms of perceptual quality in particular for avoiding the appearance of artifacts and preserving smooth gradients and continuous sharp contours. Another significant advantage is a high interpretability of transform coefficients in terms of edges and contours.

On the other hand, the efficiency of the scheme for such a diversity of image processing tasks (edge extraction, denoising and compression) is an additional argument in favor of the hypothesis that similar processes could take place in the primary visual cortex.

It can result interesting to compare the present model with another biologically inspired model called *divisive normalization* [213]. In such model, non-linear interactions between neighboring cells allow an important decorrelation of the transform domain (the scheme is currently employed with (bi-)orthogonal wavelets or overcomplete transforms such as steerable pyramids). Decorrelation and redundancy reduction are tasks supposedly performed by the visual system [7, 3]. Nevertheless it should be taken into consideration that decorrelation by itself does not necessarily implies redundancy reduction. E.g. white noise is decorrelated but it is not sparse and it contains the maximum quantity of information. Thus the "whitening" aimed by the decorrelation and the divisive normalization which does not zero the majority of coefficients could reveal in some cases pathological and could miss to reduce the quantity of information. The proposed model is then different in different aspects: first in the objectives since it aims at sparsifying the transform domain and not at decorrelating it. And second in the approach, since divisive normalization is presented as a general mechanism plausibly present at any level of the cortical pathways while the proposed model is based on the priori information on the type of data to be analyzed (i.e. contours), it is then a specific mechanism adapted to a particular neural function.

Among further improvements, end-stopping operators dealing with the extraction of junctions, corners and crossings could be implemented, which could improve the performance in the proximity of such features. Another unresolved difficulty is to deal in a non-iterative manner with pure sinusoidal patterns and textures which do not induce local maxima in the log-Gabor modulus. Many improvements are also possible in all the different steps of the algorithm, in particular to improve the selection of coefficients by a more sophisticated combination of saliency measurements (chain length, presence of parent coefficients and coefficient amplitude), or to exploit the predictability of the coefficients across scales for image compression.

Chapter 5

Contour completion using tensor voting

"Une lutte contre la pauvreté serait possible. [...]. Elle lutterait contre la pauvreté d'un mot, richesse, auquel a été retranché l'essentiel de son sens. La vraie richesse humaine, plutôt que dans l'accumulation individuelle ou collective délirante d'argent et de biens matériels, n'est-elle pas dans une capacité constante à acquérir et partager de la connaissance, de la sagesse, des sciences, de l'intelligence, de l'imaginaire, de la beauté, de la puissance poétique, une haute conscience de l'autre ? Si tout à coup c'était bien cela la richesse, on verrait qu'elle est mieux répartie, mieux partagée que ce que l'on croit, comment chacun en a mais aussi en est un morceau."

Jean-Michel Bruyère et Issa Samb, La guerre aux Pauvres, 2002.

Tensor voting methods have been developed in a series of papers by Medioni and coworkers during the last years. The method has been proved efficient for grouping and feature extraction and has been applied successfully in a diversity of applications as contour and surface inferences, motion analysis, etc. We present here two studies on improvements of the method. The first one consists in iterating the tensor voting process, and the second one integrates curvature information. In contrast to other grouping methods, tensor voting claims the advantage to be non-iterative. Although non-iterative tensor voting methods provide good results in many cases, the algorithm can be iterated to deal with more complex or more ambiguous data configurations.

We present experiments that demonstrate that iterations substantially improve the process of feature extraction and help to overcome limitations of the original algorithm. As a further contribution we propose a curvature improvement for tensor voting. Unlike the curvature-augmented tensor voting proposed by Tang and Medioni, our method evaluates the full curvature, sign and amplitude. Another advantage of the method is that it uses part of the curvature calculation already performed by the classical tensor voting, limiting the computational costs. Some new curvature-modified voting fields are also proposed. Results show smoother curves, a lower degree of artifacts and a high tolerance against scale variations of the input. The methods are finally tested under noisy conditions showing that the proposed improvements preserve the noise robustness of the tensor voting method.

The present study have been presented in [58, 57]. Tensor voting method aims at grouping edge segment in continuous contours. It completes then the Chapter 3 on edge extraction, and could also serve as a basis for the Chapter 4 on sparse representation from edges. Nevertheless at this stage of implementation tensor voting methods were not integrated neither with the edge extraction nor with the sparse coding. First because the tensor encoding of edges is different to the log-Gabor representation, but also because the proposed method for edge extraction is sufficiently efficient to deal with natural images without necessity of grouping separated edge segments. Indeed edge extraction proposed in the Chapter 3 already achieves continuous representation of contours in most cases.

5.1 Introduction

Perceptual contour extraction methods aim at drawing and completing contours in images on a similar manner a human would do. Methods are often based on Gestalt psychology [99] but also in recent psychophysical experiments [55, 84, 122] and physiological studies of the visual cortex [94, 145]. Based on such studies some algorithms achieve perceptual contour extraction by simulating the behavior of visual cortex neural networks [221, 112, 131, 79, 135, 63, 208]. Other methods as tensor voting [129, 201, 77] implement the Gestalt laws. Alternatively there exist also algorithms based on curves of least energy [179] or stochastic motion [218].

Tensor voting algorithms have been developed by Medioni and coworkers as an efficient method for contour extraction and grouping. Full descriptions of the algorithms can be found in, e.g. [129, 201, 77], and a comparison of tensor voting with other algorithms can be found in [218].

Tensor voting is a non-iterative procedure, in the sense that the original scheme implements only 2 steps of voting, claiming that no further iterative cycles are needed. Many other methods for perceptual grouping, e.g. [221, 112, 131, 79, 135, 63], refine the results by iterative loops. Thus, the first aim of this study is to investigate how the incorporation of iterations in the voting mechanism can improve the results of tensor voting. Some basic examples are analyzed and a statistical study on the influence of iterations is set up in a simple case.

A curvature improvement has already been proposed by Tang and Medioni [195] for 3D images. It computes the sign of curvature to modify the voting fields. For the 2D case, we propose a more sophisticated calculation of the curvature information with a low computational cost. Instead of the sign of curvature, the proposed method evaluates the full curvature information using part of the calculations previously performed by the tensor voting. We adopt a curvature compatibility approach that was described by Parent and Zucker [143]. A statistical evaluation is presented and the methods are finally tested with complex data and in presence of noise. The present work builds upon and further extends the preliminary investigations presented in [57]. New results have been included here for a more complete evaluation and a comparison of the different methods. In particular an evaluation of the noise robustness have been set up. Note also that, very recently Tong and Tang [200] proposed tensors of curvature, showing the increasing interest of curvature evaluation for tensor voting.

Unlike Medioni and coworkers who usually apply tensor voting to volumetric data (3D images) or even N-dimensional features [196, 136], we preferred to restrict the present work to 2D images for the following reasons: (1) 2D examples are easier to understand, which facilitates the study of the basic properties of the underlying approach. (2) Contour extraction in 2D images still remains an unsolved problem, which makes the presented approach an interesting alternative to state-of-the-art algorithms. (3) 2D applications have lower computational cost, which is important since tensor voting methods are computationally expensive.

Section II briefly introduces the tensor voting method. Section III presents a study on iterated tensor voting and the section IV describes improvements that can be achieved when both curvature information and iterations are used. Section V evaluates the robustness of the methods in presence of noise. Finally some concluding remarks are drawn in section VI.

5.2 A brief introduction to tensor voting

Tensor voting algorithm (referred as TV in the following) encodes a local description of the curves at each point of the image by a symmetric positive 2x2 tensor. Those tensors permits to represent an evaluation of the tangential orientation, of the uncertainty on that tangential orientation and of the possible presence of a junction in that point.

Tensors can be diagonalized, their eigenvalues are denoted by λ_1, λ_2 with $\lambda_1 \geq \lambda_2 \geq 0$. The corresponding eigenvectors are denoted by e_1, e_2 . Tensors are decomposed as follows:

$$T = (\lambda_1 - \lambda_2)e_1e_1^T + \lambda_2I \quad (5.1)$$

where I is the identity matrix. The first term is called the *stick component*, where e_1 corresponds to an evaluation of the tangential orientation of the curve segment. The *stick saliency* $\lambda_1 - \lambda_2$ gives a confidence measure for the presence of a curve. The second term is called the *ball component*, and its saliency λ_2 measures the confidence to have a junction. Note that the ball components can also denote an uncertainty on the tangential orientation.

For binary images, tensors can be initialized as ball tensors, i.e. as the identity matrix I at locations that contain data points and null tensors at other locations.

The classical tensor voting algorithm proceeds by two voting steps in which each tensor propagates to its neighborhood. Stick tensors propagate mostly in the direction of e_1 . The region of propagation is defined by the stick voting field which decays as a function of the distance and curvature (see Eq. 5.3 and Fig. 5.3.h). Stick voting fields are built by following the Gestalt laws in perception. These establish criteria for grouping together points that lay on a curve: "good continuation", proximity, favoring low curvatures over large ones and consistency of the curvature [129]. Ball tensors, on the other hand, isotropically propagate in all directions decaying with distance. After the tensor propagation, all contributions are summed up to define new tensors that will be used for the next step. That summation can be considered as averaging or *voting*. The first voting step is referred as *sparse vote* because the vote is performed only on points where tensors are not zero. The second voting step is called *dense vote* because the vote is accomplished at every location. After all voting steps are completed, curves are extracted at local maxima of stick saliency along the normal direction to stick components. Note that a threshold τ is necessary to eliminate low-saliency local maxima. Fig. 5.1 summarizes the different steps of the algorithm.

The tensor voting algorithms will not be described more in detail here. For a more

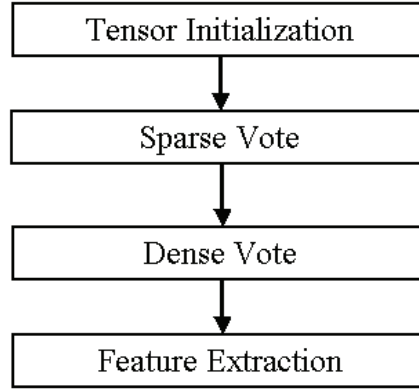


Figure 5.1: The classical tensor voting method is composed of four steps. (1) Tensor initialization, (2) sparse voting, (3) dense voting, and (4) feature extraction.

in depth study the reader can refer to [129, 201, 77, 195].

The novel contributions described in the following are represented Fig. 5.2 in bold-face characters. They consist in iterating the sparse voting step (section III) and integrating curvature information for using curved voting fields (section IV).

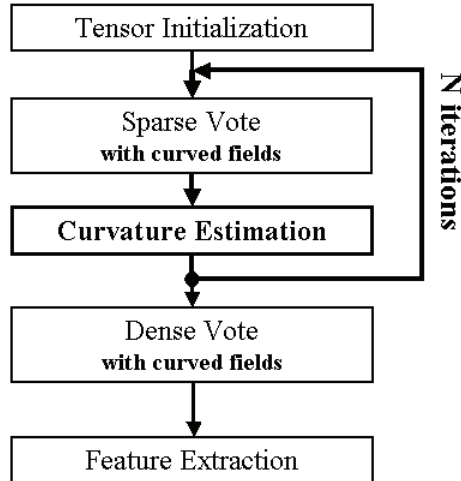


Figure 5.2: The new contributions (in boldface characters) consist in iterating the sparse voting step, evaluating the local curvature and incorporation that curvature information into the voting fields.

5.3 Iterated tensor voting

5.3.1 Example

The original tensor voting approach [129] is particularly efficient for grouping sparse data-points that are separated by gaps of comparable size. A free parameter σ_s , called the scale factor (see Eq. 5.3), has to be minimally adjusted to the inter-distance between points. If σ_s is miss-adjusted, extraction results loose quality: if σ_s is too small points will not be grouped; if σ_s is too big the grouping is less selective.

Fig. 5.3.a shows an example where points are situated along a spiral. Because the distance between two following points is smaller than the distance between two successive arcs and assuming the curvature is low, the Gestalt rules predict that the points have to be circularly grouped along the spiral but not radially. Because the gaps between points vary along the spiral, it is not possible to adjust σ_s for proper extraction of the whole spiral. As it is shown in Fig. 5.3.c, σ_s need to be small, i.e. around 5, for grouping the central part of the spiral. But only a large σ_s , around 11, is able to group the external part of the spiral (Fig. 5.3.i). Between these values (σ_s around 8, Fig. 5.3.f) neither the central part nor the external one are grouped properly.

The accuracy of voted tensors can be checked comparing their orientation with the one of the ideal spiral defined by its mathematical equations. We will consider a tensor as misoriented if the angular error exceed $\pi/10$. With $\sigma_s = 5, 8$ and 11 , respectively 5, 5 and 16 tensors are misoriented (over the 57 tensors situated in the data points).

5.3.2 Iterative tensor voting (IT) method

One objective of the sparse voting step is to evaluate the orientation (tangent) of the curve in each data point, while the dense voting step mostly bound those points (fill the gaps) by propagating the tensors thanks to the voting fields. In principle both sparse and dense voting steps could be iterated, nevertheless it has been chosen not to iterate the dense voting step for the following reasons: (1) The dense voting step has a much higher computational complexity, then its iteration could be extremely costly. (2) The dense voting step frequently produces some irrelevant votes which themselves would propagate through iterations, yielding divergent and unrealistic results. (3) In contrast the sparse voting step only modifies the eigenvalues and eigenvectors of the existing data points, refining them through mutual voting. But it does not add any

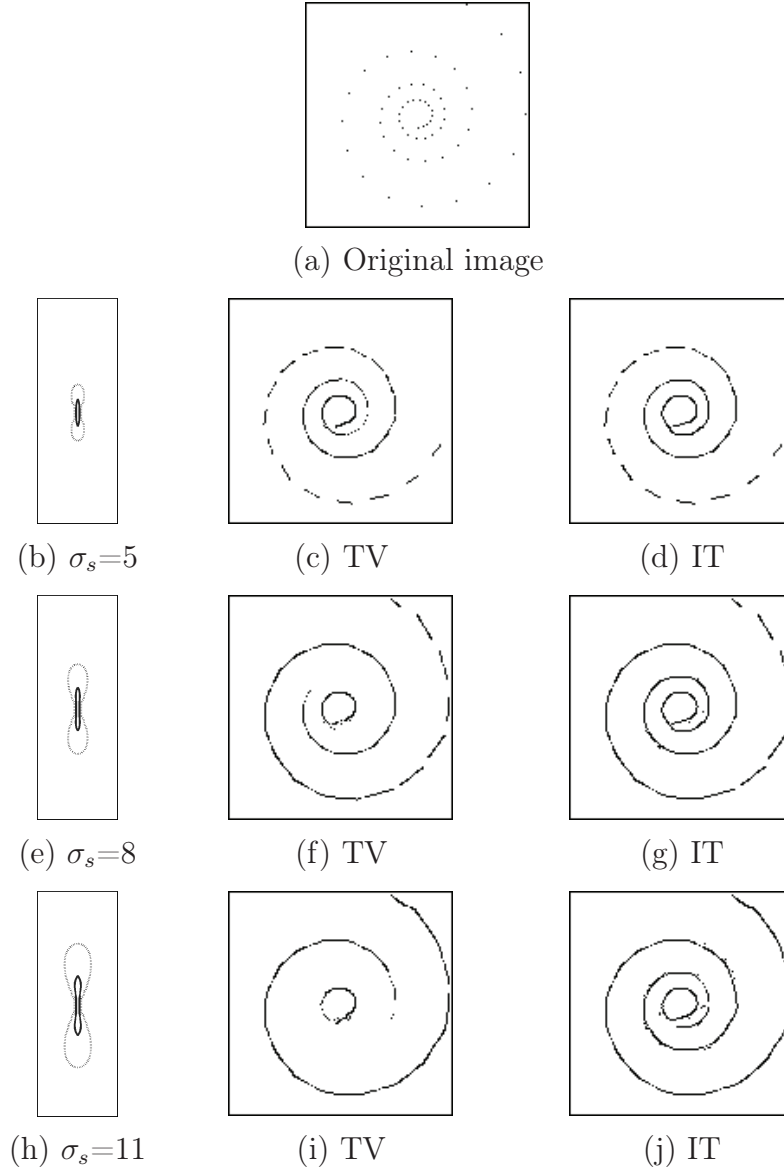


Figure 5.3: Example tested for different values of the scale factor σ_s with the non-iterative method (TV) and with 12 iterations (IT method). **a.** Data points are located along a spiral. **b, e, h.** Contours of the voting field for $\sigma_s = 5, 8$ and 11 are drawn at 50% (solid line) and 5% (dash-dot line) of the maximum value (voting fields are defined by the Eq. (5.3)). **c, f, i.** Extraction results with TV. **d, g, j.** with IT.

tensor in other points. No outlier point appear which is important for the stability of the iterative process.

A divergence of the eigenvalue amplitudes could nevertheless occur by iterating the sparse vote. This can be counteracted by a normalization stage keeping constant the sum of eigenvalues of each tensor. This normalization is applied after each iteration.

Therefore, as it is shown in Fig. 5.2, the sparse voting stage is the only step to be iterated. For n iterations, $n - 1$ sparse votes and one dense vote are realized.

As a consequence, the iterations will mainly refine the orientations of the tensors (i.e. the tangents of the curves), but due to the lack of dense voting iterations, the range of propagation will not be increased by the iterative process.

Fig. 5.3.d, g, j shows the results for 12 iterations with $\sigma_s = 5, 8$ and 11 respectively. Iterations do not fill the gaps between points where σ_s is too small, that is, as predicted, iterations do not increment the range of propagation (Fig. 5.3.d, g). On the contrary, iterations refine tensor orientation so that the central part of the spiral is accurately extracted even for large values of σ_s (compare the central parts of the Fig. 5.3.g, j to the ones in the Fig. 5.3.f, i). Iterations thus allow to extract the entire spiral for $\sigma_s = 11$ (Fig. 5.3.i) while non-iterative tensor voting fails. Note that all tensors are accurately oriented for all $\sigma_s = 5, 8$ and 11 .

5.3.3 Statistics on the influence of iterations

A very basic 3x3 array of points, inspired from [93] and shown in Fig. 5.4.a, is used to evaluate the effect of iterations on tensor voting. Vertical and horizontal distances between points are denoted by Δx and Δy respectively. In the following, Δx will be chosen smaller than Δy . In such case, following the Gestalt rule of proximity, points have to be grouped vertically (in opposition if $\Delta x > \Delta y$, points would have to be grouped horizontally). A quality measure of good tensor orientation can be defined as:

$$Q_1 = -\log_{10} \left[\frac{1}{9} \sum_{i=1}^9 \left(1 - \frac{T_i(1,1)}{S_i} \right) \right] \quad (5.2)$$

where i indexes the 9 points of the array. T_i is the tensor of the point i , S_i is the sum of eigenvalues of T_i and $T_i(1,1)$ the vertical component of the tensor T_i .

As vertical lines have to be extracted, tensors are correctly oriented if they have a form close to $T_i = S_i \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. In such case $\sum_{i=1}^9 (1 - \frac{T_i(1,1)}{S_i})$ is close to zero, providing a high value for Q_1 . It can be considered that tensors are misoriented and that the extraction has failed when $Q_1 < 1$ (in such case it can be derived from Eq. 5.2 that the mean error of orientation for stick tensors is larger than 18 degrees). Also when $Q_1 > 2$, tensors can be considered as well orientated and the structure as correctly extracted (the mean error of orientation is then smaller than 6 degrees).

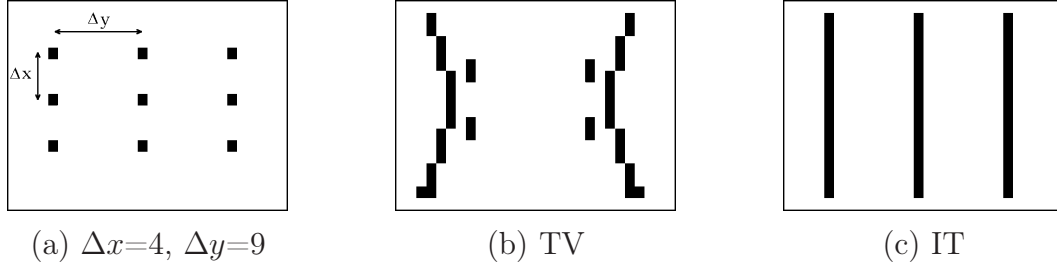


Figure 5.4: Systematic evaluation of the influence of iterations using a three by three array of points. **a.** Original array of points separated by $\Delta x=4$, $\Delta y=9$ **b.** Contours extracted with TV: central points are not grouped, lateral points are grouped but not in strictly vertical lines, moreover some artifacts appear ($Q_1=0.40$). **c.** The structure is well extracted by IT (10 iterations): points are grouped in vertical lines ($Q_1=2.38$). This example corresponds to the case 1 described in the following. $\sigma_s = 10$ and τ is the same for both insets b and c. Note that, because σ_s is larger than Δy and Δx , σ_s is too large to discriminate adequately between horizontal and vertical grouping in only one sparse voting iteration.

5.3.4 Results

Figure 5.6 presents the results obtained for different parameters Δx , Δy and n (number of iterations). For all cases the scale factor σ_s is fixed to 10. Again, it is necessary to remark that in this study we are only considering cases where $\Delta x < \Delta y$.

- **Case 1:** If $\Delta y \simeq \sigma_s$ (that is if $8 < \Delta y < 12$) TV fails to extract the curves. In contrast, iterations progressively improve the extraction quality to obtain the correct structure. This can be observed in Fig. 5.6.a in the center part, Fig. 5.6.b in the left part and also Fig. 5.4.a, b and c.
- **Case 2:** Similarly if $\Delta x \simeq \Delta y$ TV fails. There is conflict between horizontal and vertical grouping, yielding to extracted curves of different orientations. The use of iterations permits to group them adequately (see Fig. 5.5 and Fig. 5.6.c left part). Note that, for the most difficult cases, a high number of iterations can be necessary (50 iterations for the configuration with $\Delta x=13$, $\Delta y=14$).
- **Case 3:** If $\Delta x < \Delta y \ll \sigma_s$ it is impossible to extract the structure even if more iterations are deployed (see Fig. 5.6.a left part), the scale factor is indeed too large to be selective enough between horizontal and vertical grouping.
- **Case 4:** TV is accurate only if $\sigma_s \ll \Delta y$: Δy is sufficiently large in comparison to σ_s (such that the algorithm is able to segregate between vertical and horizontal

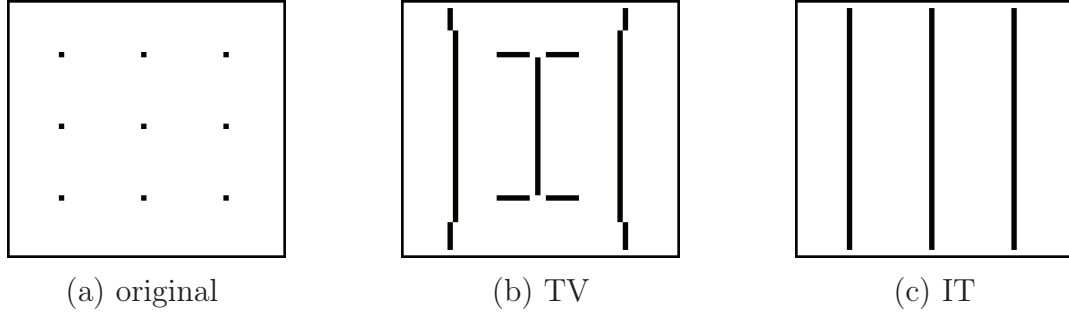


Figure 5.5: **a.** Array with $\Delta x=13$, $\Delta y=15$. **b.** Results for TV. **c.** Results for IT (10 iterations). Those results correspond to the case 2, details are given in the text.

grouping) and if there is no intrinsic conflict between vertical and horizontal grouping (that is where Δy and Δx are sufficiently different, i.e. $\Delta x \ll \Delta y$). Even in those cases the measurement Q_1 shows an improvement on the extraction for any increase of the number of iterations.

In conclusion, only if the features to be extracted are sufficiently simple, such that no conflict between different orientations of grouping appear, the non-iterative algorithm would suffice for correctly extracting image features. For more complicated cases, when some conflict between grouping orientations is present or when the scale factor σ_s is not precisely adjusted, more than two iterations are required. Moreover, it has been shown that in almost all cases iterations do not impair the quality of the results. Rather iterations refine the final structures even in simple cases.

5.3.5 Further remarks on the use of iterations

The number of iterations can be considered as an additional parameter of the algorithm. A procedure could also be implemented for stopping the iterations when the results do not differ much from one iteration to the next [226]. For all examples presented here a fixed number of iterations has been used. Generally, if data does not contain special difficulties (e.g. competition between directions of grouping or scale parameter disadjustment), a small number of iterations, i.e. between 6 to 10 iterations, has been shown to be sufficient.

The computational cost of deploying additional iterations is small in comparison to the total cost. The reason is that iterations are done only during the sparse voting step which computational costs are highly reduced in comparison with the dense voting step.

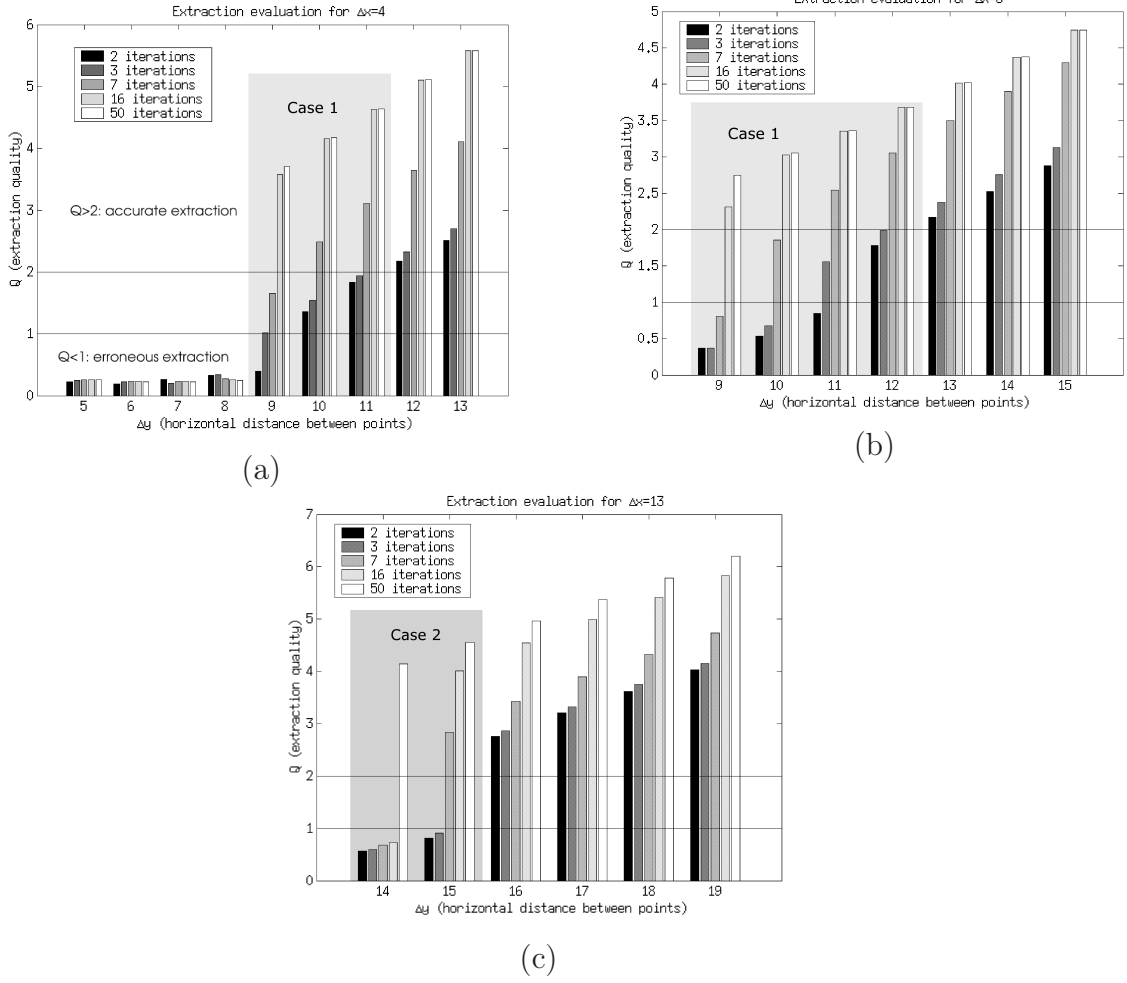


Figure 5.6: Extraction quality of the Fig. 5.4 for different parameters Δx and Δy . The number of iterations n is indicated by different gray shades in the bars (two iterations bar corresponds to TV which uses two voting steps). $\sigma_s = 10$ and the threshold τ are held constant for the entire experiment. $Q_1 < 1$ indicates the extraction has failed and $Q_1 > 2$ that it is accurate. **a.** Δx fixed to 4 and $5 \leq \Delta y \leq 13$. **b.** $\Delta x=8$ and $9 \leq \Delta y \leq 15$. **c.** $\Delta x = 13$ and $14 \leq \Delta y \leq 19$.

The problem of scale parameter adjustment could alternatively be solved by multi-scale analysis as recently described by Tong et al. [201].

5.4 Curvature improvement

5.4.1 Method

In addition to the iterative improvement we also propose a novel mechanism for extracting and incorporating curvature information. This last improvement introduces a curvature calculation and modified stick voting fields.

The curvature is evaluated in each voting point by averaging over all receiver points the curvature calculation ρ already computed by the classical tensor voting. In the classical tensor voting, a voter A *votes* by adding in any point B a tensor of an amplitude described by the following stick voting field:

$$V(A, B) = \exp \left(-\frac{s(A, B)^2}{\sigma_s^2} - \frac{\theta(A, B)^2}{\sigma_\theta^2} \right) \quad (5.3)$$

where $s(A, B)$ is the length of the circular arc AB which is tangent to $\vec{e}_1(A)$ in A (see Fig. 5.7.a). θ is the angle between the vectors $\vec{e}_1(A)$ and \vec{AB} . The scale factor σ_s and σ_θ are constants. This type of voting field has been described e.g. in [77]. Fig. 5.3.b, e, h shows the contours of those voting fields for different values of σ_s .

Eq. 5.4 expresses the geometric relation between the curvature $\rho(A, B)$, the length $s(A, B)$ of the arc AB and the angle θ (see Fig. 5.7.a):

$$s(A, B) = \frac{2 \theta(A, B)}{\rho(A, B)} \quad (5.4)$$

Thanks to Eq. 5.4, the Eq. 5.3 can be expressed as a function of $\rho(A, B)$:

$$V(A, B) = \exp \left(-s(A, B)^2 \left(\frac{1}{\sigma_s^2} + \frac{\rho(A, B)^2}{4 \sigma_\theta^2} \right) \right) \quad (5.5)$$

Another field described by Eq. 5.6 and Fig. 5.7.c is also frequently used by Medioni and coworkers [129]:

$$V(A, B) = \exp \left(-\frac{s(A, B)^2 + c \rho(A, B)^2}{\sigma_s^2} \right) \quad (5.6)$$

Nevertheless it is not employed in this study: the former field of Eq. 5.3 has the advantage that the angular range σ_θ of the field does not require to be readjusted when σ_s is changed, whereas the value of c in Eq. 5.6 has to be manually adjusted for each σ_s value.

To permit inflexion points and changes of curvature, the curvature of any voter A is calculated separately for both half planes $P_+(A)$ and $P_-(A)$ defined respectively by $P_+(A) = \{B \in P, (\vec{e}_1(A), \vec{AB}) > 0\}$ and $P_-(A) = \{B \in P, (\vec{e}_1(A), \vec{AB}) < 0\}$, where P denotes the whole image plane (see Fig. 5.7.f). The weighted average over each half plane gives $\gamma_i(A)$ (where $i = +$ or $-$), which is a curvature evaluation at the point A :

$$\gamma_i(A) = \frac{\sum_{B \in P_i(A)} (\lambda_1(B) - \lambda_2(B)) V(A, B) \rho(A, B)}{\sum_{B \in P_i(A)} (\lambda_1(B) - \lambda_2(B)) V(A, B)} \quad (5.7)$$

where $\lambda_1(B)$, $\lambda_2(B)$ are the eigenvalues of the tensor B . The weighted average is very similar to the “voting” used in tensor voting: the same weighting functions composed by the voting fields V and the stick saliency $\lambda_1 - \lambda_2$ are used.

The γ_i determined at one iteration, can then be used in the next iteration for modifying the stick voting fields, that is accomplished by extending Eq. 5.5 as follows (Note that the sign of $\gamma_i(A)$ has to be changed from one iteration to the next if the orientation of the eigenvector e_1 varies more than $\frac{\pi}{2}$.):

For any $B \in P_i(A)$,

$$V(A, B) = \exp \left(-s^2 \left(\frac{1}{\sigma_s^2} + \frac{(\rho(A, B) - \gamma_i(A))^2}{4 \sigma_\theta^2} \right) \right) \quad (5.8)$$

Some examples of such curvature-modified voting fields are shown Fig. 5.7. e, f, g (for a comparison see the classical voting field in Fig. 5.7.b).

The ball components of significant level denotes the presence of a junction or an uncertainty on the tangent orientation. In both cases the tangent estimation given by the first eigenvector can not be trusted, and as a consequence the curvature calculation is not reliable. It is then preferable not to use curved voting fields that is to consider γ_i as zero. Thus for all the following simulations we set $\gamma_i(A) = 0$ for any point A which verifies $\frac{\lambda_1(A)}{\lambda_2(A)} < 3.5$ (threshold empirically chosen).

The method follows the “voting” methodology. The curvature is computed by averaging weighted by the same voting fields V and saliencies λ_1 , λ_2 as in tensor voting. It should help to preserve the good properties of the tensor voting, like the robustness to noise. The curvature improvement does not entail an important additional computational cost in comparison to the classical method. It uses the same kind of averaging operations as the tensor voting and reuses calculations already done. E.g. in the curvature calculation of Eq. 5.7 all variables λ_1 , λ_2 , V and ρ are already computed by the classical tensor voting.

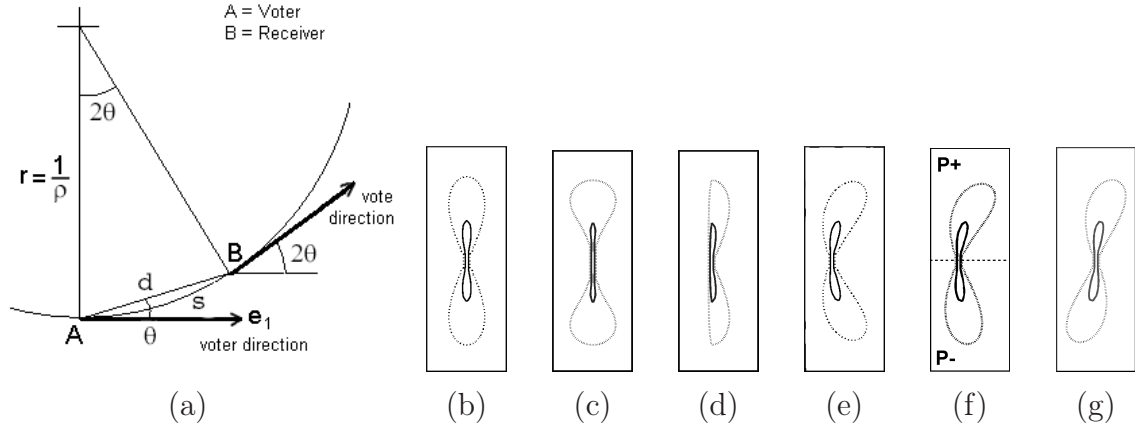


Figure 5.7: **a.** Tensor voting fields are built by calculating the distance d , the angle θ , the arc longitude s and the curvature ρ between the voter A oriented by its first eigenvector e_1 and the receiver B . In CI the curvature is evaluated in the voter A by averaging ρ over all receivers. **b.** Classical voting field without curvature. Contours are drawn at 50% and 5% of the maximum value, $\sigma_s = 15$ for all voting fields of the figure and $\sigma_\theta = \pi/6$ for insets b,d,e,f,g. **c.** Stick voting field defined by Eq. 5.6 (with $\sigma_s = 15$ and $c = 210,000$). **d.** TM voting field modified by the sign of the curvature [195]. **e.** Symmetric curved voting field with curvatures $\gamma^+ = \gamma^- = .06$. **f.** Curved voting field with different curvature in both half planes, $\gamma^+ = .09$ and $\gamma^- = .03$. **g.** Curved voting field with inflexion, $\gamma^+ = .06$ and $\gamma^- = -.06$.

Note also that an increased number of iterations is necessary to refine the curvature calculation. Six iterations have been seen to be sufficient unless the structure to extract presents some special complexity (e.g. competition between directions of grouping).

In all examples and both for iterated and curvature improvements, the iterative process has been seen to converge. Nevertheless it is not to exclude that in particular data configurations the process would not converge but for example oscillate between two states. Then, and in spite of the usually good behavior of the iterations, no proof of convergence is provided here.

In the following, the curvature improvement (CI) will be compared with the non-iterative tensor voting (TV) and with Tang and Medioni's method taking into account the sign of curvature (TM) [195]. In this last method, that has been developed for 3D images, the sign of curvature is employed to decide if the entire voting field or only half of it should be used (see Fig. 5.7.d for an example of a half voting field). Note one more sparse voting step has to be implemented to make the curvature calculation possible. (Results of the Tang and Medioni's method presented here are obtained by our own

implementation adapted for 2D images).

5.4.2 Statistical study

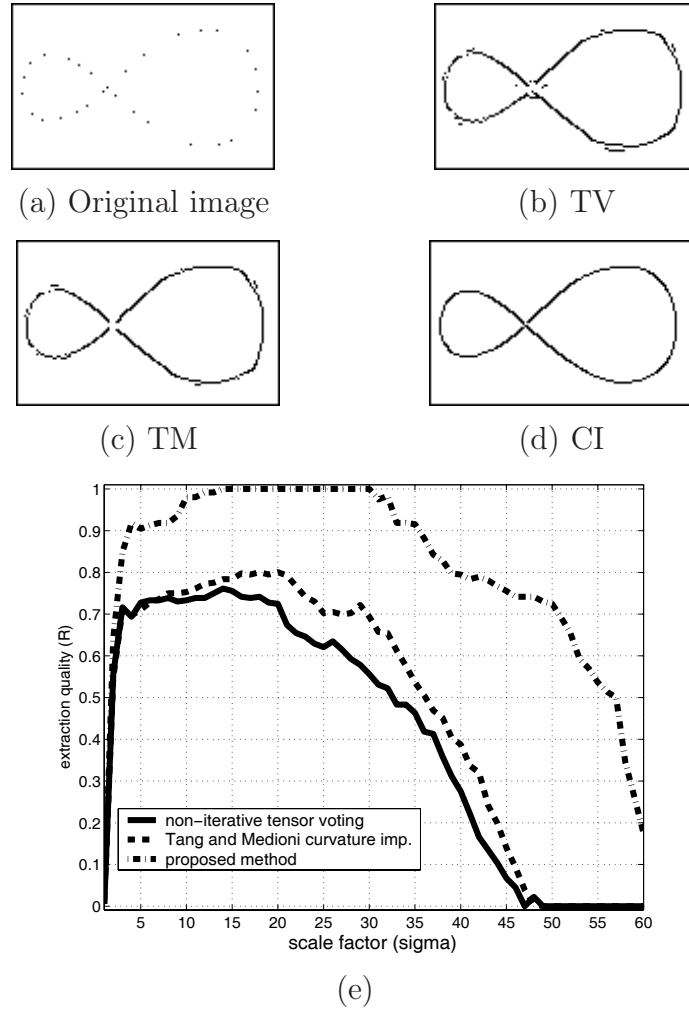


Figure 5.8: Comparison between methods. **a.** The input image is composed by a sparse set of dots dispersed along a curve in form of 8. **b., c. and d.** Extraction results with, respectively, TV, TM [196], and CI (6 iterations). In insets b., c. and d. $\sigma_s = 20$ and τ is adjusted for obtaining the best result following the Q_2 calculation. In all cases the curve is extracted, but TV and TM produces artifacts. CI produces the best result i.e. without artifacts and with smooth curves. **e.** Results for σ_s varying between 1 and 60.

Fig. 5.8.a shows an image composed by sparse points randomly chosen along an

eight-shape curve defined by the following equation:

$$\begin{cases} x = \alpha(t) \cos^2(t) \\ y = \alpha(t) \cos^2(t) \sin(t) \end{cases}, t \in [-\frac{\pi}{2}, \frac{3\pi}{2}[$$

$$\text{with } \alpha(t) = \begin{cases} 75 & \text{if } t \in [-\frac{\pi}{2}, \frac{\pi}{2}[\\ -41 & \text{if } t \in [\frac{\pi}{2}, \frac{3\pi}{2}[\end{cases} \quad (5.9)$$

75 and 41 are respectively the width in pixels of the right and the left loops of the curve.

In Fig. 5.8.a the distance between points vary between 2 to 26 pixels. This example is used for comparing the three versions of the algorithm. For different values of the scale factor σ_s , the extraction quality Q_2 is measured following:

$$Q_2 = \frac{TP - FN}{NP} \quad (5.10)$$

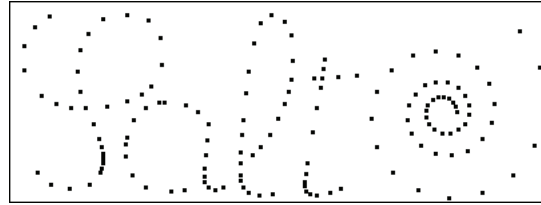
where TP (True Positive) is the number of points correctly extracted along the curve tolerating a deviation of one pixel from the ideal curve defined by Eq. 5.9, FN (False Negative) is the number of points erroneously extracted outside the curve (i.e. artifacts, outliers), and NP (Number of Positive) is the total number of points along the ideal curve (Eq. 5.9). If $Q_2 = 1$ the extraction is exact and if $Q_2 = 0$ the extraction is totally wrong, which means that there are more artifacts than points on the curve. For each experiment the threshold τ has been automatically adjusted in order to obtain the best results (i.e. the highest Q_2 value). Fig. 5.8.b-d shows results for the 3 methods with $\sigma_s = 20$.

All versions of the algorithm require a σ_s value higher than a minimum threshold for extracting the curve ($\sigma_s \geq 3$ in the present case). A smaller value of σ_s does not group the points together. On the other hand, the number of outliers increase for large values of σ_s . For TV, the number of misplaced points increases significantly ($Q_2 < 0.7$) when $\sigma_s > 20$. Similar level of outliers appear with TM when $\sigma_s > 30$, whereas for CI the quality does not fall off before $\sigma_s > 50$.

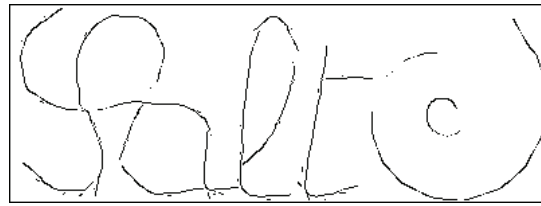
TV and TM adequately extracts the contours, but always with artifacts, for σ_s between 3 and 20 and between 3 and 30, respectively. CI provides similar results (that is $Q_2 > 0.7$) for σ_s between 3 and 50, giving in addition a better results for any σ_s value. Moreover it also yields smoother slopes. It is also remarkable that no artifact appear in the present example for any σ_s between 15 and 30.

In conclusion, CI yields smoother curves, produces much less artifacts and outliers, and increases the range of σ_s value in which the extraction is accurate. Moreover those advantages are obtained through a small additional computational cost.

5.4.3 Hand-written text example



(a) Original



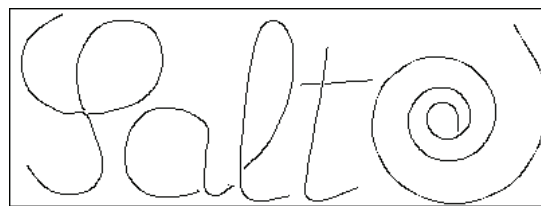
(b) TV



(c) TM



(d) IT



(e) CI

Figure 5.9: A hand drawn example. **a.** Test image “Salt”. For a better visualization points are magnified by a factor of 3. **b.** Extraction results with TV. **c.** with TM (our implementation). **d.** with IT (10 iterations). **e.** with CI (10 iterations).

Fig. 5.9 shows another example of contour extraction with the four versions of the algorithm: TV, TM, IT with 10 iterations and CI (also with 10 iterations). The 147x439

pixel image “Salt” (Fig. 5.9.a) is composed of sparse points along handwritten characters and along the spiral described in the first example Fig. 5.3. Same parameters are used for each method ($\sigma_s = 15$, $\sigma_\theta = \pi/6$ and τ are the same for all experiments).

Small improvements are observed using TM instead of TV (Fig. 5.9.c). The reason of this improvement could reside in part in the additional voting step that TM implements, since each additional voting step helps in orienting the tensors.

IT (Fig. 5.9.d) closes the curves better than TV and TM (Fig. 5.9.b and Fig. 5.9.c). Thus, it appears that even in such a case which do not seem to contain competitions between orientations of grouping, iterations are required to refine the orientations of the tensors.

With CI (Fig. 5.9.e) extracted contours of the curves are much smoother, and the level of artifacts is highly reduced. As a consequence, CI extracts the spiral much more accurately as TV and IT methods in Fig. 5.3.

For regions with straight segments and junctions both CI and IT behaves similarly. Hence, CI does not impair the results even for data which does not include curves. This allows then to use CI for any kind of images.

5.5 Robustness to noise

5.5.1 Evaluation methods

Noise robustness is evaluated on images of the type presented in Fig. 5.8 where different noise densities are added. Moreover the average gap g between data points varies between 2 and 20 points (see Fig. 5.10.a, b, c). σ_s is fixed to 20 for all experiments.

For each value of g , the number of noise points $N(g)$ supported by the methods is evaluated through the quality measurement Q_2 (Eq. 5.10). N is defined as the number of noise points for which $Q_2 = 0.5$ in the mean. For each value of g , 250 successive experiments with a different number of noise points R are realized. Results are plotted as points in the space (Q_2, R) , then the best line $(l_g)Q_2 = aR + b$ fitting the experimental points is extracted by linear regression and N is finally determined at half the high (i.e. for $Q_2 = \frac{1}{2}$).

Following the usual notation of Medioni and coworkers papers, the noise robustness

will be first expressed as the ratio M between the number of supported noise points N over the number of data points D (that is expressed as a percentage of noise points):

$$M = \frac{N}{D} \quad (5.11)$$

5.5.2 Measurement based on average gaps

We argue that the M measurement is not really appropriate since data points are concentrated along a 1D curve while noise points are dispersed throughout the whole 2D plane. Moreover it has been shown by Kovacs and coworkers [100] that the critical factor for humans in detecting a contour embedded in noise is the relative distance between elements of the contour and elements of the noise. We will then express the noise robustness as M' , the relative of average gap (or distance) between data points and between noise points.

Note first that the average gap (expressed in units of length) can not be the inverse of the density since the density of noise is expressed in points per surface unit. A detailed calculation of M' is given as follows:

Let be P the image plane and S the number of points of P . Let's consider for each noise point C_i the region P_i defined as the set of points that are closer to C_i than to any other noise point C_j (i.e. the Voronoi cell [49]). The mean area s of the P_i regions is:

$$s = \frac{S}{N} \quad (5.12)$$

Approximating P_i as a disk of area s , we obtain h , the mean gap between two noise points as the diameter of such disk:

$$s = \pi \left(\frac{h}{2} \right)^2 \quad (5.13)$$

From Eq. 5.12 and Eq. 5.13 we deduce:

$$h = 2\sqrt{\frac{S}{\pi N}} \quad (5.14)$$

Let l be the number of points of the curve to extract. The mean gap g between data points is directly defined by:

$$g = \frac{l}{D} \quad (5.15)$$

The average gap between data point over the average gap between noise points is then:

$$M' = \frac{g}{h} = \frac{l}{2} \sqrt{\frac{\pi}{S}} \frac{\sqrt{N}}{D} \quad (5.16)$$

If l is unknown M' should be expressed proportionally to $\frac{\sqrt{N}}{D}$:

$$M' \propto \frac{\sqrt{N}}{D} \quad (5.17)$$

Note that M' is a function of $\frac{\sqrt{N}}{D}$ whereas M is a function of $\frac{N}{D}$. The square root is suited to appropriately take into account the 2D repartition of noise points in relation with the 1D repartition of data points.

5.5.3 Results on noise robustness

Results presenting M versus g are shown in Fig. 5.10.d. Note first that M dramatically depends of the average gap between points (g): when gaps are small ($g=2$) the robustness M is very high (around 300% i.e. the methods are able to extract the curves in presence of 3 noise points for each data point) and it is decaying strongly when g gets higher (around 200% for $g=4$ and 20% for $g=12$, that is five time less noise than data). This represents a variation in the noise robustness by a factor of 15. M' plotted in Fig. 5.10.e shows more consistent results under variation of the average data gap g . $M' \in [0.28, .65]$ for all methods and for any $g \in [2, 12]$, i.e. the methods are able to extract the curves if the average noise gap h is from 1.5 to 3 times larger than the average data gap g . The variation of the noise robustness is only a factor of 2 which is more realistic considering the intuitive difficulty of extracting the curve in the situations presented Fig. 5.10. a, b, c.

Kovacs and coworkers [100] shown that humans are able to extract the contour formed by unoriented elements only if the distance between data elements is larger than the distance between noise elements (that is $M' \leq 1$). The M' measurement permits then a comparison with the performance of humans. The best rate is 0.55 for the classical tensor voting and 0.65 for the curvature improvement, to be compared with 1 for humans.

A related trouble of the measurement M is that it could let think that tensor voting methods are able to extract the structures where there is more noise than reliable data. That is true whether we forget that the image is inhomogeneous, e.g. in the image Fig. 5.10.a there are 3 times more noise points than data points. But this does not take

into account that the data points are concentrated in the curves while noise points are dispersed. Indeed the measurement M' shows (see Fig. 5.10.e.) that the average gap between data points need to be at least 1.5 times lower than the average gap between noise points to extract the structure. Hence it appears that the tensor voting methods succeed only if there is significantly more data points than noise points around the curves to extract.

It is not surprising that results provided by all methods decay when gaps increase because there are less points for "voting" (the noise robustness is based on the "voting" process which is similar to an averaging). Results also decrease for g small (i.e. smaller than 4 to 7, depending on the method) what could be explained by the discrete grid: for highest data densities, the rounding off positions complicate the extraction of tangents.

The experiments also demonstrate that IT and CI yield similar and even slightly better results than TV and TM methods, particularly when the average gap between data points is larger. Even for low density of data ($g = 20$), CI is still able to deal with a noise gap h 2.5 times larger than the data gap g , whereas TV fails to extract the curve for any $g \geq 15$. As a conclusion, the described improvements do not impair the noise robustness nor the ability to discard outliers of the tensor voting method but even improve them.

5.6 Conclusion

This study shows first that iterations are useful for tensor voting, particularly for extracting correct contours in difficult or ambiguous situations like competition between grouping orientation or scale parameter misadjustment. The iterations refine and improve the extraction quality yielding to more accurate results.

Further experiments extending the method by the incorporation of curvature information show better results especially for curved features. This extension provides smooth curves and reduces the level of artifacts, besides the fact that it also increases the robustness of the method to scale parameter misadjustment.

The proposed improvements have low computational cost, and they preserve and enhance the good robustness to noise of the original method.

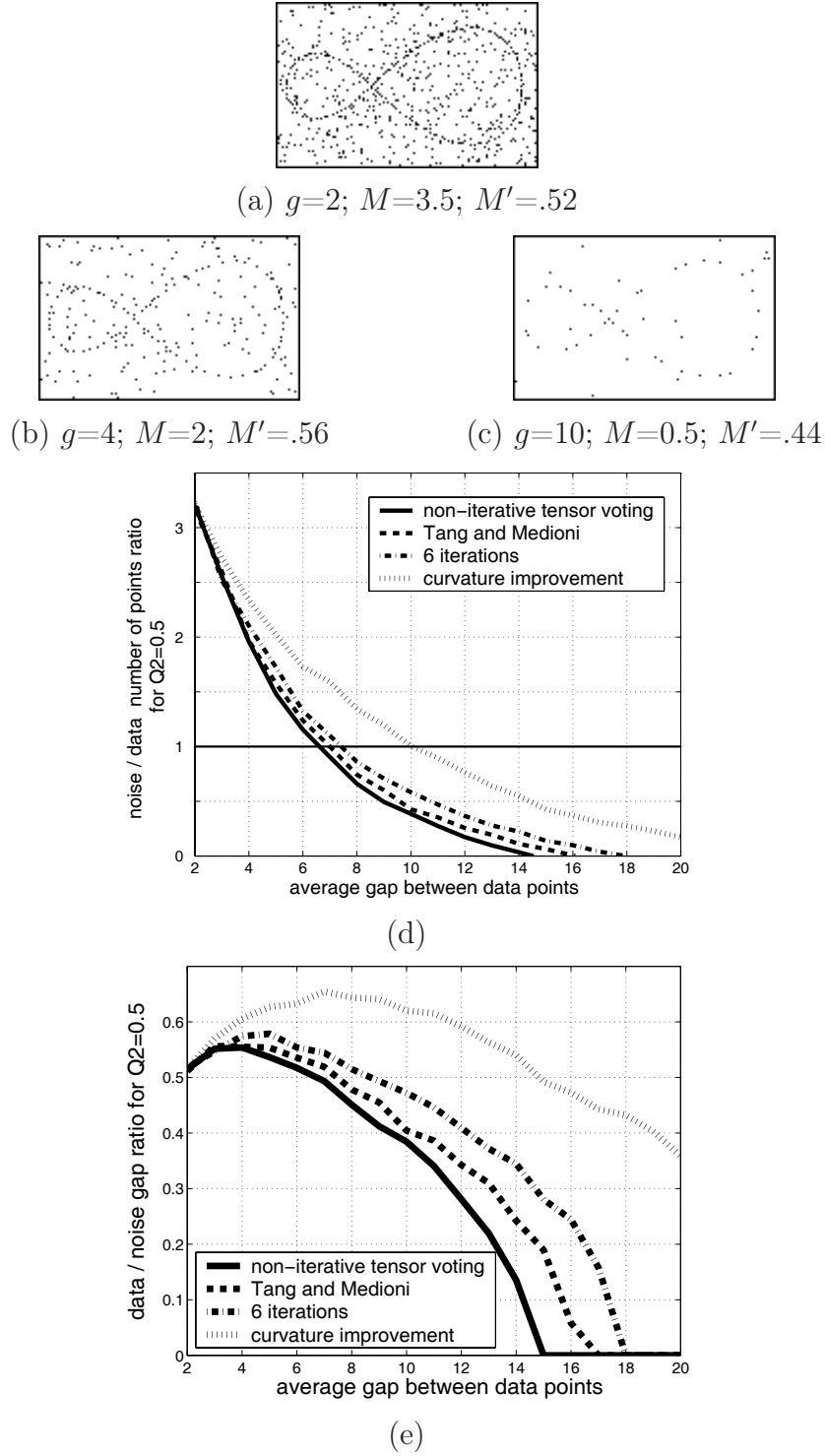


Figure 5.10: Ratio of noise supported by the different methods of tensor voting. **a.** Image with 2 pixels gaps between data points ($g=2$) and $M=3.5$ noise points for each data point. **b.** $g=4$ and $M=2$. **c.** $g=10$ and $M=0.5$. Images a, b and c contain the maximum quantity of noise for which algorithms are able to extract the structure. **d.** Results showing the ratio of noise points (M) supported for each method. **e.** Ratio of the average gaps (M') supported by each method.

Chapter 6

Conclusion

"Nous sommes perdus si nous oublions que les fruits sont à tout le monde et que la terre n'est à personne."

Jean Jacques Rousseau

The present thesis proposes new contributions for overcomplete transformations inspired by the functional architecture of the primary visual cortex. On one hand the efficiency of overcomplete transformations in image processing applications and on the other hand the difficulty to optimize with systematic methods overcomplete and non-linear transformations, conduced to build the proposed multiresolution on available knowledge about psychophysics of the perception, physiology and models of primary visual cortex and statistics of natural images. We show how studies of these fields can serve as guidelines for the construction of highly efficient image processing multiresolution transforms.

The first part of this thesis (chapter 2) proposes a self-invertible log-Gabor wavelet transform which mother wavelet is inspired from the classical receptive field of V1 simple cells; and the scale/orientation arrangement is also consistent with the knowledge on simple cell populations. This transform is highly sensitive to edges and lines of the natural images and show remarkable good results in feature to noise segregation by hard thresholding in comparison to other multiresolution transforms often used in denoising (wavelets, translation invariant wavelets, steerable pyramids).

The second part of the thesis (chapter 3) presents an edge extraction method as a synthesis of existing energy models, most of them being already biologically inspired. The advantages of the proposed method is that it detects both lines and edges, what for example the popular Canny edge extractor is not able to do; and it permits to recon-

struct the image from the extracted edges through the sparse coding scheme developed in the chapter 4 of this thesis.

The chapter 4 proposes a new sparse coding scheme based on the edge extraction framework proposed in chapter 3. We propose the hypothesis that the most important image features can be coded by solely the transform coefficients located at local maxima of energy, that is the edges and lines. The residual, composed of the coefficients located out of the local maxima, is projected on the local maxima coefficients. The image can then be accurately reconstructed from a small subset of the coefficients. The model codes particularly adequately the contours, but can be blind to some textures and not deal optimally with crossing or junctions. Therefore a separate processing for such features could be used advantageously to complete the present scheme, although some few iterations of the sparse coding are also able to overcome the problem.

The sparse coding by contour extraction brings with him the important advantage that the sparse coefficients are located along smooth continuous contours. This allows to apply an additional transform for a further decorrelation of the image information. It consists in a *chain coding* which encodes predictively the position and the value of the coefficients inside the transform domain, yielding higher compression rates.

The chapter 5 is dedicated to contour grouping, that is to form continuous contours from separated edge segments. Such method uses the *tensor voting* framework which is based on Gestalt psychology. We propose improvements consisting in the deployment of iterations and the use of an evaluation of the curvature for improving the integration of contours.

As a conclusion, the present thesis proposes new contributions in multiresolution transform which could be used as alternatives to the most popular families of wavelets or oriented bandpass transforms for low-level processing of natural images. Although the transform is overcomplete, important sparsity can be achieved through the use of the proposed dedicated sparse coding method. Thus, the transform benefits both for a excellent specificity to features of natural image and for sparseness (low information redundancy), it is then suitable for many usual low-processing tasks such as denoising, feature extraction, image fusion, image compression.

6.1 Further work

The present thesis open many new lines of research in the different aspects of the implementation. The self invertible log-Gabor wavelet transform could be optimized by a more systematic test of the free parameters (bandwidth in scale and orientation, number of orientations). Also since physiological study states that there exist cells with different bandwidth ranges [40], it could be implemented together various family of filters, i.e. various mother wavelets, with different bandwidths and frequency/orientation arrangement.

The edge extraction method could be improved by the incorporation of endpoints detectors to refine the extraction in proximity of junctions, crossing and line-endings, as proposed in [83]. For edge extraction under noisy conditions, a rigorous probabilistic approach could permit a better segregation between existing contours and noise points aligned by chance. The tensor voting approach could be extended to the log-Gabor wavelet framework for binding together separated contour segments in applications of occluded contour completion.

The sparsification algorithm could be upgraded by the development of non iterative projection techniques and the optimization of the choice of regularly spaced coefficients along the contours. Moreover a separated texture encoding should advantageously be implemented.

Promising open issues lie on further decorrelation by predictive contour encoding. The improvement of the actual chain coding can follow two directions: first the encoding of whole contour parts from its endpoints and tangent and curvature parameters. Second by a predictive coding of the scales exploiting the high degree of similarity of the features across scales. Such improvements should yield much better compression rates and should offer efficient descriptors for object recognition.

Bibliography

- [1] E.H. Adelson and J.R. Bergen. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am.*, 2:284–321, 1985.
- [2] T.D. Albright and G.R. Stoner. Contextual influences on visual processing. *Annu. Rev. Neurosci.*, 25:339–379, 2002.
- [3] J.J. Atick. Could information theory provide an ecological theory of sensory processing? *Network*, 3:213–251, 1992.
- [4] F. Attneave. Some informational aspects of visual perception. *Psychological Review*, 61:183–193, 1954.
- [5] J.F. Aujol and B. Matei. Structure and texture compression. Technical report, INRIA, Research Report 5076, January 2004.
- [6] S. Barber and G.P. Nason. Denoising real images using complex-valued wavelets. In K.V. Mardia R.G. Aykroyd and M.J. Langdon, editors, *Stochastic Geometry, Biological Structure and Images*, pages 91–95. Department of Statistics, University of Leeds, 2003.
- [7] H.B. Barlow. The coding of sensory messages. In W. H. Thorpe and O. L. Zangwill, editors, *Current Problems in Animal Behavior*, pages 331–360. Cambridge University Press, 1961.
- [8] H.B. Barlow. Redundancy reduction revisited. *Network: Computation in Neural Systems*, 12:241– 253, 2001.
- [9] P. Bayerl and H. Neumann. Disambiguating visual motion through contextual feedback modulation. *Neural Computation*, 16(10):2041–2066, 2004.
- [10] M. F. Bear, B. W. Connors, and M. A. Paradiso. *Neuroscience: Exploring the Brain*. Lippincott Williams & Wilkins, Baltimore, USA, 2nd edition, 2002.
- [11] A.J. Bell and T.J. Sejnowski. The ‘independent components’ of natural scenes are edge filters. *Vision Research*, 37(23):3327–3338, 1997.
- [12] C. Blakemore and A. Tobin. Lateral inhibition between orientation detectors in the cat’s visual cortex. *Exp. Brain Res.*, 15:439–440, 1972.

- [13] W. Bosking, Y. B. S. Zhang, and D. Fitzpatrick. Orientation selectivity and the arrangement of horizontal connections in the tree shrew striate cortex. *J. Neuroscience*, 17:2112–2127, 1997.
- [14] R. W. Buccigrossi and E. P. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE Trans. Image Proc.*, 8(12):1688–1701, December 1999.
- [15] P.J. Burt and E.H. Adelson. The Laplacian pyramid as a compact image code. *IEEE Trans. Comm.*, 31(4):532–540, 1983.
- [16] J. Canny. A computational approach to edge detection. *IEEE Trans. PAMI.*, 8(6):679–698, 1986.
- [17] M. Carandini, J.B. Demb, V. Mante, D.J. Tolhurst, Y. Dan, B.A. Olshausen, J.L. Gallant, and N.C. Rust. Do we know what the early visual system does? *J. Neurosci.*, 25(46):10577–10597, 2005.
- [18] J.R. Cavanaugh, W. Bair, and J.A. Movshon. Nature and interaction of signals from the receptive field center and surround in macaque v1 neurons. *J. Neurophysiol.*, 88(5):2530–2546, 2002.
- [19] L. M. Chalupa and J. S. Werner. *The Visual Neurosciences*. MIT Press, Cambridge, 1th edition, 2003.
- [20] S. G. Chang, B. Yu, and M. Vetterli. Adaptive wavelet thresholding for image denoising and compression. *IEEE Trans. on Image Proc.*, 9(9):1532–1546, 2000.
- [21] S. G. Chang, B. Yu, and M. Vetterli. Spatially adaptive wavelet thresholding with context modeling for image denoising. *IEEE Trans. on Image Processing*, 9(9):1522–1531, 2000.
- [22] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM Jour. on Sc. Computing*, 20(1):33–61, 1999.
- [23] Y. Choe and R. Miikkulainen. Contour integration and segmentation in a self-organizing map of spiking neurons. *Biological Cybernetics*, 90:75–88, 2004.
- [24] Y. Choe and S. Sarma. Relationship between suspicious coincidence in natural images and oriented filter response distributions. Technical report, Texas A&M University, Department of Computer Science, 2003-8-4.
- [25] N. Chomsky. *Language and Problems of Knowledge: The Managua Lectures*. The MIT Press, 1987.
- [26] O. Christiansen. *Time-frequency analysis and its applications in denoising*. PhD thesis, Dept. of Informatics, University of Bergen, Norway, 2002.

- [27] D.A. Clausi and M.E. Jernigan. Designing Gabor filters for optimal texture separability. *Pattern Recognition*, 33:1835–1849, 2000.
- [28] R.R. Coifman and D. Donoho. Translation-invariant de-noising. In A. Antoniadis and G. Oppenheim, editors, *Wavelets and statistics, Lecture Notes in Statistics 103*, pages 125–150. Springer Verlag, NY, 1995.
- [29] R.R. Coifman and M.V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. Inf. Theory*, 38(2):713–718, 1992.
- [30] D. M. Coppola, H.R. Purves, A.N. McCoy, and D. Purves. The distribution of oriented contours in the real world. *Neurobiology*, 95(7):4002–4006, 1998.
- [31] G. Cristóbal, S. Fischer, M. Forero-Vargas, R. Redondo, and J. Hormigo. Texture segmentation and analysis using local spectral methods. In *Int. Symp. on Signal Proc. and its Applications, Paris*, pages 129–132, 2003.
- [32] G. Cristóbal and R. Navarro. Space and frequency variant image enhancement based on a Gabor representation. *Patt. Rec. Letters*, 15(3):273–277, 1994.
- [33] I. Daubechies. *Ten lectures on wavelets*. SIAM, Philadelphia, PA, 1992.
- [34] J. Daugman. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Res.*, 20:847–856, 1980.
- [35] J. Daugman. Spatial visual channels in the Fourier plane. *Vision Res.*, 24(9):891–910, 1984.
- [36] J. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A*, 2(7):1160–1169, 1985.
- [37] J. Daugman. Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression. *IEEE Trans. Acoust. Speech Signal Proc.*, 36(7):1169–1179, 1988.
- [38] R. Deriche. Using canny’s criteria to derive a recursively implemented optimal edge detector. *Int. Journal of Computer Vision*, 1(2):167–187, 1987.
- [39] U.Y. Desai, M.M. Mizuki, I. Masaki, and B.K.P. Horn. Edge and mean based image compression. Technical Report AIM-1584, Nagoya University, 27, 1996.
- [40] R. L. DeValois, D. G. Albrecht, and L. G. Thorell. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res.*, 22:545–559, 1982.
- [41] M.N. Do and M. Vetterli. The contourlet transform: An efficient directional multiresolution image representation. *IEEE Trans. on Image Proc.*, 14(12):2091–2106, 2005.

- [42] A. Dobbins, S.W. Zucker, and M.S. Cynader. Endstopping and curvature. *Vision Res.*, 29(10):1371–1387, 1989.
- [43] E. Doi and M. S Lewicki. Relations between the statistical regularities of natural images and the response properties of the early visual system. *Japanese Cognitive Science Society, SIG P&P*, pages 1–8, 2005.
- [44] D. Donoho. De-noising by soft-thresholding. *IEEE Trans. Inf. Theory*, 41(3):613–627, 1995.
- [45] D. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization. *Proc. Natl. Acad. Sci. USA*, 100(5):2197–2202, 2003.
- [46] D. Donoho and A.G. Flesia. Can recent innovations in harmonic analysis ‘explain’ key findings in natural image statistics. *Network: Computation in Neural Systems*, 12(3):371–393, 2001.
- [47] D.L. Donoho. Wavelet shrinkage and W.V.D. – A ten-minute tour. Technical Report 416, Stanford University, Department of Statistics, 1993.
- [48] G. Dreyfus, M. Samuelides, J. Martinez, M. Gordon, F. Badran, S. Thiria, and L. Hérault. *Réseaux de neurones - Méthodologies et applications*. Eyrolles, 2nd edition, 2004. ISBN : 2-212-11464-8.
- [49] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd ed.)*. Wiley Interscience, 2000.
- [50] M.P. Eckert and A.P. Bradley. Perceptual quality metrics applied to still image compression. *Signal Proc.*, 70:177–200, 1998.
- [51] J. H. Elder. Are edges incomplete? *Int. Jour. of Comp. Vis.*, 34(2-3):97–122, 1999.
- [52] J. H. Elder and A.J. Sachs. Psychophysical receptive fields of edge detection mechanisms. *Vision Res.*, 44(8):795–813, 2004.
- [53] D.J. Field. Relation between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4(12):2379–2394, 1987.
- [54] D.J. Field. What is the goal of sensory coding? *Neural Computation*, 6(4):559–601, 1994.
- [55] D.J. Field, A. Hayes, and R. Hess. Contour integration by the human visual system: Evidence for a local ‘association field’. *Vision Research*, 33(2):173–193, 1993.
- [56] S. Fischer. Modelos perceptuales en la cuantificación y codificación de imágenes. Master’s thesis, ETSIT-UPM, 2000.

- [57] S. Fischer, P. Bayerl, H. Neumann, G. Cristóbal, and R. Redondo. Are iterations and curvature useful for tensor voting? In T. Pajdla and J. Matas, editors, *Proc. Int. Conf. on Artificial Neural Networks, ECCV 2004, LNCS 3023*, pages 158–169, 2004.
- [58] S. Fischer, P. Bayerl, H. Neumann, R. Redondo, and G. Cristóbal. Iterations and curvature improvement for tensor voting. *EURASIP: Image Processing*, submitted.
- [59] S. Fischer, G. Cristóbal, and R. Redondo. Sparse overcomplete Gabor wavelets based on local competitions: A model for simple cells? In *ECOVISION Workshop (Isle of Skye)*, 2004.
- [60] S. Fischer, G. Cristóbal, and R. Redondo. Sparse overcomplete Gabor wavelet representation based on local competitions. *IEEE Trans. on Image Proc.*, 15(2):265–272, 2006.
- [61] S. Fischer, L. Perrinet, R. Redondo, and G. Cristóbal. Efficient representation of natural images by local cooperation. In *European Conf. On Visual Perception, A Coruña, Spain*, august 2005.
- [62] S. Fischer, R. Redondo, and G. Cristóbal. Sparse edge coding using overcomplete Gabor wavelets. In *IEEE Int. Conf. on Image Proc.*, volume 1, pages 85–88, 2005.
- [63] S. Fischer, R. Redondo, L. Perrinet, and G. Cristóbal. Sparse Gabor wavelets by local operations. In G. Linan-Cembrano; R.A. Carmona, editor, *Proc. SPIE, Bioengineered and Bioinspired Systems II*, volume 5839, pages 75–86, 2005.
- [64] S. Fischer, R. Redondo, L. Perrinet, and G. Cristóbal. Sparse approximation of images inspired from the functional architecture of the primary visual areas. *EURASIP JASP, special issue on Image Perception*, in press.
- [65] S. Fischer, F. Sroubek, L. Perrinet, R. Redondo, and G. Cristóbal. Self-invertible 2d log-Gabor wavelets. *Int. Jour. of Comp. Vision*, submitted.
- [66] R.A. Frazor and W.S. Geisler. Local luminance and contrast in natural images. *Vision Res.*, page In Press, 2006.
- [67] H. Freeman. On the encoding of arbitrary geometric configurations. *IRE Trans. on Electronic Computers*, pages 260–268, 1961.
- [68] D. Gabor. Theory of Communication. *J. Inst. Electr. Eng.*, 93:429–457, 1946.
- [69] W.S. Geisler, J.S. Perry, B.J. Super, and D.P. Gallogly. Edge co-occurrence in natural images predicts contour grouping performance. *Vision Res.*, 41(6):711–724, 2001.
- [70] G. Gilboa, N. Sochen, and Y. Zeevi. Image enhancement and denoising by complex diffusion processes. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(8):1020–1036, 2004.
- [71] E. B. Goldstein. *Sensation and perception*. Brooks/Cole, 5th edition, 1999.

- [72] R. Gribonval and P. Vandergheynst. On the exponential convergence of matching pursuits in quasi-incoherent dictionaries. *IEEE Trans. on Information Theory*, 52(1):255–261, 2006.
- [73] C. Grigorescu, N. Petkov, and M. A. Westenberg. Contour detection based on nonclassical receptive field inhibition. *IEEE Trans. on Image Proc.*, 12(7):729–739, 2003.
- [74] M.H. Gross and R. Koch. Visualization of multidimensional shape and texture features in laser range data using complex-valued Gabor wavelets. *IEEE Trans. Visual. and Comput. Graphics*, 1(1):44–59, 1995.
- [75] S. Grossberg, E. Mingolla, and J. Williamson. Synthetic aperture radar processing by a multiple scale neural system for boundary and surface representation. *Neural Networks*, 8(7-8):1005–1028, 1995.
- [76] S. Grossberg and J.R. Williamson. A neural model of how horizontal and interlaminar connections of visual cortex develop into adult circuits that carry out perceptual groupings and learning. *Cerebral Cortex*, 11:37–58, 2001.
- [77] G. Guy and G. Medioni. Inference of surfaces, 3d curves, and junctions from sparse, noisy, 3d data. *IEEE Trans. P.A.M.I.*, 19(11):1265–1277, 1997.
- [78] A. B. Hamza, H. Krim, and G. Unal. Unifying probabilistic and variational estimation. *IEEE Signal Proc.*, 19(5):37–47, 2002.
- [79] T. Hansen, W. Sepp, and H. Neumann. Recurrent long-range interactions in early vision. In S. Wermter et al., editor, *Emergent neural computational architectures, LNAI 2036*, pages 127–138, 2001.
- [80] R.M. Haralick. Digital step edges from zero crossing of second directional derivatives. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6(1):58–68, 1984.
- [81] F. Heitger. Feature detection using suppression and enhancement. Technical Report TR-163, Communication Technology Laboratory, Swiss Federal Institute of Technology, Zürich, 1995.
- [82] F. Heitger, L. Rosenthaler, R. Von der Heydt, E. Peterhans, and O. Kubler. Simulation of neural contour mechanisms: from simple to end-stopped cells. *Vision Res.*, 32(5):963–981, 1992.
- [83] F. Heitger, R. Von der Heydt, E. Peterhans, L. Rosenthaler, and O. Kubler. Simulation of neural contour mechanisms: representing anomalous contours. *Image and Vision Computing*, 16:409–423, 1998.
- [84] R.F. Hess, A. Hayes, and D.J. Field. Contour integration and cortical processing. *J. Physiol. Paris*, 97(2-3):105–19, 2003.

- [85] J.C. Horton and D.L. Adams. The cortical column: a structure without a function. *Philos Trans R Soc Lond B Biol Sci.*, 360(1456):837–62, 2005.
- [86] P.G. Howard. The design and analysis of efficient lossless data compression system. Technical Report CS-93-28, Dept. of Computer Science (Brown University), 1993.
- [87] D. Hubel. *Eye, Brain, and Vision*. WH Freeman. Sc. Am. Lib. Series, New York, 1988.
- [88] D. Hubel and T. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *J. Physiology (London)*, 160:295–306, 1962.
- [89] A. Hyvarinen, P. Hoyer, and M. Inki. Topographic independent component analysis. *Neural Computation*, 13(7):1527 – 1558, 2001.
- [90] L. Iverson and S.W. Zucker. Logical/linear operators for image curves. *IEEE Trans. P.A.M.I.*, 17(10):982–996, 1995.
- [91] J. Jones and L. Palmer. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J. of Neurophysiology*, 58(6):1233–1258, 1987.
- [92] J.-K. Kamarainen, V. Kyrki, and H. Kalviainen. Invariance properties of gabor filter-based features - overview and applications. *IEEE Trans. Image Processing*, 15(5):1088–1099, 2006.
- [93] D.H. Kandel and R.H. Wurtz. Constructing the visual image (Chapter 25). In McGraw-Hill, editor, *Principles of neural science*, pages 492–506. Kandel, E. R. and Schwartz, J. H. and Jessell, T. M., 4th edition, 2000.
- [94] M.K. Kapadia, G. Westheimer, and C.D. Gilbert. Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *Jour. Neurophysiology*, 84(4):2048–2062, 2000.
- [95] G. Kayaert, I. Biederman, and R. Vogels. Shape tuning in macaque inferior temporal cortex. *Jour. of Neuroscience*, 23(7):3016–3027, 2003.
- [96] N.G. Kingsbury. Complex wavelets for shift invariant analysis and filtering of signals. *Jour. of Applied and Comput. Harmonic Analysis*, 10(3):234–253, 2001.
- [97] Z.F. Kisvarday, J.M. Crook, P. Buzas, and U.T. Eysel. Combined physiological-anatomical approaches to study lateral inhibition. *J. Neurosci.*, 103(1):91–106, 2000.
- [98] A. Koschan and M. Abidi. Detection and classification of edges in color images. *IEEE Signal Proc. Magazine*, 22(1):64– 73, 2005.
- [99] I. Kovacs. Gestalten of today: early processing of visual contours and surfaces. *Behav. Brain Res.*, 82(1):1–11, 1996.

- [100] I. Kovacs, U. Polat, and A. M. Norcia. Breakdown of binding mechanisms in amblyopia. *Invest. Opth. Vis. Sci.*, 37(3):670, 1996.
- [101] P. Kovesi. Phase preserving denoising of images. In *Australian Patt. Recog. Soc. Conf. DICTA. Perth WA.*, pages 212–217, 1999.
- [102] P. Kovesi. Phase congruency detects corners and edges. In *Australian Patt. Recog. Soc. Conf. DICTA. Sydney WA.*, pages 309–318, 2003.
- [103] V. Krüger. *Gabor wavelet networks for object representation*. PhD thesis, Christian-Albrechts-University Kiel, Technical Faculty, 2001.
- [104] N. Kruger. Collinearity and parallelism are statistically significant second order relations of complex cell responses. *Neural Proc. Letters*, 8:117–129, 1998.
- [105] N. Kruger, M. Pötzsch, and G. Peters. Principles of cortical processing applied to and motivated by artificial object recognition. In R. Baddeley, P. Hancock, and P. Foldiak, editors, *Information Theory and the Brain*, pages 223–228. Cambridge University Press, 1998.
- [106] W. Lawton. Applications of complex valued wavelet transforms to subband decomposition. *IEEE Trans. Signal Processing*, 41:3566–3568, 1993.
- [107] E. Le Pennec and S. Mallat. Sparse geometrical image approximation with bandelets. *Image Proc.*, 14(4):423–438, 2004.
- [108] T.S. Lee. Image representation using 2D Gabor wavelets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(10):959–971, 1996.
- [109] T.S. Lee. Computations in the early visual cortex. *J. Physiology*, 97:121–139, 2003.
- [110] G.E. Legge and J.M. Foley. Constrast masking in human vision. *J. Opt. Soc. Am. A*, 12:1458–1471, 1980.
- [111] X. Li and M.T. Orchard. Spatially adaptive image denoising under overcomplete expansion. In *Int. Conf. on Image Proc.*, volume 3, pages 300–303, 2000.
- [112] Z. Li. A neural model of contour integration in the primary visual cortex. *Neural Computation*, 10:903–940, 1998.
- [113] Z. Lin, A. Sluzek, and M.D. Saiful Islam. An adaptive edge preserving variational method for color image regularization. In S. Li, F. Pereira, H.-Y. Shum, and Tescher A. G., editors, *Proc. SPIE, Visual Communications and Image Processing*, volume 5960, pages 2034–2045, 2005.
- [114] M.S. Livingstone and D.H. Hubel. Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *J. Neurosci.*, 7:3416–3468, 1987.

- [115] K. Lorenz. *Evolution and Modification of Behavior*. University Of Chicago Press, Chicago, IL, 1st edition, 1965.
- [116] D.M. Malioutov, M. Çetin, and A.S. Willsky. Optimal sparse representations in general overcomplete bases. In *IEEE Int. Conf. on Acoustics, Speech, and Signal Proc., Montreal, Canada*, volume 2, pages II-793-796, 2004.
- [117] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1999. ISBN: 0-12-466606-X.
- [118] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. on Signal Proc.*, 41(12):3397-3415, 1993.
- [119] S. Mallat and S. Zhong. Characterization of signals from multiscale edges. *IEEE Trans. PAMI*, 14:710-732, 1992.
- [120] J. Malo, A. Felipe, M.J. Luque, and J.M. Artigas. On the intrinsic two-dimensionality of the CSF and its measurement. *Journal of Optics*, 25(3):93-103, 1994.
- [121] J. Malo and J. Gutiérrez. V1 non-linearities emerge from local-to-global non-linear ICA. *Network: Computation in Neural Systems*, accepted october 2005.
- [122] S. Mandon and A. Kreiter. Rapid contour integration in macaque monkeys. *Vision Res.*, 45(3):291-300, 2005.
- [123] R. Manduchi, P. Perona, and D. Shy. Efficient deformable filter banks. *IEEE Trans. on Sign. Proc.*, 46:1168- 1173, 1998.
- [124] S. Marcelja. Mathematical description of the responses of simple cortical cells. *J. Opt. Soc. Am. A*, 70(11):1297-1300, 1980.
- [125] D. Marr and E. Hildreth. Theory of edge detection. *Proc. R. Soc. London B*, 207:187-217, 1980.
- [126] D.R. Martin, C.C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Trans. P.A.M.I.*, 26(5):530-549, 2004.
- [127] L.M. Martinez and J.-M. Alonso. Complex receptive fields in primary visual cortex. *The Neuroscientist*, 8:317-331, 2003.
- [128] G.S. Masson. From 1D to 2D via 3D : surface motion segmentation for gaze stabilisation in primates. *Jour. of Physiology (Paris)*, 98:35-52, 2004.
- [129] G. Medioni, M.-S. Lee, and C.-K. Tang. *A computational framework for feature extraction and segmentation*. Elsevier Science, mar. 2000.
- [130] Y. Meyer. *Wavelets. Algorithms & applications*. SIAM, Philadelphia, PA, 1993.

- [131] E. Mingolla, W. Ross, and S. Grossberg. A neural network for enhancing boundaries and surfaces in synthetic aperture radar images. *Neural Networks*, 12(3):499 – 511, 1999.
- [132] J. Morlet, G. Arens, E. Fourgeau, and D. Girard. Wave propagation and sampling theory. *Geophysics*, 47:203–236, 1982.
- [133] M.C. Morrone and D.C. Burr. Feature detection in human vision: a phase-dependent energy model. *Proc. R. Soc. London B*, 235:221–245, 1988.
- [134] O. Nestares, R. Navarro, J. Portilla, and A. Taberner. Efficient spatial-domain implementation of a multiscale image representation based on Gabor functions. *Jour. of Electr. Imag.*, 7(1):166–173, 1998.
- [135] H. Neumann and E. Mingolla. Computational neural models of spatial integration in perceptual grouping. In T.F. Shipley and P.J. Kellman, editors, *From Fragments to Objects - Segmentation and Grouping in Vision*, pages 353–400. Elsevier Science, 2001.
- [136] M. Nicolescu and G. Medioni. Layered 4D representation and voting for grouping from motion. *IEEE Trans. P.A.M.I.*, 25(4):492–501, 2003.
- [137] F. Nietzsche. *Twilight of the Idols*. 1888.
- [138] J. F. Norman, F. Phillips, and H. E. Ross. Information concentration along the boundary contours of naturally shaped solid objects. *Perception*, 30:1285–1294, 2001.
- [139] B. Olshausen and D. Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Res.*, 37:3311–3325, 1997.
- [140] B. Olshausen and D. Field. Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 14:481–487, 2004.
- [141] B. Olshausen and D.J. Field. Wavelet-like receptive fields emerge from a network that learns sparse codes for natural images. *Nature*, 381:607–609, 1996.
- [142] C.C. Pack, B.R. Conway, R.T. Born, and M.S. Livingstone. Spatiotemporal structure of nonlinear subunits in macaque visual cortex. *Journal of Neuroscience*, 26(3):893–907, 2006.
- [143] P. Parent and S. Zucker. Trace inference, curvature consistency, and curve detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11:823–839, 1989.
- [144] A. Pasupathy and C Connor. Population coding of shape in area V4. *Nature Neuroscience*, 5(12):1332–1338, 2002.
- [145] A. Pasupathy and C. E. Connor. Responses to contour features in macaque area V4. *J. Neurophysiol.*, 82(5):2490–2502, 1999.
- [146] I. P. Pavlov. *Conditioned reflexes*. Routledge and Kegan Paul, London, 1927.

- [147] A.E.C Pece. The problem of sparse image coding. *Jour. Math. Imaging and Vis.*, 17(2):89–108, 2002.
- [148] L. Peotta, L. Granai, and P. Vandergheynst. Image compression using an edge adapted redundant dictionary and wavelets. *EURASIP SP, special issue on Sparse Approximations in Signal and Image Processing*, 86(3):444–456, 2006.
- [149] P. Perona. Steerable-scalable kernels for edge detection and junction analysis. *Image and Vision Computing*, 10(10):663–672, 1992.
- [150] P. Perona and J. Malik. Detecting and localizing edges composed of steps, peaks, and roofs. Technical report, MIT- CICS Technical Report, 1991.
- [151] L. Perrinet. Feature detection using spikes : the greedy approach. *J. Physiology (Paris)*, 98(4-6):530–9, 2004.
- [152] L. Perrinet, M. Samuelides, and S. Thorpe. Sparse spike coding in an asynchronous feed-forward multi-layer neural network using matching pursuit. *Neurocomputing*, 57C:125–34, 2002.
- [153] L. Perrinet, M. Samuelides, and S. Thorpe. Coding static natural images using spiking event times: do neurons cooperate? *IEEE Trans. on Neural Networks*, 15(5):1164–75, 2004.
- [154] N. Petkov and P. Kruizinga. Computational models of visual neurons specialised in the detection of periodic and aperiodic oriented visual stimuli: Bar and grating cells. *Biol. Cybern.*, 76:83–96, 1997.
- [155] A.J. Pinho and L.B. Almeida. A review on edge detection based on filtering and differentiation. *Electrónica e Telecomunicações - Revista do DETUA*, 2(1):113–126, 1997.
- [156] U. Polat and C.W. Tyler. What pattern the eye sees best. *Vision Res.*, 39(5):887–895, 1999.
- [157] D.A. Pollen and S.F. Ronner. Phase relationships between adjacent simple cells in the visual cortex. *Science*, 212:1409–1411, 1981.
- [158] J. Portilla, R. Navarro, O. Nestares, and A. Tabernero. Texture synthesis-by-analysis based on a multiscale early-vision model. *Opt. Eng.*, 35(8):1–15, 1996.
- [159] J. Portilla and E. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *Int. Jour. of Comp. Vis.*, 40(1):49–70, 2000.
- [160] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Trans. Image Proc.*, 12(11):1338–1351, 2003.
- [161] R. Prenger, M.C.-K. Wu, S.V. David, and J. L. Gallant. Nonlinear v1 responses to natural scenes revealed by neural network analysis. *Neural Netw.*, 17(5-6):663–679, 2004.

- [162] J.M.S. Prewitt. Object enhancement and extraction. In B.S. Lipkin and A. Rosenfeld, editors, *Picture Processing and Psychopictorics*. Academic Press, 1970.
- [163] N.S. Priebe and D. Ferster. Direction selectivity of excitation and inhibition in simple cells of the cat primary visual cortex. *Neuron*, 45:133–145, 2005.
- [164] M. Pötzsch, N. Krüger, and C. Malsburg. Improving object recognition by transforming Gabor filter responses. *Network: Computation in Neural Systems*, 7(2):341–347, 1996.
- [165] R. Rangarajan, R. Venkataramanan, and S. Shah. Image denoising using wavelets. Technical report, College of Engineering, University of Michigan, 2002.
- [166] R.P.N. Rao, B.A. Olshausen, and M.S. Lewicki, editors. *Probabilistic Models of the Brain: Perception and Neural Function*. MIT Press, 2002.
- [167] L. Rebollo-Neira and D. Lowe. Optimised orthogonal matching pursuit approach. *IEEE Signal Proc. Letters*, 9(4):137–140, 2002.
- [168] R. Redondo and G. Cristóbal. Lossless chain coder for gray edge image. In *IEEE Int. Conf. on Image Proc.*, pages II:201–204, 2003.
- [169] R. Redondo, S. Fischer, L. Perrinet, and G. Cristóbal. Modeling of simple cells through a sparse overcomplete Gabor wavelet representation based on local inhibition and facilitation. In *European Conf. On Visual Perception, A Coruña, Spain*, august 2005.
- [170] R. Redondo, F. Sroubek, S. Fischer, and G. Cristóbal. A multiresolution-based fusion scheme through log-Gabor wavelets and a multisize windows technique. *Information Fusion*, submitted.
- [171] D.L. Ringach. Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Jour. of Neurophysiology*, 88(1):455–463, 2002.
- [172] D.L. Ringach. Mapping receptive fields in primary visual cortex. *J. Physiol.*, 558(3):717–728, 2004.
- [173] Y.M. Ro, M. Kim, H. K. Kang, B.S. Manjunath, and J. Kim. MPEG-7 homogeneous texture descriptor. *ETRI Journal*, 23(2):41–51, 2001.
- [174] E.T. Rolls. Invariant object and face recognition. In Werner JS and Chalupa LM, editors, *The Visual Neurosciences*, pages 1155–1178. MIT press, Cambridge, Mass., 2003.
- [175] L. Rosenthaler, F. Heitger, O. Kubler, and R. von der Heydt. Detection of general edges and keypoints. In *European Conf. on Computer Vision*, pages 78–86, 1992.
- [176] B.W. Rust and H.E. Rushmeier. A new representation of the contrast sensitivity function for human vision. In H.R. Arabnia, editor, *CISST'97*, pages 1–15, 1997.
- [177] N.C. Rust, O. Schwartz, J.A. Movshon, and E.P. Simoncelli. Spatiotemporal elements of macaque v1 receptive fields. *Neuron.*, 46(6):945–956, 2005.

- [178] P. Seriès, J. Lorenceau, and Y. Frégnac. The silent surround of V1 receptive fields: theory and experiments. *J. Physiol. Paris*, 97:453–474, 2003.
- [179] E. Sharon, A. Brandt, and R. Basri. Completion energies and scale. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(10):1117–1131, 2000.
- [180] J. L. Silván and B. Esclante. The multiscale hermite transform for local orientation analysis. *IEEE Trans. on Image proc.*, 15(5):1236–1253, 2006.
- [181] E. Simoncelli. Statistical modeling of photographic images. In Alan Bovik, editor, *Handbook of Image and Video Processing*, pages 431–441. Academic Press, 2nd edition, 2005.
- [182] E.P. Simoncelli, W.T. Freeman, and D. J. Heeger. Shiftable multiscale transforms. *IEEE Trans. Inf. Theory*, 38(2):587–607, 1992.
- [183] B.F. Skinner. *The Behavior of Organisms: An Experimental Analysis*. Appleton-Century-Crofts, New York, 1938.
- [184] A. Skodras, C. Christopoulos, and T. Ebrahimi. The JPEG 2000 still image compression standard. *IEEE Signal Processing Magazine*, 18(5):36–58, 2001.
- [185] S. M. Smirnakis, A.A. Brewer, M.C. Schmid, A.S. Tolia, A. Schüz, M. Augath, W. Inhoffen, B.A. Wandell, and N.K. Logothetis. Lack of long-term cortical reorganization after macaque retinal lesions. *Nature*, 435:300–307, 2005.
- [186] S.M. Smith and J.M. Brady. Susan - a new approach to low level image processing. *Int. Journal of Computer Vision*, 23(1):45–78, 1997.
- [187] I. Sobel. An isotropic 3x3 image gradient operator. In H. Freeman, editor, *Machine Vision for Three-Dimensional Scenes*, pages 376–379. Academic Press, 1990.
- [188] F. Sroubek and J. Flusser. Multichannel blind iterative image restoration. *IEEE Trans. Image Proc.*, 12(9):1094–1106, 2003.
- [189] J. L. Starck, E. J. Candès, and D. L. Donoho. The curvelet transform for image denoising. *IEEE Trans. on Image Proc.*, 11(6):670–684, 2002.
- [190] J. L. Starck, M. Elad, and D. L. Donoho. Image decomposition via the combination of sparse representations and a variational approach. *IEEE Trans. on Image Proc.*, 14(10):1570–1582, 2005.
- [191] D.D. Stettler, A. Das, J. Bennett, and C.D. Gilbert. Lateral connectivity and contextual interactions in macaque primary visual cortex. *Neuron*, 36:739–750, 2002.
- [192] C.F. Stromeyer and B. Julesz. Spatial-frequency masking in vision: critical bands and spread of masking. *J. Opt. Soc. Am. A*, 62(10):1221–1232, 1972.

- [193] K. Tanaka. Inferotemporal response properties. In Werner JS and Chalupa LM, editors, *The Visual Neurosciences*, pages 1151–1164. MIT press, Cambridge, Mass., 2003.
- [194] Y. Tanaka and D. Sagi. Long-lasting, long-range detection facilitation. *Vision Res.*, 38:2591–2599, 1998.
- [195] C.-K. Tang and G. Medioni. Curvature-augmented tensor voting for shape inference from noisy 3d data. *IEEE Trans. P.A.M.I.*, 24(6):858–864, 2002.
- [196] C.-K. Tang, G. Medioni, and M. Lee. N-dimensional tensor voting and application to epipolar geometry estimation. *IEEE Trans. P.A.M.I.*, 23(8):829 – 844, 2001.
- [197] C. Taswell. The what, how and why of wavelet shrinkage denoising. *Computing in Science and Engineering*, pages 12–19, 2000.
- [198] A. Thielscher and H. Neumann. Neural mechanisms of human texture processing: Texture boundary detection and visual search. *Spatial Vision*, 18:227–257, 2005.
- [199] S.J. Thorpe and M. Fabre-Thorpe. Seeking categories in the brain. *Science*, 291(5502):260–263, 2001.
- [200] W.-S. Tong and C.-K. Tang. Robust estimation of adaptive tensors of curvature by tensor voting. *IEEE Trans P.A.M.I.*, 27(3):434–449, 2005.
- [201] W.-S. Tong, C.-K. Tang, P. Mordohai, and G. Medioni. First order augmentation to tensor voting for boundary inference and multiscale analysis in 3d. *IEEE Trans. P.A.M.I.*, 26(5):594–611, 2004.
- [202] J.A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. on Inf. Theory*, 50(10):2231– 2242, 2004.
- [203] D. Tschumperlé and R. Deriche. Vector-valued image regularization with PDE’s : A common framework for different applications. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(4):506–517, 2005.
- [204] T. Tversky, W.S. Geisler, and J.S. Perry. Contour grouping: closure effects are explained by good continuation and proximity. *Vision Res.*, 44(24):2769–2777, 2004.
- [205] M. Ursino and G.E. La Cara. A model of contextual interactions and contour detection in primary visual cortex. *Neural Netw.*, 17(5-6):719–35, 2004.
- [206] M.J. Van der Smagt, C. Wehrhahn, and T.D. Albright. Contextual masking of oriented lines: Interactions between surface segmentation cues. *J. Neurophysiol.*, 94:576 – 589, 2005.
- [207] J. H. Van Hateren and A. Van Der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc Roy Soc Lond B*, 265:359–366, 1998.

- [208] R. VanRullen, A. Delorme, and S.J. Thorpe. Feed-forward contour integration in primary visual cortex based on asynchronous spike propagation. *Neurocomputing*, 38-40(1-4):1003–1009, 2001.
- [209] B.T. Vincent, R.J. Baddeley, T. Troscianko, and I.D. Gilchrist. Is the early visual system optimised to be energy efficient? *Network: Computation in Neural Systems, special issue on Sensory Coding and the Natural Environment*, 16(2/3):175–190, 2005.
- [210] W.E. Vinje and J.L. Gallant. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287(5456):1273–1276, 2000.
- [211] R. von der Heydt. Image parsing mechanisms of the visual cortex. In Werner JS and Chalupa LM, editors, *The Visual Neurosciences*, pages 1139–1150. MIT press, Cambridge, Mass., 2003.
- [212] R. von der Heydt and E. Peterhans. Mechanisms of contour perception in monkey visual cortex. i. lines of pattern discontinuity. *Journal of Neuroscience*, 9:1731–1748, 1989.
- [213] M.J. Wainwright, O. Schwartz, and E. P. Simoncelli. Natural image statistics and divisive normalization: Modeling nonlinearity and adaptation in cortical neurons. In R. Rao, B. Olshausen, and M. Lewicki, editors, *Probabilistic Models of the Brain: Perception and Neural Function*. MIT Press, Boca Raton, FL, 2002.
- [214] M. Wakin, J. Romberg, H. Choi, and R. Baraniuk. Image Compression using an Efficient Edge Cartoon + Texture Model. In *Data Compression Conf.*, Snowbird, UT, April 2002.
- [215] Y. Wang and B. Lohmann. Multisensor image fusion: Concept, method and applications. Technical report, Univ. Bremen., 2000.
- [216] A.B. Watson. The cortex transform: rapid computation of simulated neural images. *Comput. Vision, Graphics and Image Processing*, 39:311–327, 1987.
- [217] BS Webb, NT Dhruv, SG Solomon, C Tailby, and P. Lennie. Early and late mechanisms of surround suppression in striate cortex of macaque. *J Neurosci.*, 25(50):11666–75, 2005.
- [218] L.R. Williams and K.K Thornber. A comparison of measures for detecting natural shapes in cluttered backgrounds. *Jour. of Computer Vision*, 34(2/3):81–96, 2000.
- [219] R. Wurtz. *Multilayer dynamic link networks for establishing image point correspondences and visual object recognition*. PhD thesis, Bochum University, 1994.
- [220] R.H. Wurtz and D.H. Kandel. Central visual pathways (Chapter 27). In McGraw-Hill, editor, *Principles of neural science*, pages 523–547. Kandel, E. R. and Schwartz, J. H. and Jessell, T. M., 4th edition, 2000.
- [221] S.-C. Yen and L. H. Finkel. Extraction of perceptually salient contours by striate cortical networks. *Vision Res.*, 38(5):719–741, 1998.

- [222] W. Zeng, S. Daly, and L. Shawmin. Point-wise extended visual masking for JPEG-2000 image compression. In *ICIP*, volume 1, pages 657–660, 2000.
- [223] C. Zetsche and F. Rohrbain. Nonlinear and extra-classical receptive field properties and the statistics of natural scenes. *Network: Comput. Neural Syst.*, 12(3):331–350, 2001.
- [224] S. Zhong and V. Cherkassky. Image denoising using wavelet thresholding and model selection. In *Int. Conf. on Image Proc.*, volume 3, pages 262–265, 2000.
- [225] N.A. Ziyad, E.T. Gilmore, and M.F. Chouikha. Dictionary approaches to image compression and reconstruction. In *Proc. Of the Iasted Int. Conf. On Signal and Image Processing, Nevada USA*, Sept. 1998.
- [226] S. W. Zucker, R. A. Hummel, and A. Rosenfeld. An application of relaxation labeling to line and curve enhancement. *IEEE Trans. Comput.*, 26:394–403, 1977.

Appendix A

Philosophical viewpoints in cognitive science

"Thus, in the early days of industrial capitalism, slave owners could and did point out that if you own a piece of machinery, you are likely to treat it with more care than if you merely rent it. Similarly, the slave owner is likely to treat his possession with more care and solicitude than the capitalist who merely rents people for his temporary purposes. Slavery, then, reflects higher moral standards than "wage slavery." No sane person would now accept this argument, though it is not entirely absurd by any means. As civilization progressed, it came to be understood that slavery is an infringement on essential human rights. We may look forward to the day when wage slavery and the need to rent oneself to survive may be seen in a similar light, as we come to have better understanding of the moral values rooted in our inner nature."

Noam Chomsky, *Language and Problems of Knowledge: The Managua Lectures* [25], p154.

Under the neuroscience modeling debates underly philosophical or even ideological aspects staying unfortunately too often hidden. A review on the underlying philosophical ideas could then benefit to the scientific debate.

- Ambientalism sometimes meet *antiracism* since it states that the cognitive processes are mostly learned, then do not depend importantly on the genetic. Innatists by their part consider that genetical differences are significative between species but not between individuals. They then do not support racism as a superficial critics let think.
- Ambientalists often consider that there is no *human nature*, and that humans are

the pure product of the environment. Innatists consider that there is a consistent, strong and innate human nature due to the genetic load of the specie.

- The ambientalist point of view that the human being is formed mainly by influence of the external world is consistent with the ideologies that intent to transform and "improve" the human life on a rational basis. Today's ambientalists are often the fervent defensors of the ideologies of *progress* and *civilization development*: the political modification of life conditions could "improve" the human life in a never-ending progress. Moreover the supposed lack of innate basis leads to depreciate the primitive forms of human life. Innatists can be more conservators. They think an unalterable human nature underlies the human life. Such human nature responsible for the richness of the human behaviors and abilities was already present before the expansion of the civilization. Therefore for them the well-being could be in the best case the understanding and the free maturing of this human nature, and some innatists can even consider that civilization progresses bring more sickness than health in the human life. Such conservative point of view towards the progress could be a reason for the distrust of many intellectuals towards innatism.
- The opposition between ambientalists and innatists is particularly evident in education where ambientalists developed the conditioning techniques trying to demonstrate that it is possible to totally "reformat" the human mind, whereas for innatists "teaching should not be compared to filling a bottle with water but rather to helping a flower to grow in its own way" (Chomsky in [25]). Many persons are afraid by the idea of a human nature since there exists a real risk that a way of living would be imposed once the "human nature" would have been described "scientifically". For innatists this fear is unrealistic: the human nature is by definition innate, it would then be meaningless to impose it.

Appendix B

Developing Big Brother

"On peut discuter des applications "bonnes" ou "mauvaises" de la recherche, soutenir que "l'outil est neutre" et l'usage seul en cause, qu'il ne faut pas "jeter le bébé avec l'eau du bain", "le bon grain avec l'ivraie", etc. mais un fait demeure indiscutable : dans un monde où s'opposent dominants et dominés, tout "progrès des connaissances" sert d'abord les dominants, leur sert d'abord à dominer, et autant que possible à rendre irréversible leur domination. Les "retombées positives" n'étant que les moindres maux dont on achète la soumission des dominés. "Progrès" curatifs et palliatifs, quand la prévention serait de renverser la domination qui provoque tant de nos maux, pour s'en rendre ensuite l'indispensable thérapeute.

Ce simple fait ne devrait-il pas conduire tout chercheur prétendument équipé d'une conscience à objecter ? Et l'ensemble de sa corporation à proclamer un moratoire sur toute recherche tant que ne serait pas réglée la question du pouvoir ?"

Simples citoyens, Nanotechnologies / Maxiservitude, Des contributions grenobloises à l'automatisation du cheptel humain.

The advances of technology and the increasing probability to achieve the building of efficient artificial vision systems that could perform accurately tasks of scene interpretation, can give rise to a justified fear. Such systems could be used for instance for systematically controlling human movements and reducing human freedom. I would like to stress here two related questions: firstly to throw doubts on the largely represented belief that the technology systematically makes the human life more pleasant and more free. And secondly to review some actual failures in the democratic systems which are sufficiently serious to prevent the institutions to confidently protect the citizens and to obey to the public opinion concerning technologic developments and usages.

It should be considered that new technologies do not necessarily improve the way of life, but in some cases, or even in many cases, they restrict and denature the human life. Let's take as example one of the inventions that strongly modified the human life during the 20th century: the automobile. It has drastically modified the way of living, the urbanism, the social interactions, the way of working and consuming, etc. It is generally thought that the automobile is a tool and that one can freely use it or not. Instead the automobile induced changes in the urbanism that made everyday's travel distances irresistibly larger and finally increased the travel time. Chemical, auditory and esthetic contaminations made the cities unpleasant, dangerous and unhealthy places. The overall ecological equilibriums are under threat. In parallel we cannot freely escape of this technology since in many cities it is necessary to drive a car to have a job, to buy everyday's food, etc. In some aspects we get the captives of that technology since we can no more live without, even if we are aware that it brings major drawbacks.

There is a real faith in the never-ending improvement of the life quality through the technologies. The objective to improve the life quality through technological solutions is unrealistic for at least three reasons: first because humans are themselves animals that requires their ecological environment to live. The technology severely deranges the ecological equilibriums reached through millions of years of co-evolution between vegetal and animal species and the physical environment. Moreover it stays very hypothetical that future technologic improvements will offer solutions for all the ecological disequilibriums that the former and actual technologies produce. Second because many human problems like the ability to live in peace, the ability not to dominate nor to plunder its neighbor do not reside in a technology but in the individual wisdom, education and knowledge (non-necessary academic). And third because, as some science fiction writers pointed out, human abilities decreases as the technology preforms their parts of everyday's tasks. Humans get highly specialized, every time more single-task and ignorant of what stay beyond their speciality: "The logical end of mechanical progress is to reduce the human being to something resembling a brain in a bottle [...] some frightful subhuman depth of softness and helplessness." (George Orwell, *The Road to Wigan Pier*, Chapter 12). Concluding, it is sure that the technology is modifying drastically the human life, but it could reveal to be more destructive and submissiveness than constructive and liberating. It is then worrying that the technology appears today like a new religion on which everyone believe hopping for a happy future. A bit more of scepticism could be salutary.

Nowadays it is no more possible (and it was perhaps never the case) to trust the political and institutional decisions for preserving everyone wellbeing. The old democracy set up in the 18th century was certainly not designed for resisting to the influences of big corporations and companies which are today more powerful than governments and states. Such companies are extremely powerful for its financial capacities, for their

influence on governments, but also for owning the mass-media. (Today's mass-media in Europe or the USA are almost exclusively controlled by a very small number of corporations. Silvio Berlusconi is a caricatural example of this concentration of financial, media-related and political power.) Therefore the corporations are not only non-democratic institutions, they also play a drastic role in making the democracy working as a coarse spectacle in which most people already lost interest.

Then, can we confidently offer to such institutions new technologies able to strongly limit the human freedom? It would be rather preferable to revendicate first a working and efficient democracy at every stages of the actual modern society, including within the corporations. This democracy would not consist in the actual spectacular elections where two parties ruled by the same financial caste face in a shallow theater. It would be a participative democracy, where all opinions can express with roughly equal media support. The media would therefore not be owned by a financial caste. And important questions on the society construction and technology development and usage can at last be discussed.