INSTITUTO DE OPTICA
"DAZA DE VALDES"
C.S.I.C.
MADRID

REPRESENTACION DE IMAGENES MEDIANTE FUNCIONES DE GABOR. MODELADO DEL SISTEMA VISUAL Y ANALISIS DE TEXTURAS

Memoria presentada en la Facultad de Ciencias Físicas
de la Universidad Complutense de Madrid
por

ANTONIO TABERNERO GALAN

para aspirar al grado de Doctor en Ciencias Físicas

MADRID 1992

Agradecimientos

Al final del trabajo quisiera mencionar a quienes me han ayudado, o han estado cerca durante su desarrollo.

En primer lugar, al Dr. Rafael Navarro, director de este trabajo, quien con su interés por el tema y su apoyo constante ha contribuido decisivamente en el desarrollo de esta tesis.

Resaltar a Al Ahumada del NASA Ames Research Center, tanto por su hospitalidad durante aquellos dos meses, como por sus enseñanzas, sin las cuales el capítulo 7 no habría visto la luz.

A los Drs. Julián Bescós y Javier Santamaría, con quienes primero contacté al llegar al instituto, y quienes me introdujeron en el campo del tratamiento de imágenes y visión.

También agradecer al Dr. Antonio Corrons, Director del Instituto de Optica, por poner a mi disposición los medios del Centro, al Prof. Dr. Eusebio Bernabeu, Director del Departamento de Optica de la Universidad Complutense de Madrid, y ponente de esta tesis.

La vida en el instituto no habría sido lo mismo sin todos los compañeros con los que he compartido, en mayor o menor medida, venturas y desventuras: Gabriel, Chelo, Jose Mari, Sesé , Mariangeles, los tres Juanes (Alberto, Manuel e Ignacio), Fidel, Santiago, Luis, Guillermo, Pedro; en especial a Jose, por sus "ayudas" en estos momentos finales. También, Luis, Jose Vicente, Pablo, etc. y cómo no, a Pili, la alegría de la tercera planta.

Y la vida fuera del instituto habría sido mucho más aburrida sin otros muchos amigos, demasiados para ser nombrados aquí, los de la Facultad y los otros, que han estado a mi alrededor durante estos últimos años. A todos ellos va dedicada esta tesis.

Resumen

Se presenta un esquema de representación de imágenes basado en funciones de Gabor, como un entorno multipropósito para aplicaciones de codificación y procesado de imágenes y visión artificial. El esquema considerado contempla las características básicas de la codificación de la información que se produce en la corteza visual, constando de 4 canales de frecuencia (distribuidos en octavas) y 4 de orientacion. Se han desarrollado dos implementaciones alternativas para el cálculo rápido de esta transformada de Gabor, prestando especial atención a la correspondiente al dominio espacial, donde se usan máscaras de tamaño reducido (7×7) y una implementación piramidal con objeto de reducir el coste computacional.

A partir de esta base común, el esquema de Gabor propuesto se ha aplicado en dos direcciones diferentes. En primer lugar se estudia su eficiencia en tareas de análisis de imágenes. Para ello se ha definido una matriz de descriptores 4×4 cuyas componentes son la salida directa de los canales del esquema de Gabor. Dicha matriz se aplica a tareas de segmentación y clasificación de texturas con excelentes resultados, así como al problema de reconocimiento invariante de imágenes rotadas o cambiadas de escala. Se presenta también un método para estimar la dimensión fractal de imágenes con una buena precisión a partir de la matriz de decriptores. En segundo lugar, dicho esquema se adopta como base para el desarrollo de un modelo realista de las etapas primarias del sistema visual humano, susceptible de ser empleado tanto para la simulación de efectos visuales, como también como base de sistemas de visión artificial. Este modelo se ha restringido aquí a la zona de la fóvea (unos 3° en el centro del campo visual). Entre los aspectos que considera están la óptica del ojo, el muestreo en la retina por una red hexagonal de espaciado variante y el modelado de las células simples de la corteza visual con funciones de Gabor. A partir de dicho modelo se reproducen fenómenos observados experimentalmente, tales como los patrones de Moiré que aparecen cuando se proyectan estímulos de alta frecuencia directamente sobre la retina, y otras ilusiones visuales. También se simula la detección de estímulos sinusoidales por el sistema visual (función de sensibilidad al contraste o CSF) y su variación con la excentricidad.

Finalmente, se propone un modelo simple del sistema visual desde la retina hasta el córtex. Dicho modelo está basado en redes neuronales y podría explicar el desarrollo y organización de los campos receptivos de células simples en el córtex, sin hipótesis previas sobre su forma, en base a procesos de aprendizaje.

Abstract

A scheme of image representation based on Gabor functions is presented as a multipurpose environment for applications involving image coding and processing and artificial vision. The proposed scheme consists of 4 frequency channels (distributed in octaves) and 4 orientation channels, taking into account the basic characteristics of the information coding that has been observed in the visual cortex. Two alternative implementations (in both domains) for the fast computation of this Gabor transform have been developed, paying special attention to the one working in the spatial domain, where masks of small size (7×7) and a pyramid implementation help to reduce the computational cost.

From this common ground, the Gabor scheme has been applied in two different directions. First, its performance in texture analysis tasks has been examined. As a descriptor for the textures, we have defined a 4×4 matrix, the components of which are the direct outputs of the 4×4 channels of the Gabor scheme at the point being analyzed. This descriptor matrix has been applied to texture segmentation and analysis with excellent results, as well as to the problem of invariant recognition of rotated or scaled images. A method for estimating the fractal dimension of images from the same descriptor matrix is also proposed

Secondly, the Gabor scheme is adopted as the departing point for a more complex and realistic model of the early stages of the human visual system, capable of being used both for the simulation of visual effects and as a base of artificial vision systems. The model is here restricted to the foveal area (the central 3° of the visual field), and we have considered the effects of the optics of the eye, the retinal sampling using a hexagonal lattice with a variant spacing, and the modelling of simple cells in the visual cortex with Gabor functions. Using this model, we have reproduced several experimentally observed phenomena, such as the "zebra patterns" that can be seen when imaging high frequency gratings on the retina bypassing the eye optics, and other visual illusions. The ability of the visual system to detect sinusoidal stimuli (contrast sensibility function, CSF) has been also simulated, as well as its variability with the eccentricity.

Finally, a simple model of the visual system based on neural networks is proposed. This model could explain the development and organization of the receptor fields of simple cells in the visual cortex as the result of a learning process. without a priori assumptions about their form.

Indice

Lista de Figuras

Lista de Tablas

Capítulo 1

Introducción

En los últimos años hemos asistido a un interés creciente en las diversas técnicas de procesado de imágenes [1][2] y sobre todo de visión artificial [3]. Las razones de este auge han sido muy diversas, pero al margen de su indiscutible utilidad, existen dos factores que han sido fundamentales: su amplio radio de aplicación en numerosos campos de la ciencia y la técnica y la aparición de sistemas de computación suficientemente potentes. Sólo con éstos ha sido posible la implementación eficiente de dichas técnicas en tiempos razonables. Para dar una idea del potencial de aplicación del procesado de imágenes, basta decir que en nuestra vida diaria más de un 80% de la información la recibimos por via visual.

Paralelamente al desarrollo de estas técnicas, estudios experimentales del sistema visual (SV) humano (y de animales superiores) han constatado que es altamente eficiente en tareas de procesado, codificación y transmisión de la información (por ejemplo, de la retina al cerebro) [4][3]. Asímismo, todos conocemos la superioridad del cerebro frente a los algoritmos actuales en tareas de análisis e interpretación de imágenes. Incluso animales inferiores como los insectos realizan eficientemente tareas dificiles de reproducir en nuestros más rápidos ordenadores.

En visión artificial, al igual que en otros muchos campos de la ciencia y de la ingeniería, un posible enfoque para resolver un cierto problema es mirar a la naturaleza, imitando la solución adoptada por los sistemas biológicos. Dichos sistemas no pueden desarrollar libremente un algorítmo "ad hoc" de características diferentes para cada aplicación concreta. En efecto, estando limitados por unas estructuras determinadas, deben buscar soluciones a los diferentes problemas en un marco único, aunque flexible. Del mismo modo podemos pensar en establecer un entorno general de codificación, procesado y análisis de imágenes, a partir del cual se desarrollarían las distintas aplicaciones, pero siempre con una base común.

Dada la innegable superioridad del SV humano sobre los algorítmos convencionales de visión artificial, es lógico pensar que basar dicho entorno en modelos del SV podría presentar numerosas ventajas.

Es alrededor de esta idea donde se situa el presente trabajo. Con estas premisas, vamos a dar a continuación una breve introducción de algunas de las características básicas del procesado de información en el SV humano, según los modelos actuales.

Es común dividir el proceso visual en etapas, las primeras comúnmente denominadas tempranas [4]. Estas son de bajo nivel, produciendose de forma automática, sin que precisen atención o procesos de razonamiento. Comprenden desde la formación de la imagen óptica, su muestreo por los fotoreceptores y el preprocesado y compresión para su transmisión a través del nervio óptico hasta la corteza visual primaria. Una vez en la corteza (o "córtex") visual, la señal está ya codificada de una forma especial, habiendose separado los distintos componentes de la información (forma y textura, color, movimiento, estereopsis, etc.). Sobre esta información operan procesos de alto nivel (análisis, reconocimiento, interpretación, etc.), que tienen lugar en capas más profundas del cerebro, siendo menos conocidos, y en los que no vamos a entrar al estar fuera del propósito de esta tesis. Vamos ahora a recordar aquellos aspectos, sólo de las etapas primarias, relevantes para nuestro trabajo.

El primer paso se produce en el ojo, que puede modelarse [5] como un sistema óptico formador de imágenes sobre la retina. Este proceso es lineal y se caracteriza a través de la distribución de energía de la imagen de un punto (PSF). Suponiendo invarianza espacial, lo cual es aproximadamente cierto localmente, una imagen extensa vendrá dada por la convolución del objeto con la PSF. Alternativamente, un sistema óptico se puede caracterizar por la transformada de Fourier de la PSF, la función de transferencia óptica (OTF) o su módulo (MTF). La MTF del ojo como sistema óptico se ha medido a través de métodos de doble paso [6] y psicofísicos, dando en ambos casos resultados concordantes [7].

Una vez formada la imagen sobre la retina, es detectada y muestreada por los fotoreceptores. Existen fundamentalmente dos tipos de fotoreceptores, conos y bastones, aunque dada su especialización en diversas condiciones de visión, se suelen estudiar por separado. El mosaico de fotoreceptores en la retina es de tipo hexagonal, preferible al rectangular por razones de eficacia en el muestreo [8]. La densidad de fotoreceptores presenta un máximo en la fóvea, o zona correspondiente al centro del campo visual, donde la agudeza visual es máxima. Al aumentar la excentricidad, el espaciado entre fotoreceptores decae de forma cons-

tante, perdiendose resolución. Este es el compromiso adoptado por el SV para resolver simultáneamente dos requerimientos contrapuestos, alta resolución y gran campo visual, con un número limitado de fotoreceptores. Los datos sobre la red de muestreo, tales como la densidad de fotoreceptores, la regularidad de la red en orientaciones o espaciado, el tamaño de los receptores, etc. han sido obtenidos tradicionalmente con técnicas fisiológicas "in vitro" [9]–[14]. Recientemente se han desarrollado técnicas no invasivas que permiten un estudio en vivo. Unas son psicofísicas, [15]–[18] basadas en la apreciación subjetiva de patrones de Moiré producidos por franjas de interferencia de alta frecuencia ("aliasing") [19]–[22]. Un método objetivo basado en la interferometría estelar de "speckle" ha sido también propuesto recientemente [23].

En la retina comienza ya la parte neural del SV. Los métodos de estudio en esta fase son fundamentalmente neurofisiológicos, en base al estudio de los registros de actividad de las diversos tipos de neuronas en las diversas etapas del proceso, aunque también pueden estudiarse mediante métodos psicofísicos. El procesado neural empieza en la misma retina, donde ya tenemos un complejo entramado de conexiones que comprende varios niveles y tipos de células [24]. El paso final son las células ganglionares cuyos axones (del orden de 10^6) forman el nervio óptico que lleva la información hasta el cuerpo geniculado lateral (CGL). Las neuronas del CGL muestran una similitud casi completa con las células ganglionares, por lo que son éstas últimas las que se suelen estudiar por su mejor accesibilidad. Una de las funciones principales del proceso retiniano sería la compresión de la información que ha de transmitirse por el nervio óptico, ya que se pasa de un total de unos 10^8 fotoreceptores a sólo unas 10^6 células ganglionares.

Es difícil, aun con los medios actuales, establecer exactamente cuáles son las conexiones entre fotoreceptores y células ganglionares, por lo que nos limitaremos a describir sus campos receptivos [25]. Su forma es la de filtros de tipo paso-banda, habiendo sido modelados por diferencias de gaussianas (DOG) u operadores laplacianos [26], presentando una región circular central rodeada de un anillo de signo opuesto. No hay evidencias de su selectividad a orientaciones. Al margen de diferencias morfológicas [27] hay dos clases, ON y OFF [28], dependiendo de si son activas a un estímulo luminoso central o viceversa, no estando aún resuelto si forman canales de información separados o se combinan de alguna forma. La relación entre células ganglionares y fotoreceptores es difícil de establecer en la fóvea [10][29][30], pero se sabe que en la periferia son las células ganglionares las que marcan el limite a la resolución espacial [31][32].

El paso posterior al CGL es el córtex visual (zona V1). De nuevo, es difícil

establecer las conexiones exactas entre ambas zonas; sin embargo, hay numerosos estudios sobre la forma de los campos receptivos de células simples en el córtex. Desde los trabajos pioneros de Hubel & Wiesel [33] se ha determinado que dichas células muestran selectividad a orientaciones y frecuencias dadas, estando por lo tanto localizadas en el dominio de Fourier [34]–[36]. Esto se confirmó con experimentos psicofísicos [37]. Los campos receptivos de estas neuronas se han modelado como filtros lineales usando diversas funciones: diferencias de gaussianas, funciones de Gabor [38]–[41], etc. El uso de filtros lineales supone una simplificación, ya que se conocen diversos tipos de células cuyas respuestas son fuertemente no lineales. Sin embargo, el papel jugado en la visión por dichos mecanismos no lineales se desconoce en su mayor parte, por lo que la mayoría de los modelos los ignoran [42][43]. Dado que cada uno de estos campos receptivos cubre sólo una pequeña área del campo visual (localización espacial), para ver la escena completa debemos disponer de muchos de ellos, cada uno muestreando una parte diferente. Tal mosaico de campos receptivos es lo que se ha denominado un canal del SV. Dichos canales se caracterizan por su posición en el espacio de Fourier (localización en frecuencias).

Finalmente, decir que estas células se encuentran agrupadas con una clara organización, donde se ha comprobado que las neuronas con similar selectividad a la orientación se agrupan en columnas que se extienden verticalmente a través del córtex [33]–[35]. De la misma forma, neuronas que procesan información de la misma área de la retina forman lo que se denomina un módulo del córtex estriado. Tales módulos se repiten a lo largo del córtex, de forma que cubren todo el campo visual. El número de neuronas dedicadas a un área de la retina no es constante, sino que disminuye con la excentricidad. En la corteza visual se da especial importancia a las zonas que tratan la información procedente de la fóvea (casi la mitad del córtex está dedicado a los 5 grados centrales del campo visual [44]). Al principio se creyó que esta especialización se generaba en el propio córtex, pero estudios recientes muestran que ésta podría aparecer ya en la retina [29]. Esta representación constituye una transformación conforme del espacio exterior (cartesiano, x,y) a unas nuevas coordenadas ($\log(r)$,θ), donde r y θ son coordenadas polares. En el dominio de Fourier aparece una representación similar, ya que los canales de frecuencia se distribuyen por octavas [45][42][39].

En los parrafos anteriores se ha descrito el tipo de información que llega a la corteza visual. Esta no parece concordar con una representación en el dominio espacial (píxeles) ni en el de Fourier (frecuencias espaciales). La primera nos da información sobre la energía de la señal en un punto dado, pero no sobre el con-

tenido en frecuencias en las inmediaciones de dicho punto. La representación de Fourier nos informa sobre la energía de una frecuencia dada, no nos dice nada acerca de su distribución espacial. Esto es debido a que las funciones base de dichas representaciones están totalmente lozalizadas en uno de los dominios, y por lo tanto, deslocalizadas en el otro. Por el contrario, en el córtex tenemos una representación conjunta, que proporciona información simultánea sobre ambos dominios (aunque debido al principio de incertidumbre, con una cierta imprecisión en ambos). Una de las primeras representaciones conjuntas propuestas fué la distribución de Wigner [46], introducida en el campo de la mecánica cuántica y posteriormente usada en procesado de imágenes [47][48]. Las funciones base de una representación conjunta renuncian a estar completamente localizadas en un dominio, lo que les permite suministrar información sobre ambos. Cohen [49] presentó una formulación general de los distintos tipos de representaciones conjuntas.

Una de las funciones que se citaron como modelos de los campos receptivos en el córtex, las funciones de Gabor, constituyen un ejemplo de base de una representación conjunta. Las funciones de Gabor, primeramente propuestas por D. Gabor [50] en acústica, son paquetes de onda con una envolvente gaussiana cuyas anchuras en ambos dominios minimizan el producto de incertidumbre [50][51]. Debido a esto fueron propuestas como una forma de empaquetamiento óptimo de la información. Además, sus perfiles gaussianos hacen posible implementaciones alternativas en ambos dominios. Recientemente han cobrado un nuevo interés al haberse propuesto, como ya se mencionó, como modelos de campos receptivos de células simples en el córtex [40]–[39]. La principal objección a una transformación de Gabor es su no completitud [52], es decir, que la imagen original no puede recuperarse exactamente a partir de la transformada. Esto no quita para que sirvan como modelo del SV, ya que las ilusiones visuales muestran que la representación en el SV tampoco es completa. No obstante, se han propuesto modelos basados en funciones ligeramente distintas que posibilitan una transformación exacta, como la transformada Cortex [53]–[55]. Igualmente, se han presentado diversas variaciones de la transformada Gabor que permiten una recuperación exacta de la señal a costa de una mayor complejidad [52][56][57], e implementaciones que tratan de minimizar los efectos de su no completitud [58].

Otro aspecto interesante de las funciones de Gabor es que permiten un procesado de tipo jerárquico o piramidal [59]. Este tipo de procesado es la base de recientes técnicas de análisis de la información, como las denominadas "wavelets" [60][61], codificación en sub-bandas [62]–[64], filtros QMF [65], etc., todas ellas

ejemplos de representaciones conjuntas. El procesado piramidal consiste en la descomposición de la imagen en diversas versiones, cada una recogiendo los detalles a una cierta escala, obteniendose así interesantes ventajas [66]. En el SV tiene lugar también un procesado de este tipo, ya que como dijimos anteriormente, los diversos canales de frecuencia se disponen en una escala logarítmica. Aunque en un principio Gabor no propuso un esquema multiresolución, las funciones de Gabor pueden ser fácilmente adaptadas a un esquema de este tipo.

Hasta ahora hemos descrito al SV como un sistema estático, con una complejidad que viene dada por el gran número de interconexiones entre las neuronas, pero que no difiere cualitativamente de otros sistemas físicos. Sin embargo, uno de los aspectos más característicos (aunque no por ello menos sorprendentes) del SV es su capacidad para aprender y responder de forma flexible a alteraciones del entorno. Por ejemplo, algunas células mueren diarimente sin que sus capacidades resulten significativamente afectadas, e incluso en casos patológicos [67] de pérdidas de fotoreceptores, la disminución de la capacidad visual es menor de la esperada. Se sabe igualmente que los fotoreceptores pueden llegar a migrar una considerable distancia durante las primeras fases del desarrollo [68], pero el SV es capaz de compensar tales desajustes. Desde el punto de vista de modelos físicos tradicionales (como pueda ser un modelo del ojo como sistema óptico) estas propiedades son difíciles de justificar. Se hace pues necesario acudir a modelos radicalmente diferentes. Las propiedades mencionadas aparecen en numerosos aspectos del funcionamiento del cerebro en muy diversas áreas, lo que hace sospechar que es inherente a la organización de dichas áreas. Esto inspiró el desarrollo de modelos de redes neuronales, un nuevo campo a caballo entre la física, las matemáticas y las ciencias de la computación. Una red neuronal [69][70] es un ensamblado de unidades sencillas ("neuronas") con una alta conectividad y trabajando en paralelo, dotada de un mecanismo de aprendizaje, que le permite llegar a realizar la tarea para la cual ha sido diseñada.

Se han presentado distintos tipos de redes neuronales que pueden explicar diversos aspectos del SV. Por una parte, redes neuronales de aprendizaje supervisado han sido propuestas para justificar la capacidad del SV de autocalibrarse, es decir, compensar los posibles desplazamientos de los fotoreceptores, sin necesidad de conocer dicha variación [71]–[74]. Otro tipo de redes neuronales son las llamadas no supervisadas, o de aprendizaje hebbiano, donde la red en sí debe encontrar patrones, redundancias, rutinas, etc. en la entrada y codificarlas de alguna manera en la salida [75]–[77]. La red debe pues mostrar un cierto grado de autoorganización, lo que tiene cierta relación con la estructura encontrada en el córtex, habiendo sido

propuestas por lo tanto como modelos para el aprendizaje de campos receptivos en el cortex [78]. Según esta hipótesis, la formación y desarrollo de los campos receptivos sería el resultado de un proceso de aprendizaje.

El hecho de que el sistema visual analize por separado distintos tipos de información (textura, color, movimiento, etc.) ha hecho que se hayan propuesto diversas aplicaciones, basadas en modelos del sistema visual, que aunque aparentemente diferentes, podrían englobarse en un entorno único. Una parte del procesado de imágenes común a todos estos tipos de información es la codificación. Este es uno de los campos que mayor desarrollo ha tenido, por la necesidad de reducir el ingente volumen de información digital requerida cuando se manejan imágenes, sobre todo en aplicaciones que incorporan color y movimiento [79]. Watson [53][43] y Daugman [56] han propuesto esquemas de proposito general basados en modelos del SV que dan buenos resultados cuando se aplican a la codificación (con factores de compresión entre 8 y 20).

En este trabajo nos centraremos en modelos de visión espacial, excluyendo la información relacionada con el color, estereopsis, movimiento, etc. Dicha información espacial comprende aspectos como la forma, la textura, etc. En particular, una de las características que el sistema visual utiliza para lograr una rápida segregación de una imagen en diferentes áreas es la textura [80]. Siendo un concepto difícil de definir, se han propuesto diversos métodos para su análisis. La mayoría de estos métodos consisten en extraer una serie de de descriptores que puedan servir para caracterizar distintas texturas [81]–[83]. De nuevo, estos descriptores pueden definirse "ad hoc", en función de su rendimiento, o basarse en modelos visuales. En particular, se ha propuesto reciemntemente el uso de descriptores basados en funciones de Gabor para el análisis de texturas [84]–[88]. Por otra parte, las características de autosimilitud e invarianza de un esquema de Gabor lo convierten en un entorno adecuado para analizar problemas de reconocimiento invariante (ante rotaciones o cambios de escala). Un aspecto particular de las invarianzas ante cambios de escala es el estudio y caracterización de imágenes en base a parámetros fractales[89]–[91]. Para la estimación de la dimensión fractal de una imagen se han propuesto métodos basados en el análisis local de Fourier (representación conjunta) y en análisis multiresolución. Dado que un esquema de Gabor basado en el SV incorpora estas dos capacidades, se puede usar con este fin [88].

Una vez establecido este marco de referencia, pasamos a exponer los objetivos del presente trabajo, y a describir las distintas partes de que está compuesto.

1.1 Objetivos y estructura de la presente tesis

La motivación principal de esta tesis es la de añadir más evidencias en favor de la hipótesis de que esquemas basados en modelos del sistema visual humano pueden constituir entornos multipropósito para aplicaciones de codificación y procesado de imágenes y visión artificial. En concreto nos proponemos en una primera etapa la definición y puesta a punto (implementación) de un esquema de representación de imágenes basado en funciones de Gabor. Dicha representación es cuasicompleta y da cuenta de las características básicas de la codificación de la información que se produce en la corteza visual.

La segunda etapa consiste en la aplicación del esquema anterior en dos direcciones diferentes. En primer lugar se estudia su eficiencia en tareas de análisis de imágenes, tales como la clasificación y reconocimiento invariante de texturas (y como caso particular de fractales). En segundo lugar, dicho esquema se usa como base para el desarrollo de un modelo realista del sistema visual humano, susceptible de ser empleado tanto para la simulación de efectos visuales, como también como base de sistemas de visión artificial. Este modelo incluye las etapas primarias del proceso visual, desde la óptica del ojo, pasando por el muestreo en la retina, hasta llegar a las células simples de la corteza visual.

Finalmente, se propone un modelo que muestra cómo procesos de aprendizaje pueden justificar el desarrollo y organización de campos receptivos en el cortex, de características similares a las de las funciones de Gabor.

En cuanto a la organización de esta memoria, hemos decidido dividirla en tres partes. La primera es la más general, presentando las nociones básicas y el esquema de Gabor desarrollado. En la segunda, se tratan aplicaciones de dicho esquema al análisis de texturas, mientras que en la tercera se desarrollan modelos más realistas del sistema visual humano. Estas dos últimas partes son lo suficientemente independientes como para permitir una lectura por separado. Así pues, la estructura de la presente tesis es la siguiente:

- Parte I
  - En el **capítulo 2** se hace una introducción a las funciones de Gabor y se establecen los parámetros del esquema de Gabor adoptado, en base a los datos conocidos del SV.
  - La implementación de la transformada de Gabor directa e inversa se explica en el **capítulo 3.** Se presentan dos implementaciones alternativas

en ambos dominios (espacial y de Fourier), así como una comparación con la transformada córtex.

- Parte II
  - En el **capítulo 4** se llevan a cabo aplicaciones de clasificación y segmentación de texturas usando una matriz de descriptores basada en el esquema de Gabor desarrollado previamente.
  - Usando los mismos descriptores, en el **capítulo 5** se estudia el reconocimiento invariante de texturas ante cambios de escala u orientación. Se presenta también un método para estimar la dimensión fractal de imágenes.
- Parte III
  - En el **capítulo 6** se presenta un modelo realista de las etapas primarias del SV restringido a la zona de la fóvea. Dicho modelo incorpora la óptica del ojo, el muestreo por una red hexagonal de espaciado variante, etc. A partir de dicho modelo se reproducen algunos fenómenos observados experimentalmente.
  - Finalmente, en el **capítulo 7** se propone un modelo del sistema visual basado en redes neuronales. Dicho modelo podría explicar el desarrollo y organización de los campos receptivos de células simples del córtex, sin hipótesis previas sobre su forma, en base a procesos de aprendizaje.

Parte I

Capítulo 2

Funciones de Gabor y elección del esquema

En este capítulo presentamos en primer lugar una definición de las funciones de Gabor, sus propiedades matemáticas más relevantes, así como su relación con el sistema visual, razonando el porqué de su elección como base de nuestro esquema. En la segunda sección hacemos una descripción detallada del esquema escogido, justificando la elección de los diferentes parámetros por su similitud con datos experimentales del sistema visual humano. Finalmente, en la sección 2.3, se incluye una comparación entre los parámetros escogidos y los de otros esquemas similares existentes en la literatura.

2.1 Funciones de Gabor

Este tipo de funciones reciben su nombre de D. Gabor, quien las propuso en 1946 [50] para el tratamiento de señales unidimensionales. Presentan algunas interesantes propiedades como el hecho de estar localizadas tanto en el dominio espacial como de frecuencias, y minimizar (en una métrica L2) la relación de incertidumbre entre su posición y frecuencia espacial. Dadas estas características, Gabor las propuso como un método óptimo de descomponer una señal en cuantos de información o "logons", para su posterior codificación o tratamiento. Mucho más recientemente, este tipo de funciones fueron introducidas en el campo de la visión (por Marcelja [38] en el caso unidimensional y Daugman [51] en el bidimensional) por su conveniencia como modelos de campos receptivos de células simples en el córtex visual [39][41]. El campo receptivo de una neurona del sistema visual es aquella zona del campo visual a la que ésta es sensible. Dicho campo suele

componerse de una serie de zonas excitatorias (respuesta positiva) e inhibitorias.

Una función de Gabor es un paquete de ondas gaussiano, es decir, una exponencial compleja, de frecuencia dada, con una envolvente gaussiana que la localiza espacialmente. En el caso bidimensional [51], la expresión de una función de Gabor centrada en el punto x_0, y_0 es:

$$g_{x_0,y_0,f_0,\theta_0}(x,y) = g_{0,0,f_0,\theta_0}(x,y) * \delta(x - x_0, y - y_0) \ ,$$

donde

$$g_{0,0,f_0,\theta_0}(x,y) = e^{-\pi a^2((x\cos\theta_0 + y\sin\theta_0)^2 + \gamma^2(y\cos\theta_0 - x\sin\theta_0)^2)} e^{i2\pi f_0(x\cos\theta_0 + y\sin\theta_0) + \phi} \ , \quad (2.1)$$

con $*$ representando un producto de convolución.

En la expresión anterior se han usado coordenadas cartesianas en el dominio espacial y polares en el de frecuencias. El uso de estas coordenadas es conveniente dado que la organización de nuestro esquema, al igual que otros similares [42][40][43] esta basado en descomponer las imágenes en canales de distinta frecuencia y orientación (f_0, θ_0). Dentro de cada canal, la posición espacial se da en coordenadas cartesianas (x_0, y_0) (aunque esto es por conveniencia, ya que por ejemplo en la retina también se suelen usar coordenadas polares -excentricidad y meridiano- para especificar la posición espacial). La función de Gabor de la Ec. (2.1) está localizada tanto en el dominio espacial (x_0, y_0) como de frecuencias (f_o, θ_0), y los parámetros que definen su forma son : a, que está relacionada con el ancho de banda radial, y γ, el factor de forma, que es la relación entre los anchos de banda angular y radial. Una $\gamma < 1$ significa que el ancho de banda radial será mayor que el angular y viceversa. Finalmente, ϕ es una fase constante que causa un desplazamiento de la exponencial compleja con respecto a la gaussiana. Esta fase va a afectar la forma y propiedades de simetría de las componentes real e imaginaria de la función de Gabor, haciendo que éstas sean de simetría par, impar, o asimétricas. El hecho de que el angulo θ_0 sea el mismo en la frecuencia y en la orientación de los ejes de la gaussiana presupone que la orientación de la onda plana sigue los ejes de la gaussiana. En la Fig. 2.1 vemos una representación esquemática de la componente real de una función de Gabor en ambos dominios. Otra interesante propiedad que acentúa el caracter dual de este tipo de funciones es que su transformada de Fourier conserva la misma expresión excepto por una fase lineal.

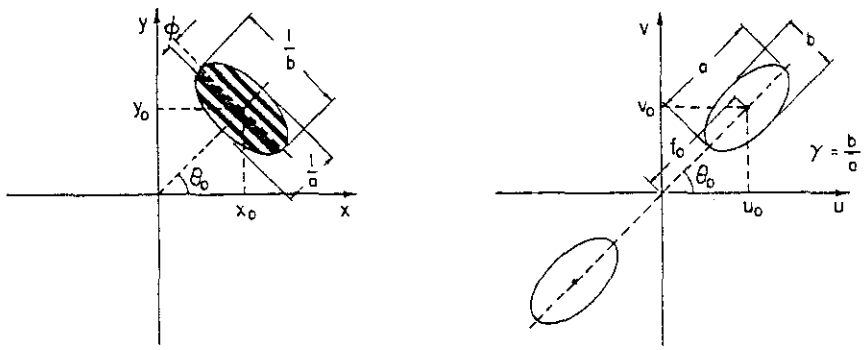


Figura 2.1: Componente real (campo receptivo par) de una función de Gabor (izquierda) y su transformada de Fourier (derecha), mostrando los principales parámetros de la Ec. (2.1).

2.1.1 Funciones de Gabor y campos receptivos

Se han encontrado [39][41] en ciertas áreas del córtex visual pares de neuronas con unas características peculiares. Algunos de estos pares tienen sus campos receptivos situados en la misma posición de la retina y están sintonizados a la misma frecuencia y orientación. Su única diferencia es en la fase ϕ, pero siempre guardando una fase relativa de aproximadamente $\Delta\phi = 90°$; es decir, están en cuadratura de fase. Por otra parte es bien conocido en teoría de la señal [92] que una función analítica (que es el caso de una función de Gabor) tiene sus partes real e imaginaria en cuadratura de fase (relacionadas a traves de una transformada de Hilbert). Todos estos detalles apuntan a que las componentes real e imaginaria de una función de Gabor podrían ser un buen modelo para los campos receptivos de dichas células corticales. Nosotros adoptaremos este modelo en nuestro esquema, por lo que a partir de ahora usaremos el termino campo receptivo (CR) para referirnos en nuestro esquema a las componentes real e imaginaria de una función de Gabor en el dominio espacial.

Aunque Pollen y Ronnen [39] indicaron que los CRs no presentaban una paridad definida o constante (el valor de ϕ está distribuido bastante arbitrariamente), en nuestro modelo tomaremos $\phi = 0$ por simplicidad. De esta forma, los CRs en nuestro esquema presentarán una simetría definida, bien par ($p = 0$, coseno) o impar ($p = 1$, seno) :

$$g_{p=0}(x,y) = e^{-\pi a^2 \cdots} \cos(2\pi f_0 \ldots)$$
$$g_{p=1}(x,y) = e^{-\pi a^2 \cdots} \sin(2\pi f_0 \ldots) \ . \qquad (2.2)$$

Por consiguiente, cada campo receptivo $g_{x_0,y_0,f_0,\theta_0,p}(x,y)$ estará etiquetado con 5 parámetros en total : su localización tanto en el dominio espacial (x_0, y_0) como espectral (f_0, θ_0) y su paridad(p). Es como si pudiesemos etiquetar cada neurona con estos parámetros, quedando así perfectamente caracterizada una vez definidos el ancho de banda a y el factor de forma γ.

2.1.2 Campos receptivos y canales de frecuencia

La respuesta $(R_{x_0,y_0,f_0,\theta_0,p})$ de un campo receptivo a una imagen concreta vendrá dada por el producto interno de las dos funciones [42]:

$$R_{x_0,y_0,f_0,\theta_0,p} = \int\int_{-\infty}^{+\infty} I(x,y) g_{x_0,y_0,\theta_0,f_0,p}(x,y) dxdy \ . \tag{2.3}$$

Combinando Ec. (2.1) y (2.2) con la expresión anterior podemos expresar $R_{x_0,y_0,f_0,\theta_0,p}$ como una convolución :

$$R_{x_0,y_0,f_0,\theta_0,p} = C_{f_0,\theta_0,p}(x_0,y_0) = (I(x,y) * g_{0,0,f_0,\theta_0,p}(x,y))(x_0,y_0) \ . \tag{2.4}$$

Segun el teorema de convolución, la Ec. (2.4) puede calcularse en el dominio de frecuencias espaciales, multiplicando la transformada de Fourier $G_{f_0,\theta_0,p}(f,\theta)$ del campo receptivo $g_{0,0,f_0\theta_0,p}(x,y)$ con la de la imagen. . Dicha función de transferencia es lo que se conoce como un canal en el sistema visual, estando caracterizada por sus propiedades de filtro en el espacio de Fourier: frecuencias de paso, anchos de banda, etc. En general hablaremos de canales en el dominio de Fourier y de campos receptivos en el espacial, mientras que la palabra filtro se reservará para una implementación particular en cualquiera de los dos dominios. Denominaremos salida de un canal $C_{f_0,\theta_0,p}(x,y)$ al conjunto de las respuestas de todas las células que comparten las mismas etiquetas de frecuencia, orientación y paridad. A veces, por una licencia del lenguaje, hablaremos de canales cuando en realidad nos estamos refiriendo a sus salidas. Así sucede cuando hablamos de los distintos canales de orientación y frecuencia de una imagen: en realidad nos referimos a los resultados de aplicar los diversos canales $G_{f_0,\theta_0,p}(f,\theta)$ a la imagen original y no a los canales en sí.

A la hora de calcular la transformación de Gabor, es decir, obtener el conjunto de coeficientes $R_{x_0,y_0,f_0,\theta_0,p}$ a partir de la Ec. (2.4) el teorema de convolución nos permite hacerlo en ambos dominios. Una posibilidad es aplicar directamente los

campos receptivos $g_{x_0,y_0,f_0,\theta_0,p}(x,y)$ a la imagen en el dominio espacial. Tambien podremos filtrar usando $G_{f_0,\theta_0,p}(f,\theta)$ en el dominio de frecuencias y evaluar posteriormente la salida de cada canal en el punto x_0, y_0. Hablaremos de las ventajas de uno u otro método en el capítulo siguiente.

Al aplicar los CRs a una imagen según las Ecs. (2.3) ó (2.4), obtenemos información codificada de una forma especial ($R_{x_0,y_0,f_0,\theta_0,p}$). Una de las características de esta representación es que es redundante, al estar la salidas de los canales pares ($p = 0$) e impares ($p = 1$) relacionadas a través de una transformada de Hilbert [92]. Varias aplicaciones pueden derivarse de esta descomposición, tales como análisis, codificación, etc. De hecho existen numerosas evidencias de que el sistema visual lleva a cabo una codificación de similares características. En nuestro modelo, cada una de estas muestras de información (llamadas cuantos o "logons", unidades óptimas de empaquetamiento de información en la teoría original de Gabor [50]) se asocia con la respuesta de una célula del cortex visual, etiquetada con una posición en ambos dominios y una paridad [39].

2.2 Parámetros del esquema

En esta sección se presenta la elección de los parámetros que definirán nuestro modelo particular. Los criterios seguidos han sido fundamentalmente dos: primeramente, adecuación a lo que se conoce actualmente del sistema visual tanto por métodos psicofísicos [42][40][93][94] [43], como fisiológicos [36][39][41], y en segundo lugar hacer nuestro modelo lo más simple posible.

Los parámetros a considerar incluyen las posiciones de nuestros campos receptivos en ambos dominios (x_0, y_0 y f_0, θ_0) y su forma (a, γ). Debe entenderse que dado que estamos tratando con una transformación conjunta en ambos dominios, estaremos limitados por el teorema de muestreo (equivalente al principio de incertidumbre), por lo que los parámetros citados anteriormente estarán ligados entre sí. Una situación con pocos canales (pocos puntos de muestreo f_0, θ_0 en el dominio de Fourier) y con un ancho de banda (a, γ) grande requeriría un muestreo fino en el dominio espacial x_0, y_0. Muchos canales con una anchura de banda pequeña precisarían de un muestreo mucho menos fino. Desde el punto de vista teórico, esta elección es arbitraria en lo que respecta a la completitud del modelo. Sin embargo sabemos que la capacidad de nuestro sistema visual en cuanto a la localización de objetos en el espacio es alta, mientras que nuestra habilidad para el análisis de frecuencias espaciales es relativamente pobre. Esto

indica una situación con relativamente pocos canales de frecuencia, y dado que estamos imitando el sistema visual adoptaremos este enfoque. Primeramente decidiremos sobre el conjunto (reducido) de canales (elección de f_0, θ). Como se mostró en la sección anterior, podemos prescindir de x_0, y_0 a la hora de definir nuestros canales ($G_{f_0,\theta_0,p}(f,\theta)$), o los campos receptivos base que se usarán en la convolución ($g_{0,0,f_0\theta_0,p}(x,y)$). Una vez diseñado un canal, el teorema de muestreo nos indicará qué espaciado es preciso aplicar en el dominio espacial para no perder informacíon. En esos puntos aplicaremos nuestros campos receptivos base para calcular los coeficientes $R_{x_0,y_0,f_0,\theta_0,p}$.

2.2.1 Frecuencias (f_0)

Hemos adoptado para nuestro esquema una disposición logarítmica de los canales de frecuencia. Consecuentemente, para garantizar un buen recubrimiento del espacio de Fourier, la anchura de banda de cada canal habrá de ser proporcional a su frecuencia. Hemos considerado tres razones fundamentales para esta elección. En primer lugar, todas las evidencias apuntan a que en el sistema visual los canales de frecuencia se distribuyen en un esquema de octavas (logarítmico) [42][39] [41][45][40]. Además, el espectro de potencias de imágenes naturales tiende (en promedio) a decaer de tal forma [95] que la energia por octava tiende a ser constante [96]. De esta forma, una distribución logarítmica en octavas es óptima para trabajar con imágenes naturales, haciendo que las diversas células del córtex den respuestas de un orden similar Por último, este esquema logarítmico permite una fácil y eficiente implementacíón piramidal . Estos sistemas piramidales tienen interesantes propiedades [66] y suponen un importante ahorro de tiempo de calculo. Un cuidadoso enfoque teórico de estos métodos multiresolución puede encontrarse en la ref. [61].

Por lo tanto, en nuestro esquema usaremos canales de frecuencia distribuidos en octavas ($\log_2$), siendo la frecuencia a la que está centrado cada canal la mitad de la del anterior. Para especificar las frecuencias de nuestros canales sólo tendremos que asignar la frecuencia maxima f_{max} y el número de canales N. Las demás frecuencias serán $f_n = 2^{-n} f_{max}$, con $n = 0, ..., N-1$. En una imagen discreta, la frecuencia más alta posible es la frecuencia de Nyquist $f_N = 1/2$ ciclos/píxel. Para nuestro esquema hemos adoptado el valor de $f_{max} = f_N/2 = 1/4$ ciclos/píxel en el canal de máxima frecuencia, para mantener la similitud entre la forma de los filtros de la frecuencia más alta y los de frecuencias inferiores. Si hubiesemos escogido $f_{max} = f_N$ la forma de este canal hubiera sido diferente de la de los otros, debido

al corte en frecuencias impuesto por el límite de Nyquist. Como consecuencia de esto, los filtros de frecuencias bajas no hubiesen podido ser generados cambiando la escala de los primeros, perdiendo así la propiedad de autosimilitud que es un aspecto fundamental en un esquema piramidal.

Para comparar las frecuencias de nuestros canales con las del sistema visual humano podemos suponer que cada píxel de la imagen corresponde a la respuesta de un cono. Entonces, la frecuencia de Nyquist del sistema visual será la correspondiente a un periodo de dos conos. Dado que la distancia entre conos es del orden de 0.5 minutos de arco [23][13][14], dicha frecuencia de Nyquist será $f_N = \frac{1}{2T} = 60$ ciclos/grado en el centro de la fovea (asumiendo un muestreo rectangular). De acuerdo con esta comparación, nuestro canal de frecuencia máxima correspondería a uno visual sintonizado a unos 30 ciclos/grado, lo que se aproxima a estimaciones experimentales [39].

En cuanto al numero de canales, se han propuesto valores entre 4 y 8. Estudios experimentales [42][39] sugieren 4 ó 6 canales en el sistema visual. Dado que pretendemos que nuestro modelo sea tan simple como sea posible, aunque sin perder el contacto con la realidad del sistema visual, hemos optado por 4 canales de frecuencia. Este es también el numero de canales usado en la mayoría de las aplicaciones y esquemas existentes [53][84][59].

2.2.2 Ancho de banda radial (a)

El parámetro a en la Ec. (2.1) determina la anchura de banda en la dirección radial de cada canal, determinando el grado de solapamiento entre canales adyacentes y el mayor o menor recubrimiento del dominio de Fourier. Como se indicó anteriormente, en un esquema logaritmico a no es constante, sino proporcional a la frecuencia central f_n del canal; de esta forma, la anchura de banda es constante cuando se expresa en octavas. Si establecemos que $a = kf_n$, es fácil mostrar que la anchura a media altura en octavas es:

$$\Delta = \log_2\left(\frac{1 + k\sqrt{\left(\frac{\ln 2}{\pi}\right)}}{1 - k\sqrt{\left(\frac{\ln 2}{\pi}\right)}}\right) , \tag{2.5}$$

donde k es el factor de proporcionalidad entre a y f_n. La expresión anterior es inmediata de obtener si consideramos la gaussiana de la Ec. (2.1) en un eje logarítmico.

Los datos experimentales [36][39][43] muestran un rango de anchos de banda que va desde 0.5 a 2 octavas, con una media de aproximadamente 1.3 octavas. Esta valor medio es además óptimo ya que esta anchura de banda hace que las envolventes gaussianas de canales adyacentes se encuentren a una altura de $\sqrt{(\frac{1}{2})}$ obteniendose así un buen recubrimiento del espectro de potencias ($\sum_i |G_i|^2 \simeq 1$). En algunos modelos se adopta el valor $\Delta = 1$ octava. En este caso el solapamiento es menor, canales contiguos se encuentran a altura $\frac{1}{2}$ y consecuentemente $\sum_i |G_i| \simeq 1$. Nosotros nos hemos decidido por este último valor, que equivale, segun Ec. (2.5) a una constante de proporcionalidad $k = \sqrt{(\pi)}/(3\sqrt{(ln2)}) = 0.71$, por dos razones principales. Primeramente, para aplicaciones de análisis de texturas (Cap. 4 y 5), un excesivo solapamiento disminuiría la potencialidad del esquema: si los canales fueran demasiado anchos, dos patrones con diferentes caracteristicas espectrales podrían excitar la misma célula, y aparecerían como idénticos desde el punto de vista de nuestro esquema. En segundo lugar, si la suma de los filtros de Gabor es próxima a la unidad la imagen podría ser recuperada con una buena aproximación, sólo con sumar la salida de los diferentes canales (Cap. 3).

2.2.3 Factor de forma (γ)

El ancho de banda angular se determina al fijar el factor de forma γ de los canales. Algunos datos experimentales [36] muestran que los canales en el sistema visual están elongados en la dirección radial ("filtros de almendra"), lo que indicaría que γ es menor que la unidad. Algunos valores citados son de 0.6 ó 0.7, aunque algunos autores [93] apuntan que el valor de γ podría variar para orientaciones específicas, proponiendo una mayor resolución angular para los ejes verticales y horizontales. Sin embargo, como en la mayoria de los modelos [43] [84], hemos escogido $\gamma = 1$ por simplicidad. Dicho valor simplifica bastante la expresión (Ec. (2.1)) de una célula de Gabor, que se convierte en :

$$g_{x_0,y_0,f_0,\theta_0,p=0}(x,y) = e^{-\pi(kf_0)^2((x-x_0)^2+(y-y_0)^2)} \cos(2\pi f_0((x-x_0)\cos\theta_0+(y-y_0)\sin\theta_0)) \quad , \tag{2.6}$$

donde hemos substituido a por kf_0 y γ por 1. De esta forma, la célula de Gabor depende ahora sólo de sus etiquetas $x_0, y_0, f_0, \theta_0, p$ (ya que $k = 0.71$).

Una vez que fijamos el valor de γ, queda indirectamente determinado también el número de canales de orientación para cada frecuencia. Como se ve en la Fig. 2.1, la anchura de banda angular $\Delta\theta$ depende de $\gamma\frac{a}{f_0}$. Dado que a es proporcional

a f_0 ($a = kf_0$), $\Delta\theta$ es constante para todas las frecuencias $\Delta\theta = \gamma k = k = 0.71$ radianes. Un recubrimiento óptimo del dominio de frecuencias espaciales se obtendrá cuando el numero de canales sea próximo a $\pi/\Delta\theta$. Para el criterio de solapamiento escogido ($\sum_{all} |G_i| \simeq 1$) esto da un valor ($\pi/\Delta\theta = \pi/0.71 = 4.4$) entre 4 y 5 canales. Para nuestro esquema hemos adoptado 4 canales de orientación por frecuencia (centradas en las 4 direcciones principales (horizontal, vertical, y las dos diagonales) [53][84] que proporcionan un buen recubrimiento del plano de Fourier.

2.2.4 Posición espacial (x_0, y_0)

Como ya se indicó, la frecuencia de muestreo del dominio espacial está determinada por las características de los distintos canales. La elección cuidadosa del intervalo de muestreo adecuado evitará tanto problemas de aliasing como de redundancia [54][55]. Los canales de alta frecuencia precisarán de un muestreo fino, mientras que para los de baja frecuencia bastará con muy pocas muestras. Lo que sí hemos de decidir es el tipo de muestreo que se llevará a cabo en nuestro esquema. De nuevo hemos optado por la simplicidad y escogido un muestreo en una red rectangular con espaciado uniforme. En el Capítulo 6, en el marco de un modelo mucho más realista de la fóvea, usaremos un muestreo hexagonal. En cualquier caso los resultados que se muestran en los siguientes capítulos no pierden generalidad por el uso de un muestreo rectangular.

En la Fig. 2.2 se muestra el aspecto final de los canales resultantes en ambos dominios. Los 4×4 los campos receptivos básicos (simetría par, $p = 0$) en el dominio espacial aparecen en la Fig. 2.2a. En la Fig. 2.2b se aprecia el grado de cubrimiento del dominio de Fourier por los 4×4 canales. Los campos receptivos de alta frecuencia (arriba en la Fig. 2.2a) muestran un tamaño pequeño, mientras que sus correspondientes canales tienen por contra el mayor soporte en el dominio de frecuencias (canales más exteriores en Fig. 2.2b). Lo que se presenta en la figura es el conjunto básico de elementos con los que se forma la base de funciones de una transformación conjunta, la transformada de Gabor. Todas estas funciones base cubren un área igual en el dominio conjunto. Esto se puede apreciar con mayor facilidad en la Fig. 2.3, donde se representa el dominio conjunto espacio (horizontal) - espectral (vertical) en el caso unidimensional. Nótese la escala logarítmica en el eje de frecuencias y cómo el muestreo en el eje espacial se ajusta al ancho de banda de cada canal.

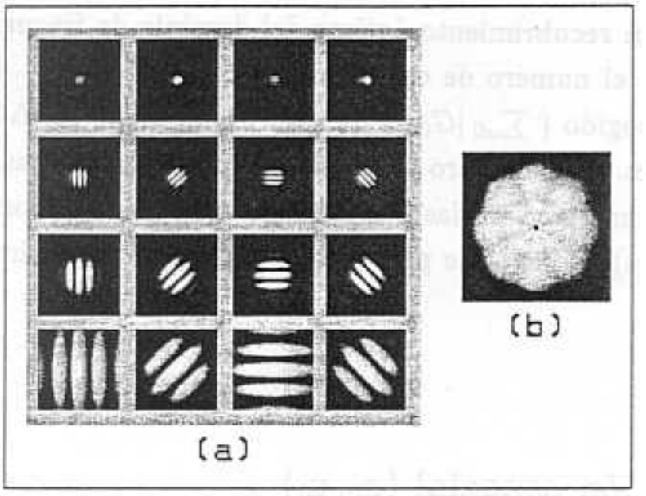


Figura 2.2: Las 4×4 funciones de Gabor básicas usadas en la transformada de Gabor: (a) los campos receptivos pares en el dominio espacial; (b) cubrimiento del dominio de Fourier por los canales de frecuencia.

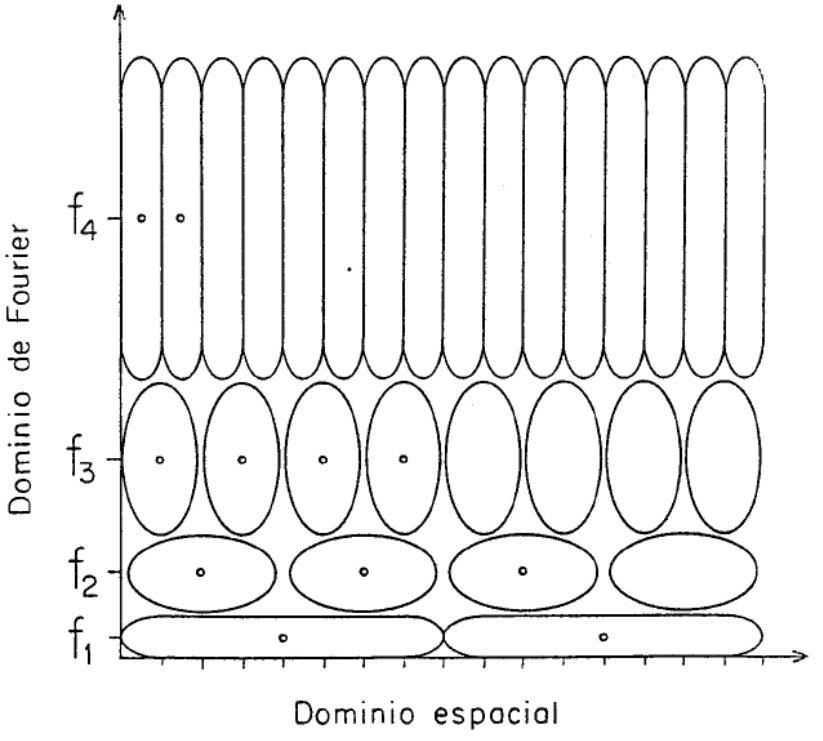


Figura 2.3: Recubrimiento del dominio conjunto (1D) con funciones de Gabor, todas ellas teniendo la misma área y minimizando el principio de incertidumbre posición espacial/frecuencia espacial.

2.3 Comparación con otros esquemas

En la Tabla 2.1 se hace una comparación entre el esquema de Gabor propuesto y otros modelos equivalentes publicados previamente. También se incluyen referencias a datos experimentales tanto psicofísicos como fisiológicos. La Tabla muestra las siguientes características: numero de canales de orientación y frecuencia, valores propuestos de frecuencias para dichos canales (en ciclos/grado en datos experimentales y ciclos/píxel en modelos que manejan imágenes discretas), anchura de banda radial y factor de forma, y si el esquema incluye o no una distribución logarítmica de los canales de frecuencia. Comparando los distintos modelos pre-

Tabla 2.1: Comparación entre diferentes modelos y datos experimentales del sistema visual.

Modelos	Log	a	Δ (oct)	γ	N Frec	Freq (cic/pix)	N orient
Cortex[53]	SI	$0.66f$	$\simeq 1$	1.17	4	3/8 ... 3/64	4
Turner[84]	SI[a]	constant	—	1	4	1/4 ... 1/32	4
Sutter[86]	SI	$0.66f$	$\simeq 1$	1	13	2/7 ... 1/112	3
QMF[65]	SI	$0.66f$	$\simeq 1$	1.6	4	3/8 ... 3/64	3
Esta tesis	SI	$0.71f$	1	1	4	3/8 ... 3/64	4
Datos	**Log**	a	Δ (oct)	γ	N Frec	Frec (cic/grad)	N orient
Fisiol.[b]	SI	$0.91f$	1.3	0.5-3[c]	4	16 ... 0.5 [d]	5-6
Psicof.[e]	SI	—	$\simeq 1$	0.5-0.7	4[e]	8 ... 1 [f]	4-8

[a] Este esquema usa una distribución logarítmica, pero a es constante, no pudiendo ser expresada en octavas.
[b] Datos de [39][36].
[c] Se han presentado resultados muy diferentes.
[d] En monos.
[e] Datos de [43][42][40][93][94].
[f] Wilson y Bergen [42] estaban limitados a buscar por debajo de 16 ciclos/grado.

sentados con datos experimentales, se observa que todos ellos tienden lógicamente a simplificar la mayoría de las características del sistema visual. El número de canales tanto en frecuencia como en orientación se reduce a 4 ó 5 [53][84]. El factor de forma γ es cercano a 1, y la anchura de banda suele ser $\Delta \simeq 1$ octava. Esa tendencia simplificadora ha sido adoptada también en nuestro esquema. Sutter et al. [86] proponen un esquema con un gran solapamiento en frecuencias y por contra, un escaso cubrimiento en orientaciones. Esto hace que su modelo,

aunque útil para análisis, no sea válido para la recuperación de la imagen. El esquema de filtros de espejo en cuadratura (QMFs) [65] proporciona una transformación no redundante, lo que le hace especialmente adecuado para aplicaciones de codificación.

Los esquemas más similares al nuestro son los propuestos por Turner [84] y la transformada córtex [53]. La principal diferencia con el primero es que aunque Turner usa una distribución logarítmica de canales de frecuencias (cuyos picos coinciden con los nuestros) la anchura de sus canales no es proporcional a su frecuencia (usa a constante). De esta manera, como el propio autor indica, el recubrimiento del plano de Fourier está lejos de ser completo, lo que impide una correcta recuperación de la imagen. La transformada cortex es también muy similar. La principal diferencia es que no usa funciones de Gabor, sino filtros especificamente diseñados para asegurar la completitud de la transformada. En un principio, puesto que fué diseñada para codificación, sólo incorporaba filtros pares (suficientes para reconstruir la imagen), pero posteriormente [54] incorporó también el uso de filtros analíticos.

Capítulo 3

Transformada directa e inversa de Gabor

En el capítulo anterior se presentaron las características concretas de las funciones de Gabor que van a ser la base de nuestro modelo. Asimismo se definió (Ec. (2.3)) cual es la respuesta de una célula, con un determinado campo receptivo modelado por una función de Gabor, a una imagen dada. Sin embargo, cuando se trata de desarrollar un modelo computacional, es fundamental que los cálculos necesarios para obtener los resultados deseados puedan ser realizados lo más eficientemente posible, con objeto de reducir el coste computacional. De ahí que la implementación práctica del modelo sea un tema muy importante. Entendemos por implementación el conjunto de algoritmos y procesos digitales necesarios para la aplicación práctica del modelo a imágenes cualesquiera. En la primera sección presentamos dos implementaciones alternativas en ambos dominios (espacial y de Fourier) para el rápido cálculo de las respuestas de los campos receptivos a una imagen dada. Se ha prestado una especial atención a la implementación en el dominio espacial (3.1.2), por ser más novedosa y presentar a nuestro juicio ventajas en ciertas aplicaciones.

Uno de los propósitos del presente trabajo es proponer el uso de un esquema de Gabor (o de modelos similares) como entornos generales de procesado de imágenes para diversas aplicaciones. Así pues, no estamos sólo interesados en extraer información sobre la imagen (análisis) sino también en aplicaciones que precisen su reconstrucción (síntesis). Por lo tanto, debemos considerar también la implementación de la transformación inversa, a lo que dedicamos la Sección 3.2, considerando aspectos como la complejidad de dicha operación, el error en la recuperación y cómo minimizarlo, etc. Finalmente, en la Sección 3.3 hacemos una comparación de nuestro esquema de Gabor y la transformada córtex [53][54], con-

siderando la recuperación de la imágen en diversas circustancias y sus posibles aplicaciones.

3.1 Implementación de la transformada directa

Hallar la transformada de Gabor de una imagen equivale a calcular el conjunto de coeficientes $(R_{x,y,f,\theta,p})$ definidos en las Ecs. (2.3) ó (2.4). A partir de ahora trabajaremos con imagenes discretas (aptas para ser manejadas por un ordenador), por lo que dicha ecuación se convierte en :

$$R_{i,j,f_m,\theta_m} = \sum_k \sum_l I(k,l)\ g_{0,0,f_m,\theta_m,p}(i-k, j-l)\ . \tag{3.1}$$

Esta expresión es una convolución discreta que podemos implementar tanto en el dominio espacial (convolucionando con el campo receptivo $g_{0,0,f_0,\theta_0,p}$) o en el de Fourier (filtrando con el canal f_0, θ_0, p), por el teorema de convolución.

La mayoría de los esquemas similares usan una implementación en el dominio de Fourier [53][54]. Sin embargo, nosotros propondremos también una implementación alternativa en el dominio espacial. Hay varias razones para considerar tal alternativa. Una razón de fondo es que el sistema visual no utiliza transformadas de Fourier, sino que opera con campos receptivos en el dominio espacial. Además, cuando se trabaja en el dominio de Fourier uno está limitado a un procesado global. Esto no tiene importancia cuando se trata de recuperar la imagen (Sección 3.2), pero no así cuando sólo se está interesado en el análisis (como en las aplicaciones presentadas en los Capítulos 4 y 5). En ese caso, una implementación espacial nos permite calcular la respuesta de nuestros campos receptivos sólo en la región o puntos en los que estamos particularmente interesados, evitandonos tener que calcular la transformación completa. Otro aspecto a considerar es que en el caso de usar un muestreo no uniforme (Capítulo 6) nos está vedada la implementación en el dominio de Fourier. El mayor inconveniente de la implementación espacial es el mayor error que se introduce en sus distintos pasos, pero como veremos a continuación tales errores son pequeños y pueden ser minimizados. Finalmente, un punto muy importante es el coste computacional como se discutirá más adelante.

3.1.1 Dominio de Fourier

La implementación en el dominio de frecuencias espaciales ha sido la más común en este tipo de transformadas[53][86]. Los distintos canales de frecuencia son inmediatamente asimilados a filtros lineales en el dominio de Fourier (en el caso del esquema de Gabor serán ventanas gaussianas). Para implementar dichos filtros basta discretizar la expresión de la transformada de Fourier del campo receptivo de la Ec. (2.6). Nótese que dicha ecuación sólo es válida si $\gamma = 1$, por lo que si queremos tener en cuenta la posibilidad de diferentes anchuras radiales y angulares deberemos referirnos a Ec. (2.1). La expresión discreta de un filtro analítico de Gabor de frecuencia f ciclos/píxel y orientación θ grados para un tamaño de imagen de $N \times N$ píxeles vendrá dada por:

$$G_{f,\theta}(u,v) = \exp(-\pi \frac{(u - (\frac{N}{2} + N f \cos(\theta)))^2 + (v - (\frac{N}{2} + N f \sin(\theta)))^2}{a^2}) \quad . \quad (3.2)$$

En esta ecuación u y v varían entre 0 y N-1 y a es la anchura de banda que es proporcional a la frecuencia, cuyo valor es de $a = 0.7 f N$ (nótese que todos las variables de frecuencia espacial llevan un factor N, indicado por el tamaño de la implementación del filtro.

En cuanto a los canales pares e impares, dado que la expresión anterior es un filtro analítico, serán las componentes real e imaginarias obtenidas una vez filtrada la imagen. Podríamos haber construido un filtro para cada paridad sin más que reflejar especularmente el filtro anterior sobre el origen [92] (con signo positivo para los canales pares y negativo para los impares). Sin embargo, usando un filtro analítico podemos calcular ambos canales con una sóla transformada compleja.

Una vez que tenemos los filtros basta con multiplicarlos sucesivamente por la transformada de Fourier de la imagen considerada y hacer la transformada inversa del producto obteniendose así las salidas de los diferentes canales. Se hace uso de un esquema en pirámide para que la salida de los canales de frecuencias más bajas se pueda calcular haciendo transformadas de Fourier de tamaño cada vez menor (ya que sus filtros tienen un menor soporte en el dominio de Fourier). Además, de esta forma se va reduciendo el número de coeficientes para los canales de frecuencia más baja. Una explicación más detallada del proceso de aplicación de los filtros puede encontrarse en [53][54].

Dado que nuestro esquema es logarítmico no hemos tomado en cuenta la frecuencia cero o componente continua de las imágenes (nótese el agujero para la

frecuencia cero en la Fig. 2.2). Hay varias maneras de tener esto en cuenta. Para posibilitar la recuperación de la imagen original (Sección 3.2) se ha incorporado un filtro residual de frecuencia nula [53]. En nuestra implementación tal filtro es simplemente una pequeña ventana gaussiana alrededor del cero de frecuencias. Como anchura de dicha gaussiana hemos empleado la del último canal de frecuencias, ya que comprobamos que cubría razonablemente la zona considerada.

3.1.2 Dominio espacial

Diseño de los "campos receptivos"

La implementación espacial se basa en el diseño de máscaras de convolución que tengan una respuesta en frecuencias lo más parecida a la de los canales deseados. La máscara resultante actuaría como un campo receptivo al aplicarse en un punto de una imagen dada. De las Ec. (2.4) y (3.1) se deduce que sólo necesitamos diseñar una pareja de filtros para cada canal $(f_0, \theta_0, p = 0, 1)$, que luego se aplicarán en los distintos puntos de la imagen. Es más, debido a la autosimilitud de nuestro esquema y al uso de un método en pirámide, sólamente tendremos que diseñar los filtros sintonizados a la frecuencia máxima $f_0 = f_{max} = 1/4$. Los correspondientes a frecuencias más bajas pueden obtenerse con un cambio de escala a partir de los del nivel inicial. De la misma forma, los filtros con diferentes orientaciones son versiones rotadas de uno de ellos. De esta forma solo tendremos que generar una pareja de filtros teniendo $x_0 = 0$, $y_0 = 0$, $f_0 = 1/4$, $\theta = 0°$ y las dos paridades: fase coseno ($p = 0$, par) y seno ($p = 1$, impar). A partir de ellos podemos generar el conjunto completo a partir de cambios de escala y rotaciones.

El tamaño del filtro que diseñemos ha de ser el resultado de un compromiso entre la tendencia a reducir el coste computacional (tamaño más pequeño) y la necesidad de mantener la deseada respuesta en frecuencias para el canal. Con este propósito se estudió cual era el tamaño mínimo para que la máscara mantuviera una respuesta en frecuencias lo suficientemente parecida a la del canal original. Para escoger el tamaño óptimo, la pareja de filtros con los parámetros antes citados se ha muestreado y cuantificado con 8 bits, anulandose sus valores fuera de una ventana de un determinado tamaño y calculandose entonces su transformada de Fourier. A continuación se han comparado los canales resultantes con los originales en base a la relación señal ruido entre la energía del error y la del canal original. Esto se ha repetido para diferentes tamaños de ventanas, desde 5×5 hasta 16×16 píxeles. Los resultados muestran que la ventana 15×15 no presenta un error

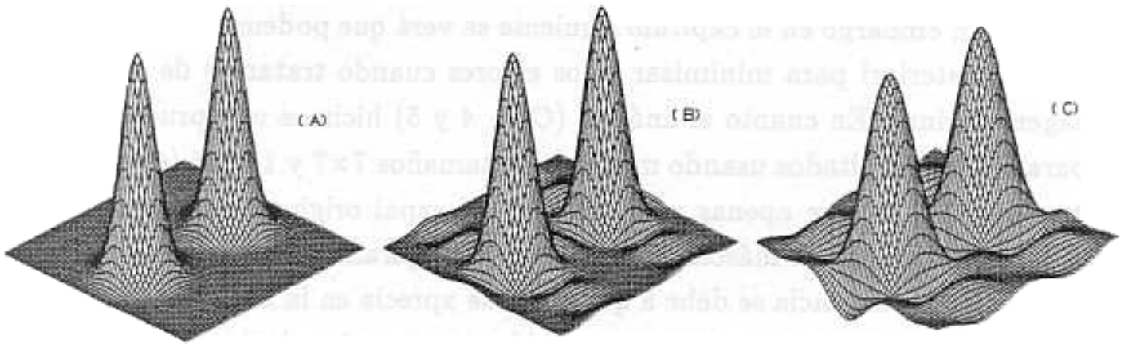


Figura 3.1: Canal teórico en el dominio de Fourier (a) y los obtenidos a partir de máscaras 9×9 (b) y 7×7 (c) píxeles.

apreciable (SNR $\simeq$ 50dB). Las máscaras con un tamaño 9×9 o mayor (SNR >= 20dB) mantienen un gran parecido con el canal original. Sin embargo, el filtro 7×7 empieza a mostrar los efectos de la aplicación de la ventana, los cuales afectan a su anchura de banda (ensanchamiento) y a la forma del canal. En la Fig. 3.1 se puede ver una comparación entre los canales obtenidos a partir de máscaras 7×7 y 9×9 con el canal teórico.

$$\begin{pmatrix} 0 & -3 & 0 & 13 & 0 & -3 & 0 \\ 0 & -8 & 0 & 16 & 0 & -8 & 0 \\ 0 & -13 & 0 & 22 & 0 & -13 & 0 \\ 0 & -16 & 0 & 26 & 0 & -16 & 0 \\ 0 & -13 & 0 & 22 & 0 & -13 & 0 \\ 0 & -8 & 0 & 16 & 0 & -8 & 0 \\ 0 & -3 & 0 & 13 & 0 & -3 & 0 \end{pmatrix} \quad \begin{pmatrix} 3 & 0 & -8 & 0 & 8 & 0 & -3 \\ 6 & 0 & -11 & 0 & 11 & 0 & -6 \\ 8 & 0 & -15 & 0 & 15 & 0 & -8 \\ 10 & 0 & -16 & 0 & 16 & 0 & -10 \\ 8 & 0 & -15 & 0 & 15 & 0 & -8 \\ 6 & 0 & -11 & 0 & 11 & 0 & -6 \\ 3 & 0 & -8 & 0 & 8 & 0 & -3 \end{pmatrix}$$

(a)

$$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 19 & 31 & 19 & 0 \\ 1 & 31 & 53 & 31 & 1 \\ 0 & 19 & 31 & 19 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

(b)

Figura 3.2: (a) Máscaras 7×7 usadas en la implementación espacial, correspondientes a los canales par (izqda) e impar (der) de frecuencia máxima y $\theta = 0°$. (b) Máscara paso-bajo 5×5 usada para evitar el aliasing.

A pesar de todo hemos adoptado finalmente la máscara de tamaño 7×7 (Ver Fig. 3.2a) para ahorrar tiempo de ordenador y también porque podía ser implementada más facilmente con el hardware de que disponíamos. Se usó una tarjeta de procesado de imágenes (MVP-AT Matrox sobre un ordenador personal de tipo AT) la cual nos permitía hacer convoluciones más rápidas si el tamaño de la máscara de convolución era 7×7 o menor. Los errores en la respuesta en frecuencias de los filtros podrían ser reducidos con un diseño más sofisticado de

los filtros. Sin embargo en el capítulo siguiente se verá que podemos utiliar ciertas técnicas a posteriori para minimizar estos errores cuando tratamos de recuperar la imagen original. En cuanto al análisis (Cap. 4 y 5) hicimos una prueba previa comparando los resultados usando m scaras de tamaños 7×7 y 15×15 (ésta última como referencia, ya que apenas se diferencia del canal original). Los resultados (ver Tabla 4.1) usando la máscara 15×15 son sólo ligeramente mejores (alrededor de un 6%). Esta diferencia se debe a que como se aprecia en la Fig. 3.1 los canales correspondientes a la máscara 7×7 han sufrido un ensanchamiento con respecto a los originales. Esto provoca una pérdida de resolución, que empeora los resultados del análisis. Sin embargo, la pequeña mejora en la discriminación obtenida en el caso 15×15 no compensa cuadruplicar el tiempo de cálculo.

En este sentido podemos ahora, una vez decidido el tamaño de la máscara en el dominio espacial, comparar las dos implementaciones en cuanto a coste computacional. El número de operaciones para cada convolución en dos dimensiones será proporcional a n^2D^2, donde n es el tamaño de la máscara de convolución y D la dimensión de la imagen. Para una transformada de Fourier (usando FFT) el tiempo necesario es proporcional a $kD^2 \log_2(D^2)$, donde un valor típico para k es del orden de 2 ó 3. Con imágenes de dimensión 128 ó 256 píxeles y un tamaño de 7 para la máscara de convolución ambos métodos tienen un coste computacional equivalente. Por lo tanto para decidir cual de ellos se quiere usar se debe pensar en las razones aducidas anteriormente en cuanto a sus ventajas e inconvenientes para las distintas aplicaciones.

Implementación piramidal

Una vez que la pareja de filtros de frecuencia más alta han sido diseñados, el resto puede obtenerse a partir de ellos, por rotación y cambio de escala. Los filtros de frecuencias bajas precisarán un tamaño de máscara mayor en el dominio espacial por lo que calcular sus respuestas via convolución llevaría un tiempo excesivo . Para evitar este inconveniente (que eliminaría la igualdad computacional de ambos métodos) se ha implementado un esquema piramidal [66]. En vez de aplicar versiones ampliadas de los filtros a la imagen original es mucho más rápido aplicar los mismos filtros a una versión submuestreada de la imagen, al igual que se hace en la pirámide laplaciana. Para evitar el aliasing cuando reducimos la imagen, aplicamos un filtro paso-bajo 5×5 (Fig. 3.2b) antes del submuestreo. Cuando este filtro se aplica a los sucesivos niveles obtenemos una pirámide gaussiana [59]. Si se usa un filtro paso-bajo ideal se puede demostrar que este proceso

es matemáticamente equivalente a aplicar los filtros ampliados a la imagen original y posteriormente submuestrear el resultado, tomando una muestra de cada dos en ambas direcciones. Esto es posible ya que los canales de Gabor de bajas frecuencias tienen una banda limitada; consecuentemente, sus salidas pueden ser submuestreadas sin causar aliasing.

Por lo tanto, el proceso completo para obtener la transformada de Gabor con esta implementación espacial consiste en las siguientes etapas:

1. Se aplica el conjunto de máscaras correspondientes a la frecuencia máxima a la imagen original. Esto da el primer nivel de la pirámide, con ocho respuestas diferentes (4 orientaciones y 2 paridades) para cada punto.

2. La imagen se convoluciona con el filtro paso-bajo gaussiano de tamaño 5×5.

3. Diezmamos la imagen tomando una muestra de cada dos en ambas direcciones.

4. Se repite los pasos anteriores $N-1$ veces, donde N es el número de canales, para obtener los N niveles de frecuencia (en nuestro caso N=4).

En todos los niveles de frecuencia obtendremos 8 respuestas (4 orientaciones y 2 paridades) para cada punto donde se aplican los filtros. El número de puntos disminuye en un factor 4 al pasar a una frecuencia más baja, de forma que el número de coeficientes para los distintos niveles es de $8N^2$, $8(N/2)^2$, $8(N/4)^2$ y $8(N/8)^2$ respectivamente.

Si observamos las máscaras de la Fig. 3.2a se aprecia que la suma de sus valores es nula, lo que indica que estos filtros no apreciarán un cambio de la intensidad media de la imagen. Esto indica que, al igual que se comentó en el caso de la implementación de Fourier, el esquema presentado hasta ahora no considera la componente continua. Necesitaremos pues añadir un residuo de baja frecuencia. En una implementación espacial éste puede calcularse fácilmente, después de haber obtenido el resto de los canales, aplicando una vez más el filtro usado para evitar el aliasing y submuestreando la imagen (es decir repitiendo pasos 2 y 3 una vez más). El número de coeficientes de este residuo será de $(N/16)^2$.

De esta forma, cuando aplicamos cualquiera de las dos implementaciones a la imagen original obtenemos una versión codificada de dicha imagen con una estructura piramidal.

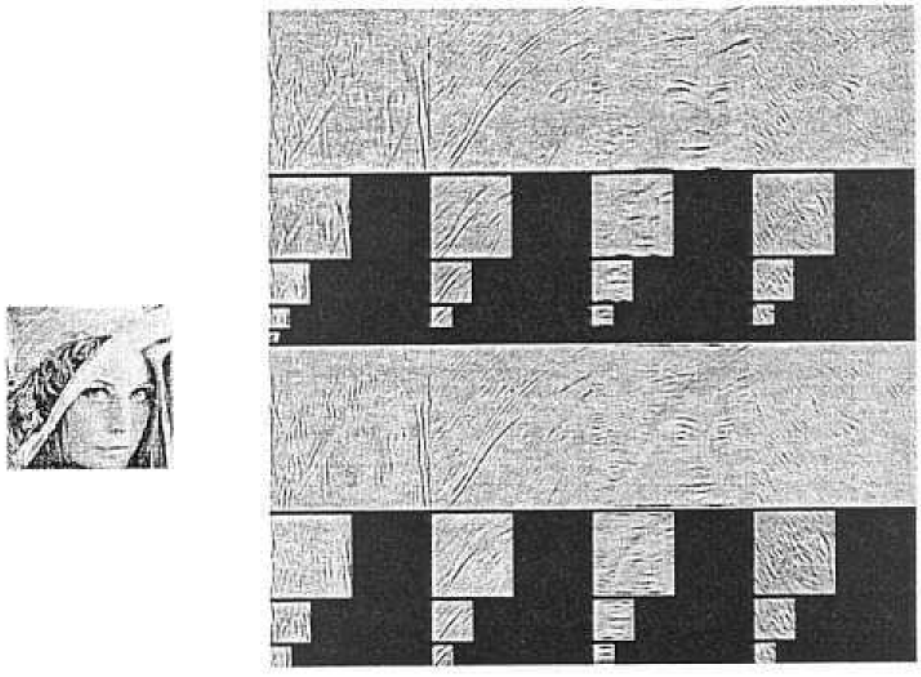


Figura 3.3: Transformada de Gabor de una imagen. El nivel de gris es proporcional al valor de cada coeficiente. Arriba se muestra la salida de los campos receptivos pares (detectores de barras) y abajo la de los impares (detectores de bordes). En medio, a la derecha, se puede apreciar el residuo de baja frecuencia

La Fig. 3.3 muestra la imagen original y la descomposión en canales resultante. En la parte superior está la descomposición obtenida con los canales pares (detectores de barras) y en la inferior la de los impares (que actúan como detectores de bordes). El residuo de baja frecuencia, apenas visible, se encuentra entre ambas. Cada nivel de frecuencia tiene la resolución adecuada a su anchura de banda. En este tipo de esquemas piramidales en vez de hablar de la respuesta en un punto dado sería mejor hablar de la respuesta en la vecindad de un punto, dado que la información de baja frecuencia es compartida por muchos píxeles. La imagen está pues codificada en una representación conjunta en ambos dominios, donde ni la posición ni la frecuencia espacial están definidas sin incertidumbre. A causa de esto, una visualización más realista sería asignar a cada bloque 16×16 de la imagen original una pirámide de coeficientes. En la base de dicha pirámide los coeficientes correspondientes a la frecuencia más alta estarían asignados a píxeles individuales, mientras que al ir disminuyendo la frecuencia, los coeficientes serían

compartidos por píxeles vecinos, hasta llegar al único coeficiente del residuo paso-bajo que es común a los 256 píxeles. Sin embargo, sobre todo al usar este esquema para análisis de texturas (Cap. 4 y 5), hablaremos a veces de la respuesta en un punto dado, o de las propiedades de una textura en un punto; en dichos casos se debe entender que tal respuesta se ha extraido a partir de un entorno local del punto considerado.

3.2 Inversión de la transformada

Como ya se indicó al presentar la transformación de Gabor (TG) (Cap. 2) hay una redundancia inherente al esquema, que hace que la imagen pueda recuperarse tanto de los coeficientes obtenidos a través de los filtros pares como de los impares. Como veremos en el capítulo siguiente el uso de ambos es importante cuando estamos interesados en el análisis, pero dado que aquí nos centraremos en la recuperación de la imagen, sólo usaremos el conjunto de coeficientes extraido con los filtros pares.

La principal objeción que se puede presentar a un esquema de Gabor es que, como se puso de manifiesto en la Fig. 2.2, su cubrimiento del espacio de Fourier no es completo. Esto implica que aún efectuando un adecuado muestreo para cada canal la reconstrucción de la imagen no será exacta. Este inconveniente ha tratado de ser resuelto con la implementación de algoritmos más sofisticados que proporcionan diversas variantes de la transformada de Gabor permitiendo una recuperación exacta de la imagen. Sin embargo, el inconveniente de estos algoritmos es que emplean procesos más complicados, como el uso de funciones biortogonales en el algoritmo generalizado de Gabor [52][57], o redes neuronales empleando operaciones de "feedback" [56]. Esto hace que dichas soluciones sean poco probables desde el punto de vista de una implementación que trate de simular lo que sucede en el sistema visual humano. Otros autores han propuesto esquemas como la transformación córtex (TC) [53][54](Ver Tabla 2.1), que son muy similares al esquema de Gabor. Esta transformación permite la recuperación exacta de la imagen pero con el uso de unos filtros que, aunque similares, son menos realistas que los de Gabor, y dif!ciles de implementar en el dominio espacial.

Sin embargo, nosotros creemos que el esquema de Gabor presentado previamente es una buena elección como base de un sistema de descomposición y reconstrucción de la imagen, a pesar de su no completitud. A continuación presentamos (3.2.1) el método para la inversión de la TG, así como una optimización para

reducir el error de reconstrucción. En la sección 3.2.2 consideramos las posibilidades de una codificación basada en nuestro esquema de Gabor, comparando los resultados de la reconstrucción con los de una tranformacion exacta como es la TC.

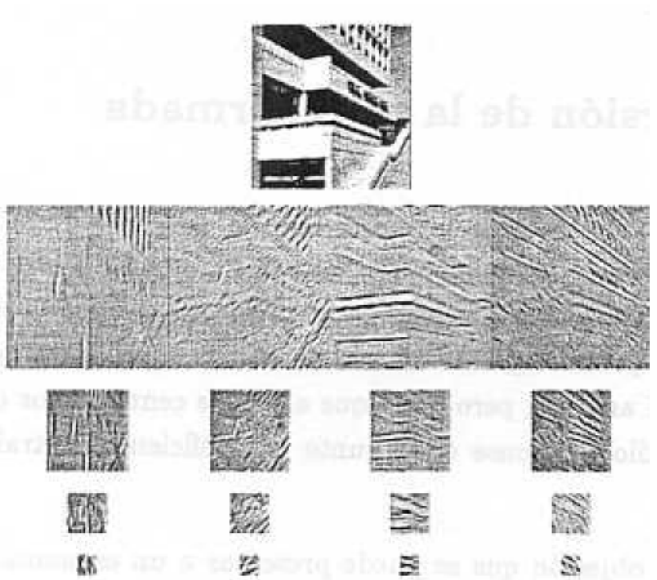


Figura 3.4: Descomposición de una imagen (arriba) en los 4×4 canales pares de la transformada de Gabor. Nótese cómo cada canal captura los detalles correspondientes a la frecuencia y orientación a la que está sintonizado

3.2.1 Reconstrucción de la imagen

Una vez generados los coeficientes pares de la transformación de Gabor ($R_{x_0,y_0,f_0,\theta_0,p=0}$) a partir de cualquiera de las dos implementaciones propuestas (convolución en dominio espacial o filtrado en el de Fourier) obtenemos una representación de la imagen que se muestra en la Fig. 3.4 (al contrario que en la Fig. 3.3 ahora sólo se muestran los canales pares). Cada canal captura los detalles de la imagen correspondientes a su frecuencia y orientación. Tenemos pues la imagen subdividida en 4 canales de orientación, 4 de frecuencia y un residuo final de baja frecuencia. El número de coeficientes empleados en la reconstrucción será la mitad del calculado anteriormente, debido a que sólo consideramos una paridad. Así pues, partiendo de una imagen de N^2 píxeles obtendremos un conjunto de

$4(N^2 + N^2/4 + N^2/16 + N^2/64)$ coeficientes, esto es, aproximadamente $16N^2/3$, algo más de 5 veces el número original de píxeles. Esta representación es por lo tanto redundante, si bien presenta otro tipo de ventajas.

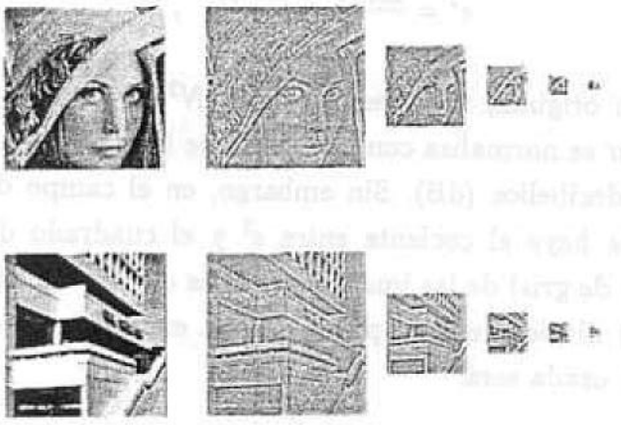


Figura 3.5: Suma de los 4 canales de orientación de cada frecuencia para las imágenes de las Fig. 3.3 y 3.4. El resultado es muy similar a la pirámide laplaciana [59]

Una de estas ventajas es que este esquema no precisa de una compleja transformada inversa. Para recuperar la imagen basta con ir sumando los distintos canales, ampliando mediante interpolación aquellos que sean de menor tamaño que el original. Para optimizar este proceso, primeramente se suman los cuatro canales de orientación para cada frecuencia. Se obtienen de esta forma 5 canales de frecuencia sin selectividad ante orientaciones, muy similares a los de la pirámide laplaciana [59]. Dichos canales se muestran en la Fig. 3.5 para las dos imágenes antes consideradas. Una vez efectuado dicho paso, se parte del canal de frecuencia más baja (residuo paso-bajo), se expande con una interpolación ideal en un factor 2 en ambas dimensiones y se suma al canal de frecuencia inmediatamente superior. Esto se repite para el canal así obtenido y se continúa el proceso hasta sumar el canal de mayor resolución, momento en el que recuperamos la imagen original. Sin embargo, la reconstrucción no es exacta, es decir, la imagen recuperada presenta errores con respecto a la original. La razón de estos errores es la no completitud de nuestro esquema, reflejada en el cubrimiento no uniforme del dominio de Fourier (Fig. 2.2). Esto es lógico si se piensa que el esquema de Gabor pretende ser un modelo sencillo del sistema visual, y este presenta también "errores" en la

reconstrucción (ilusiones visuales).

El criterio objetivo más común para cuantificar el parecido entre dos imágenes consiste en calcular el error cuadrático medio (e^2) [1] entre ambas:

$$e^2 = \frac{\sum_{i,j}(x_{ij} - \mathbf{x}_{ij})^2}{N^2} \quad , \tag{3.3}$$

donde x es imagen original, $\mathbf{x}$ la recuperada y N^2 el número de píxeles. Generalmente este error se normaliza con la energía de la señal original, expresandose dicho cociente en decibelios (dB). Sin embargo, en el campo de la codificación, tradicionalmente se haya el cociente entre e^2 y el cuadrado del rango máximo (número de niveles de gris) de las imágenes con las que tratamos. En nuestro caso el rango es de 256 niveles de gris, por lo que la expresión en dB de la relación señal/ruido (SNR) usada será:

$$SNR = -10 \log_{10} \frac{e^2}{256^2} \quad . \tag{3.4}$$

De esta forma una mayor relación señal/ruido nos indica una mayor fidelidad en la reconstrucción de la imagen. Sin embargo, aunque este parámetro cuantifica el grado de exactitud en la recuperación, nos dice bien poco sobre la forma de reducirlo (ya que al ser un indicador global no especifica donde se comete el error). Con la intención de minimizar dicho error calcularemos primeramente el equivalente de la función de tranferencia de modulación (MTF) para ambas implementaciones de la transformada de Gabor. En el caso de una tranformación exacta, dicha función debería ser la unidad para todas las frecuencias; esto, sumado al hecho de que podemos variar libremente las ganancias de cada canal antes de recuperar la imagen, nos sugiere un simple método de optimizar la reconstrucción. Bastaría con ver el aspecto de la función de tranferencia en ambas implementaciones y tratar de ecualizar las distintas frecuencias, asignando pesos diferentes a los distintos canales, hasta conseguir una función de transferencia que fuese lo más uniforme y próxima a la unidad posible. Para calcular dicha función de tranferencia hemos introducido como entrada al sistema un ruido blanco. Dado que el módulo de la transformada de Fourier de un ruido blanco es la unidad, la transformada de Fourier de la imagen recuperada será equivalente a la función de transferencia de modulación del sistema. La línea discontinua en las Figs. 3.6a y 3.6b muestra la función de transferencia en la implementación en el dominio espacial y de Fourier respectivamente. Son distintas puesto que en la implementación espacial se introducen errores adicionales debido al uso de aritmética entera, máscaras de

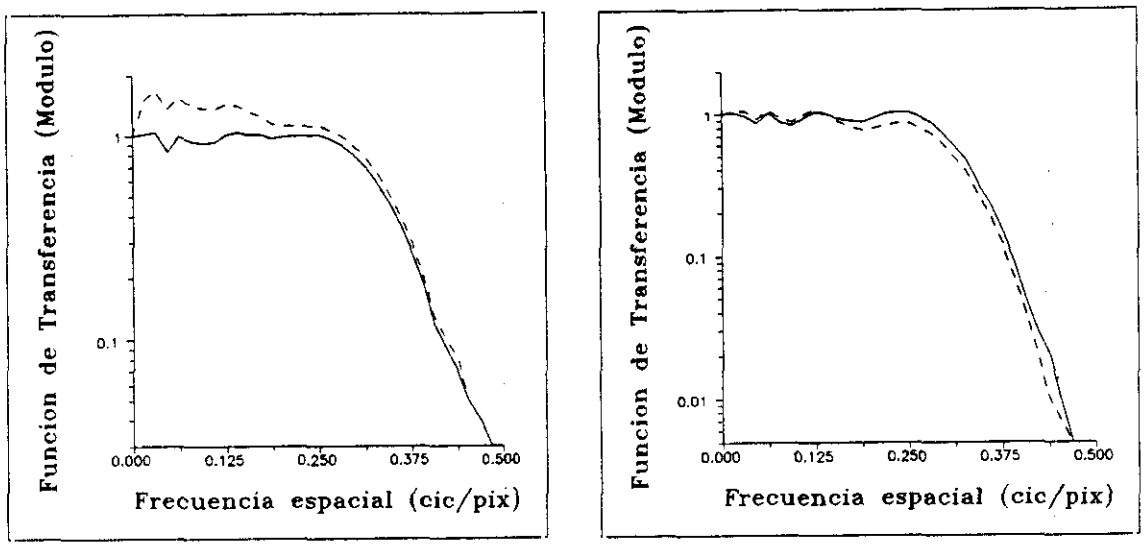


Figura 3.6: Funciones de transferencia de modulación para las implementaciones en el dominio espacial (izqda.) y de Fourier (derecha). La línea discontinua corresponde al caso de ganancias iguales para los distintos canales. La continua muestra el resultado de usar diferentes pesos (0.45, 0.60, 0.65, 0.90 en la implementaci"n espacial, 0.93, 0.93, 0.93, 1.15 en la de Fourier) para cada canal.

tamaño reducido, la reducción de la imagen en vez de aplicar filtros mayores, etc. En ambas figuras se muestran perfiles radiales, habiéndose promediado en orientaciones. Su alejamiento del valor unidad y sus oscilaciones ponen de manifiesto que la recuperación de la imagen no va a ser exacta.

Para encontrar las distintas ganancias con las que tratamos de ecualizar en lo posible dichas funciones se usó un método empírico. Como resultado de esta búsqueda se encontró que unas ganancias (para los cuatro canales de frecuencia en orden creciente) de 0.45, 0.60, 0.65 y 0.90 para la implementación espacial y de 0.93, 0.93, 0.93 y 1.15 para la de Fourier reducían el error de manera notable, principalmente en el caso de la implementación espacial. El residuo final paso-bajo no fué alterado. Las líneas contínuas en las Fig. 3.6a y 3.6b muestran las funciones de transferencia corregidas con dichas ganancias. Ahora se ve como tienden a la unidad, excepto por pequeñas indentaciones y sobre todo por la caída a frecuencias altas, ya que nuestro modelo no considera el residuo de altas frecuencias [53]

Sin embargo, mostramos a continuación que aunque estos criterios objetivos (aspecto de la función de transferencia, relación señal/ruido) den claras diferencias entre el original y la imagen recuperada, la calidad de la reconstrucción es buena bajo criterios subjetivos, siendo díficil distinguir visualmente el original de

(a) (b) (c)

Figura 3.7: Imagen original (a) y reconstrucciones a partir de las implementaciones de la transformada de Gabor en el dominio de Fourier (b) y espacial (c).

la imagen recuperada. En la Fig. 3.7a se muestra la imagen original, y en la 3.7b y 3.7c las recuperaciones a través de las implementaciones de Fourier y espacial respectivamente, con las ganancias antes citadas. En ambos casos la relación señal/ruido (Ec. (3.4) de las reconstrucciones es de 26 dB. La mayor parte del error de reconstrucción es debido al residuo de altas frecuencias que no se considera. Esto se comprueba considerando la recuperación a partir de una tranformación córtex despreciando tal residuo (ver apartado siguiente). En dicho caso (donde el error es sólo debido a la pérdida de altas frecuencias, ya que la transformación es exacta) la relación señal/ruido sólo aumenta a 27 dB. Podríamos considerar dicho residuo en nuestra transformación, pero su valor va a ser generalmente despreciable. Dicho residuo puede llegar a ser nulo en casos reales de imágenes obtenidas a través de sistemas ópticos (con el consiguiente corte en frecuencias) si se hace un muestreo adecuado.

3.3 Comparación con la transformada córtex (TC)

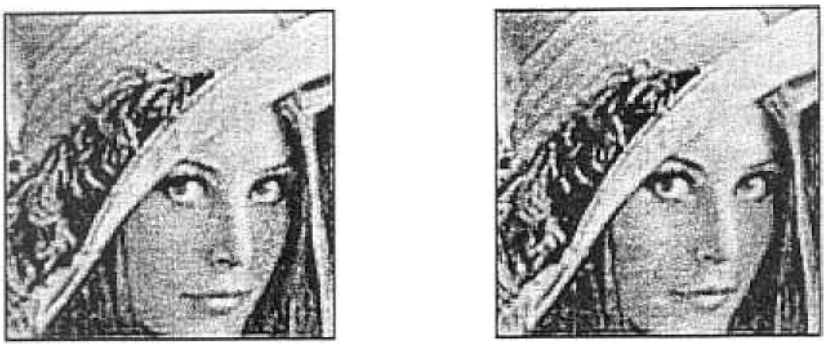
La codificación de imágenes es un aspecto cada vez más fundamental en numerosas aplicaciones de comunicaciones. El manejo de cantidades cada vez mayores de información, así como la alta dimensionalidad de las imágenes, hacen que sea imprescindible reducir la cantidad de información a transmitir. Considerando el típico caso de una emisión de TV digital donde estan involucrados del orden de 500000 bytes por plano de color; con unas 30 imagenes por segundo, el flujo de información a mantener es del orden de 50 Mbytes/seg. Cualquier proceso que permitiera reducir en un factor significativo tal cantidad supondría un considerable

ahorro [79].

En general, la mayoría de los métodos propuestos (wavelets [60][61], sub-bandas [62][63][64], filtros de espejo en cuadratura -QMFs- [65], etc.) están diseñados para constituir un código compacto, poco redundante, de forma que los nuevos coeficientes se calculan en una base ortogonal, no aumentándose la dimensionalidad de la imagen original. Otro enfoque diferente a estos es el uso de esquemas basados en el sistema visual, como la transformada cortex [54], o el esquema de Gabor aquí propuesto. Estos modelos consisten en la descomposición de la señal en distintos canales de frecuencia y orientación, lo que les hace presentar un cierto grado de redundancia, similar a la que parece existir en el sistema visual, lo que les pone en desventaja para aplicaciones de codificación. Sin embargo, considerando que una codificación ideal (admitiendo pérdidas) es aquella que sólo conserva la información a la cual el observador humano es sensible, la ventajas de este tipo de esquemas son obvias. Si la transformación modela de alguna forma lo que sucede en el sistema visual estaremos en una situación óptima para dar mayor o menor importancia (o en términos de codificación, mayor o menor número de bits) a un canal o a otro). Watson demostró [54] que esquemas redundantes como la TC pueden ser la base de un esquema de codificación competitivo con otros no redundantes. Sin embargo, a pesar de sus similitudes con la TC, la tranformación de Gabor ha sido tradicionalmente desdeñada para aplicaciones de codificación debido a su no exactitud. Sin pretender desarrollar un esquema de codificación completo compararemos algunos aspectos de ambas transformadas y expondremos las razones por las que creemos que la transformada de Gabor es también apropiada como base de un esquema de codificación.

3.3.1 Comparación bajo criterios objetivos y subjetivos

Para la comparación hemos usado una implementación de la TC [54] con el mismo número de canales (4×4) que el esquema de Gabor e igual posición de estos en el espacio de Fourier. La anchura a media altura de los canales en ambos casos es también igual, aunque por su naturaleza, los filtros de la TC presentan una caída más brusca que nuestras gaussianas. Debido a la similitud entre los resultados obtenidos a partir de las dos implementaciones de la transformada de Gabor (TG), a partir de ahora sólo se presentarán los de la implementación espacial.



(a) (b)

Figura 3.8: Comparación entre las reconstrucciones obtenidas a partir de la transformada córtex (a) y la de Gabor (b). En ambos casos el residuo de alta frecuencia no se ha utilizado en la recuperación.

En la Fig. 3.8 se comparan las dos recuperaciones, habiendose despreciado en ambas el residuo de alta frecuencia. La relación señal/ruido es ligeramente mejor para la TC, 27.6 dB frente a los 26.4 dB de la TG, pero la apariencia visual es casi idéntica. Dado que el receptor final de las imágenes va a ser el ojo humano, creemos que no sólo han de considerarse criterios objetivos (que por otra parte dan sólo pequeñas diferencias). En este sentido, en la Fig. 3.8 la mayoría de los observadores encuentran difícil escoger entre ambas reconstrucciones. Este resultado indica que aunque las funciones de Gabor han sido comunmente descartadas por su falta de completitud, el esquema aquí propuesto es "cuasicompleto". Esto es, los efectos de la falta de exactitud son pequeños medidos con un criterio objetivo e irrelevantes ante criterios subjetivos de percepción visual.

Otro aspecto importante cuando se trata de llevar a cabo una codificación es minimizar la cantidad de información necesaria para reconstruir la imagen. Dos factores son fundamentales a la hora de determinar la información necesaria: el número de coeficientes empleados y cuántos bits (unidades de información) son necesarios para codificar cada coeficiente. La cantidad de información total (en bits) será pues el producto de estos dos factores. El número mínimo de bits necesarios para codificar un conjunto de coeficientes una vez cuantificados en n niveles (admitiendo un cierto error)viene dado su entropía E, definida [2] como:

$$E = -\sum_{i=0}^{n-1} p(i) \ln(p(i)) \ , \tag{3.5}$$

donde $p(i)$ es la probabilidad de que un coeficiente haya sido cuantificado dentro del

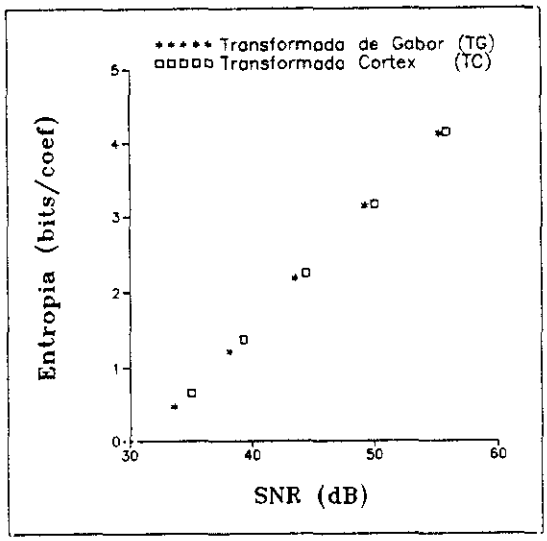


Figura 3.9: Número mínimo de bits necesarios para codificar el conjunto de coeficientes del nivel de máxima frecuencia, frente al error cometido en la cuantificación (en dB), para ambas transformadas (TG y TC)

nivel i. Por otra parte, el número de coeficientes empleados para un determinado canal es proporcional a su ancho de banda (un ancho de banda pequeño posibilita un muestreo más espaciado, y por lo tanto un menor número de coeficientes). Compararemos pues ambas transformadas teniendo en cuenta estos dos factores.

En la Fig. 3.9 se representa el número de bits necesarios para codificar los coeficientes del nivel de resolución máximo frente al error cometido al hacer dicha cuantificación (en dB) para ambas transformaciones (usando un cuantificador ideal). Las gráficas para otros niveles de frecuencia muestran un comportamiento similar. Se observa que para un mismo grado de error, la entropía es muy similar en ambos casos, por lo que ambas transformadas parecen equivalentes en esta primera comparación.

Las anchuras de banda de los canales se han comparado de la forma siguiente. Primeramente se calculó el área que en el espacio de Fourier encerraba el 99% de la energía de un canal de la TC. A continuación se comprobó qué porcentaje de la energía de un canal similar de Gabor estaba dentro de ese área, obteniendose un resultado de un 90%. Aunque la diferencia no es muy grande, sí que implica una mayor redundancia en el caso de la TG, que hace que necesite la transmisión de un número ligeramente superior de coeficientes. Sin embargo, como veremos en la sección siguiente esta redundancia incrementa la robustez de la codificación.

Por otra parte, la reciprocidad de las funciones de Gabor bajo transformaciones de Fourier las hace especialmente apropiadas para ser usadas como base de representaciones conjuntas. Finalmente, la caída suave de los filtros de Gabor (causa de su mayor anchura de banda) disminuye considerablemente los efectos de "ringing" en los casos de reconstrucciones con pérdidas. Como se verá a continuación esta propiedad hace que constituyan un código robusto.

3.3.2 Comparación en términos de robustez de la codificación

En el apartado anterior comparabamos ambas transformaciones en cuanto a la eficiencia, es decir, la capacidad de minimizar la cantidad de información necesaria. Sin embargo otra característica fundamental de una codificación es su robustez, entendiéndose por tal la capacidad de presentar una buena reconstrucción aun con fallos en la transmisión de información, es decir, con pérdidas parciales (una situación común en muchas aplicaciones). Como se aprecia en la Fig. 3.10 los métodos de codificación piramidal son robustos ante pérdidas de canales de alta frecuencia. La fila superior presenta reconstrucciones a partir de la TC, en las que vamos eliminando progresivamente canales de alta frecuencia: con sólo tres canales (izquierda, SNR = 23.0 dB), con dos (centro, SNR = 20.2 dB), y finalmente sólo con un canal y el residuo de baja frecuencia (derecha, SNR = 17.8 dB). En la fila inferior se muestran los mismos resultados a partir de una TG (SNR = 21.8, 19.3 y 17.4 dB, respectivamente). A pesar de la baja relación señal ruido que presentan las recuperaciones a partir de pocos canales, esta figura ilustra la robustez de este tipo de representaciones. Nuestro sistema visual puede hacer un buen trabajo de reconocimiento de la cara incluso cuando sólo dos canales están disponibles (lo que comprende unicamente un 6% del número total de "logons" o coeficientes. Piénsese cual sería el aspecto de la imagen si un porcentaje similar de información hubiese sido eliminada en la representación espacial -píxeles- de la imagen). La Fig. 3.10 ilustra también otro punto muy importante: en el caso de pérdidas de información, ninguna codificación es exacta, por lo que la robustez de la codificación no depende de su completitud inicial. Comparando las dos filas de la Fig. 3.10 podemos observar algunos efectos de "ringing" cerca de los bordes, que son más aparentes en la fila superior. De nuevo, aunque la relación señal ruido (criterio subjetivo) es mejor para la TC (27.6, 23.0, 20.2 y 17.8 dB frente a 26.4, 21.8, 19.3 y 17.4 dB), un criterio visual puede preferir la TG cuando ha habido pérdidas.

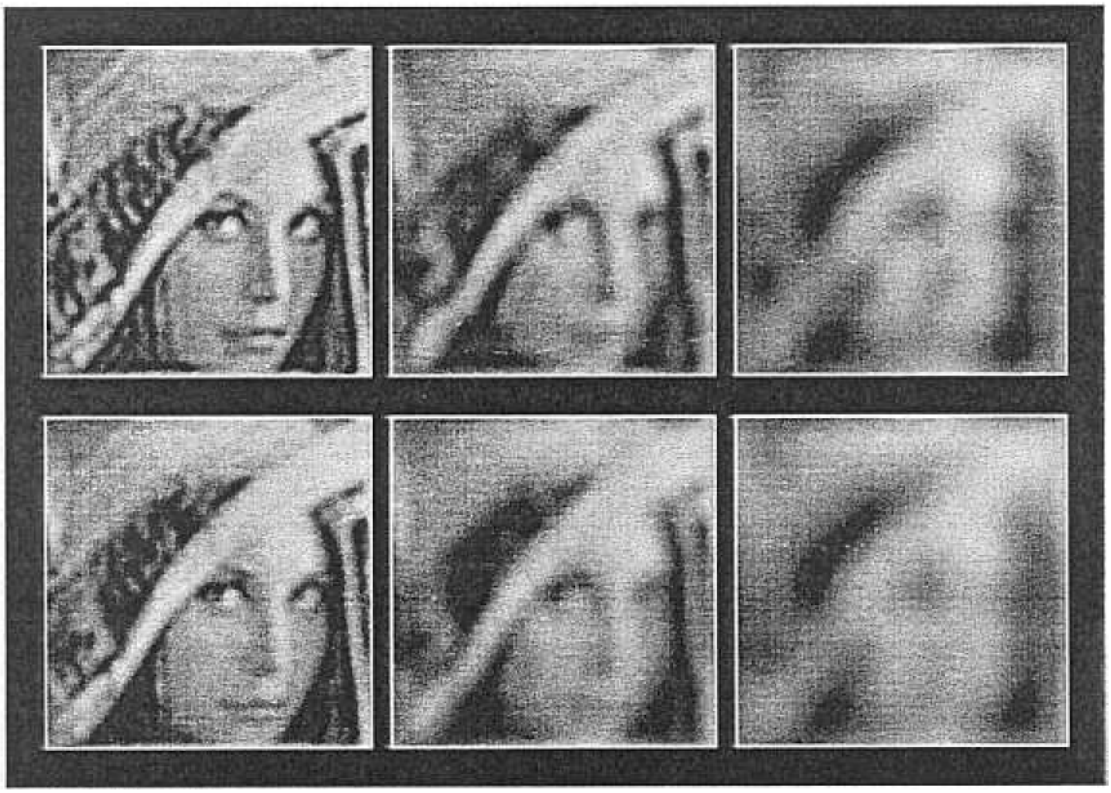


Figura 3.10: Reconstrucción de una imagen con pérdidas parciales de canales de frecuencia a partir de la TC (arriba) y la TG (abajo). En la primera columna se muestra la reconstrucción con sólo tres canales, dos en la segunda y uno en la tercera.

Este hecho se vuelve a poner de manifiesto en la Fig. 3.11, donde tenemos diversos casos de omisión de canales de orientación. La primera columna (izquierda) corresponde a la eliminación de un sólo canal de orientación (45°) en el canal correspondiente a la segunda frecuencia más alta; la segunda y la tercera corresponden a pérdidas de un canal de orientación (0° y 135° respectivamente) para todas las frecuencias. La fila superior corresponde a la TC y la inferior a la TG. Se ve en estos ejemplos que la TC sufre en general más degradación (ver columna derecha), siendo la TG más robusta. Esto se debe al hecho de que cuando se pierde un canal en la TC, al haber menos solapamiento entre canales, queda un hueco abrupto en el espacio de Fourier. Por el contrario, en la TG este hueco es parcialmente llenado por la información aportada por el solapamiento de los canales vecinos.

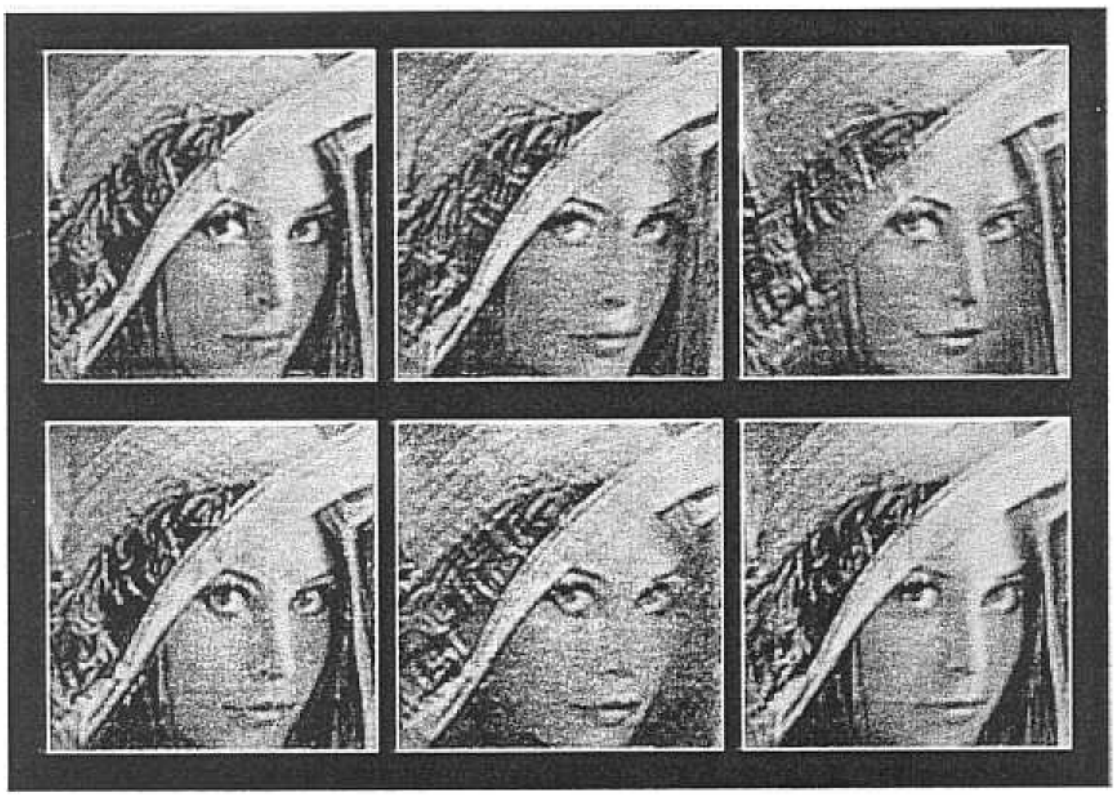


Figura 3.11: Efectos de pérdidas de canales de orientación en la TC(arriba) y TG (abajo). En la primera columna se ha perdido la orientación de 45° sólo en una frecuencia f_2. En la segunda y tercera, una orientación (0° y 135°, respectivamente) se ha eliminado en todas las frecuencias.

Esto hace que en el caso de la TG las posibles pérdidas de información sean más facilmente recuperables (por ejemplo con filtros inversos de Wiener o similares) que en el caso de la TC.

En conclusión, creemos que la transformada de Gabor presenta una serie de ventajas con respecto a otras, mientras que en contra tiene que su inversión no es exacta. Frente a tal crítica, hemos mostrado que con una adecuada elección de las funciones de Gabor y una minimización de los errores en la reconstrucción, los efectos de su falta de completitud son pequeños y perceptualmente casi inapreciables. La comparación con la transformada córtex muestra las posibilidades de un esquema de Gabor como base de aplicaciones de codificación.

Parte II

Capítulo 4

Segmentación y clasificación de texturas

La textura es una de las más importantes e intuitivas propiedades de las imágenes; sin embargo, es también una de las más difíciles de definir. De hecho, incluso la palabra textura se usa aquí fuera de su contexto original, que refiere a una sensación táctil. Se han hecho muchos intentos, la mayoría sin éxito, para caracterizar completamente las texturas a través de un conjunto único y reducido de parámetros. Uno de estos intentos fué la famosa conjetura de Julesz [80], que proponía que dos texturas que compartieran sus estadísticas de primer y segundo orden serían indistinguibles para un observador humano. Esto hubiera reducido el análisis de la textura al estudio de dichos parámetros estadísticos, pero desafortunadamente se han presentado varios contraejemplos [97].

Dado que la propia definición de la textura es todavía un problema abierto, se han propuesto una gran variedad de métodos diferentes para estudiar las prodiedades de las imágenes en cuanto a su textura. Estos técnicas se pueden agrupar en dos clases principales [81]. Una comprende los métodos que usan un enfoque estadístico, consistente en la definición de un conjunto de parámetros (llamados características o descriptores) que se extraen de la textura. Por el contrario, el enfoque estructural considera a las texturas como repeticiones de un patrón básico de acuerdo con ciertas reglas (que componen lo que se denomina una gramática).

Los métodos estadísticos se pueden clasificar a su vez teniendo en cuenta el tipo de descriptores que usan. Muchos de ellos están basados en el estudio de parámetros estadísticos de primer (niveles de gris [82], camino recorrido [98], diferencias de niveles de gris [99], etc) o segundo (matrices de co-ocurrencia [83]) orden. Por otra parte, algunas de las características intuitivas de las texturas, como el

grano o la direccionalidad, pueden estudiarse fácilmente en el dominio de frecuencias espaciales. Por esta razón se han propuesto diversos métodos basados en descriptores espectrales. Algunos de ellos usan descriptores globales [100], pero son mayoría los que se basan en descriptores locales. Los métodos basados en el análisis local de Fourier tales como la distribución de Wigner [47][48] o las funciones de Gabor [84][87][85][86], pertenecen a esta última clase, y usan descriptores que combinan información tanto del dominio espacial como del de Fourier. Un estudio útil comparando algunos de estos métodos se encuentra en [99]. Recientemente, se han propuesto también descriptores de texturas basados en la posible geometría fractal de las imágenes [89]–[91].

La existencia de tantos y tan diversos métodos indica que ninguno de ellos constituye en realidad una solución definitiva, y que por lo tanto, el problema está aún sin resolver. Por otra parte, el sistema visual humano ofrece un alto rendimiento (superior a cualquiera de los métodos citados) en el reconocimiento y discriminación de texturas. Consecuentemente, se podría pensar que un método basado en un modelo del sistema visual daría buenos resultados. Dado que las funciones de Gabor son un buen modelo de los campos receptivos de células corticales [39][41], no es de extrañar que hayan sido propuestas recientemente por varios autores [84]–[87] para el análisis de texturas.

En este Capítulo aplicamos el sencillo modelo del sistema visual descrito en los capítulos precedentes para este propósito. En la sección 4.1 describimos el conjunto de descriptores que vamos a emplear para caracterizar las texturas. A continuación se estudian las prestaciones de dichos descriptores cuando se aplican a diversas tareas de análisis de imágenes, como son la segmentación (4.2) y la clasificación (4.3).

4.1 Matriz de descriptores

Un problema común en el análisis de texturas es encontrar un conjunto reducido de descriptores capaz de discriminar entre diversas texturas. Esta extracción de descriptores es generalmente necesaria, puesto que la mayoría de los métodos propuestos manejan una gran cantidad de información que necesita ser reducida para poder ser tratada más fácilmente. Por ejemplo, la salida de los métodos estadísticos de segundo orden son matrices de co-ocurrencia [83], cuyo tamaño típico es 256×256. Métodos basados en la distribución de Wigner [47] suelen dar lugar a matrices 32×32 ó 64×64 por cada punto analizado. Por ello, estos métodos

requieren un procesado previo para obtener finalmente un número reducido de descriptores, del orden de 10 o menos [83]. Algunos métodos basados en funciones de Gabor también recurren a esta extracción de descriptores. Una consecuencia de esto es que tal proceso no es reversible, es decir, es útil para aplicaciones de análisis, pero las texturas no pueden ser recuperadas a partir de esos descriptores, ya que con la reducción del número de descriptores se ha eliminado parte de la información.

La idea clave de nuestro enfoque es no usar un conjunto de parámetros especialmente definidos para el análisis de texturas y escogidos precisamente por su rendimiento en unas tareas específicas, sino todo lo contrario. Lo que aquí proponemos es el uso de un modelo simplificado del sistema visual (Capítulo 2,[53]) como entorno de propósito general. Así, el análisis de texturas sería sólo una aplicación concreta del modelo. En este capítulo usaremos la salida del esquema de Gabor propuesto en los capítulos anteriores ($R_{x_0,y_0,f_0,\theta_0,p}$) directamente como conjunto de descriptores de la textura.

Como se mostró en el Capítulo 3, la salida de nuestro modelo es algo similar a una pirámide de coeficientes asociada a un área local de la imagen, en la que las respuestas de canales de baja frecuencia son compartidas por una serie de píxeles. Sin embargo, en este tipo de aplicaciones de análisis lo que se busca es asociar a cada punto de la imagen original un conjunto de descriptores. Lo que haremos es asignar a cada píxel (x, y) una matriz compleja 4×4, $M_{x,y}$, cuyos coeficientes reales e imaginarios serán las salidas de los 4×4 canales de frecuencia y orientación de paridad par e impar respectivamente:

$$M_{x,y}(f, \theta) = R_{x,y,f,\theta,p=0} + iR_{x,y,f,\theta,p=1} \ . \tag{4.1}$$

Naturalmente, esto implica que los componentes de nuestra matriz de descriptores para frecuencias bajas serán compartidos por píxeles vecinos. Tal matriz es suficientemente pequeña como para ser utilizada como conjunto de descriptores, constituyendo a su vez, como ya se apuntó, una codificación especial de la imagen, por lo que sus usos no están limitados al análisis. Además, algunos campos receptivos de células simples del cortex presentan perfiles similares a las funciones de Gabor [39][41], lo que sugiere que el sistema visual podría usar un tipo de información similar para efectuar su propio análisis de texturas.

Como ya se indicó en los capítulos precedentes, hay una redundancia de información en nuestro esquema, dado que la imagen original puede ser recuperada a partir sólo de los canales de simetría par (de hecho, la salida de los canales

pares e impares está relacionada a través de una transformada de Hilbert [92]). Sin embargo, hemos constatado que el uso de ambos tipos de canales mejora notablemente los resultados en el caso del análisis. Para evitar redundancias y al mismo tiempo maximizar las capacidades de discriminación entre texturas usaremos sólo el módulo de los elementos complejos de la matriz de la Ec. (4.1). Esta simplificación en el número de descriptores se llevó a cabo tras comparar los resultados obtenidos con la matriz de los módulos y los de la matriz compleja en una tarea de clasificación. El resultado fué que las mejoras en la discriminación usando ambas componentes fueron menores del 10% (ver B-distancias en Tabla 4.2), no siendo pues interesante duplicar el número de descriptores para mejorar tan poco los resultados. Por el contrario, en la misma Tabla 4.2 se aprecia que sí que es importante el uso del módulo de ambas componentes en vez de sólo una de ellas. Esto se debe a que calcular el módulo equivale a una demodulación de la salida del canal. Considerese el caso de una textura con una alta periodicidad, cuya frecuencia corresponda a uno de los canales de nuestro modelo. La salida de dicho canal (es decir, los valores de una de las componentes de la matriz de descriptores) consistirá en algo muy parecido a la textura original, una sucesión de valores altos y bajos. Si se intenta clasificar la textura en base a dicho descriptor obtendremos un mal resultado, ya que valores distintos del descriptor corresponden a la misma textura. Para eliminar este problema deberíamos demodular la salida del canal, centrandolo en la frecuencia cero. Un resultado equivalente (y más sencillo de obtener) consiste en calcular el módulo, ya que $|H \exp(-i\mathbf{f})|$ (módulo del canal tras ser demodulado) es igual a $|H|$ (módulo del canal original).

El uso del módulo viene también refrendado por el hecho de que se han propuesto ejemplos [101] que sugieren que en el sistema visual se lleva a cabo un proceso no lineal, operando sobre la respuesta de las células simples del córtex. Algunos autores [85][86] proponen un cálculo de la energía, enteramente análogo a lo que nosotros hacemos. Sin embargo, recientemente se ha observado [101] que la extracción del módulo no es probablemente lo que sucede en el sistema visual, proponiendose algunos otros mecanismos no lineales. A pesar de esto, tales consideraciones no afectarían a la bondad general de nuestro enfoque, consistente en el uso directo de las salidas de un modelo de sistema visual para tareas de análisis de texturas.

Como ya se comentó en el Capítulo 3, nuestro esquema no toma en cuenta la frecuencia cero. Para la recuperación de la imagen se implementó un canal adicional paso bajo. En nuestras aplicaciones de análisis hemos preferido incorporar dicha información de baja frecuencia como una normalización de los distintos

elementos de nuestra matriz de descriptores. Para ello, se dividió la respuesta de cada célula de Gabor por la respuesta de un filtro de tamaño similar, pero centrado en la frecuencia cero. Esta normalización hace que nuestra matriz de descriptores sea invariante ante un factor en la iluminación. Siguiendo la analogía con el sistema visual este proceso de normalización daría cuenta de los mecanismos de adaptación. Sin embargo, en las tareas que se presentan en éste y el siguiente capítulo, hemos comprobado que no hay apreciables diferencias entre usar o no estos canales normalizados. Probablemente la normalización no es esencial, dado que la digitalización de las imágenes precisa de un proceso previo de adaptación a la iluminación: usamos una cámara con un control de ganancia automática, diafragma ajustable y similares condiciones de iluminación. De esta forma, las imágenes así obtenidas están casi óptimamente cuantificadas en 256 niveles (un bajo rango dinámico frente al encontrado en la vida real), por lo que es innecesario un control digital posterior. En lo que sigue, se han usado los canales normalizados, aunque como hemos dicho, no hay apenas diferencias con respecto a los originales.

La Fig. 4.1 muestra una comparación entre las matrices de descriptores medias de las texturas paja y mar, del album de Brodatz [102] (ver Fig. 4.2a). Las filas representan canales de frecuencia, desde $\frac{1}{4}$ ciclos/píxel en la superior hasta $\frac{1}{32}$ ciclos/píxel en la inferior. Las columnas corresponden a las diferentes orientaciones (0°,45°,90°, y 135°). Podemos ver una clara relación entre las propiedades de las texturas y sus matrices de descriptores. La matriz de la textura paja tiene mayores valores que la del mar, debido a un mayor contenido de energía en frecuencias medias y altas, lo que se traduce en respuestas mayores a los filtros de Gabor.

$$\begin{pmatrix} 1.11 & 0.28 & 0.34 & 0.40 \\ 2.26 & 0.52 & 0.44 & 0.82 \\ 3.12 & 0.77 & 0.56 & 1.10 \\ 2.47 & 0.75 & 0.54 & 0.94 \end{pmatrix} \qquad \begin{pmatrix} 0.12 & 0.15 & 0.41 & 0.19 \\ 0.14 & 0.35 & 0.62 & 0.25 \\ 0.17 & 0.23 & 1.02 & 0.31 \\ 0.15 & 0.24 & 1.22 & 0.41 \end{pmatrix}$$

PAJA MAR

Figura 4.1: Matrices de descriptores características de las texturas de Brodazt paja (a) y mar (b), obtenidas a partir de las texturas mostradas en la Fig. 4.2a

Asímismo, los valores más altos en la textura paja ocurren en la primera columna (orientación horizontal) reflejando la alta direcionalidad de esta textura. El máximo aparece en la tercera fila (frecuencia de 1/16 ciclos/píxel), lo que indica

una distancia media entre pajas de unos 16 píxeles. El mar presenta valores más bajos al tener una estructura espacial más suave. Los valores máximos aparecen en la tercera columna (vertical), debido a la dirección de las olas.

Estos simples ejemplos muestran cómo nuestra matriz de descriptores refleja las características de la textura, por lo que cabe esperar una buena discriminación al usar los métodos apropiados. En las siguientes secciones se estudian las posibilidades del esquema de Gabor en tareas de segmentación y clasificación de texturas. Para estas tareas hemos usados métodos convencionales bien conocidos, ya que nuestra intención no es la de diseñar nuevos algoritmos al respecto, sino probar la utilidad de un esquema de propósito general para el problema particular del análisis de texturas.

4.2 Segmentación

Una importante tarea en el análisis de imágenes, previa a procesados de nivel superior, es la segmentación. Consiste en la división de la imagen en varias regiones de características comunes, de acuerdo a ciertas consideraciones previas. Los métodos más usuales se basan en técnicas y algoritmos de "clasificación no supervisada" [103], mediante la cual la imagen se subdivide en regiones homogeneas con respecto a características prefijadas. En nuestro caso, las características escogidas son los elementos de la matriz de descriptores definida en la sección anterior.

Entre las diversas técnicas que se emplean para este tipo de segmentación, hemos escogido el algoritmo denominado de K medias (K-means) [103][90], debido a su simplicidad. Este algoritmo se basa en la minimización de la suma de la distancia entre los puntos pertenecientes a una clase y el centro de dicha clase. El centro de una clase es el valor medio de los puntos que pertenecen a ella. Al aplicar este algoritmo, los puntos de entrada se agrupan en clases homogéneas. Una descripción más completa del proceso puede encontrarse en las referencias antes citadas. Hay varias formas de mejorar los algoritmos de K medias : permitir que las clases se unan y separen durante el proceso, estudiar más cuidadosamente el tipo de métrica a usar dependiendo de la aplicación, etc [103]. Sin embargo, ninguna de estas posibilidades han sido usadas en nuestro caso. Nuestra intención no era tanto diseñar un algoritmo optimizado, y por lo tanto más complejo, sino mostrar las posibilidades de los descriptores de Gabor en tareas de discriminación entre texturas. Consecuentemente, hemos usado una implementación sencilla, lo que permite suponer que empleando un algoritmo de segmentación más sofisticado

sería posible mejorar los resultados. Se ha empleado una distancia euclídea en el espacio de 16 dimensiones de la matriz de descriptores. Así pues, dados dos puntos de la imagen, caracterizados por sus matrices de descriptores M_1 y M_2 (Ec. (4.1), la distancia $d_{(1,2)}$ entre ellos usada por el algoritmo será:

$$d_{(1,2)} = (\sum_{f,\theta}(M_1(f,\theta) - M_2(f,\theta))^2)^{\frac{1}{2}}. \tag{4.2}$$

Los pasos seguidos para probar la capacidad del esquema de Gabor en la segmentación de texturas han sido los siguientes:

1. Se compusieron diversas imágenes de tamaño 128 × 128 píxeles combinando 2 ó 4 texturas naturales extraidas del album de Brodatz [102]. La Fig. 4.2a contiene dos texturas (mar [D37] y paja [D15]) y la Fig. 4.3a cuatro texturas (mar, paja, tela de algodón[D77] y arena [D29]).

2. Aplicamos el esquema de Gabor a cada imagen y calculamos la matriz de descriptores en un subconjunto de puntos. Concretamente, se obtuvieron muestras espaciadas 8 píxeles en ambas direcciones, por lo que al ser las imágenes de tamaño 128×128, se extrajeron un total de 16×16 = 256 matrices de descriptores por cada imagen. La razón de este submuestreo fué únicamente la reducción del coste computacional.

3. Los descriptores así calculados constituyen el conjunto de puntos que se suministra al algoritmo de segmentación. Como ya se ha comentado, se usó una distancia euclídea, pero previamente se normalizaron los 16 componentes de la matriz de forma independiente entre 0 y 1. De esta forma todos los componentes tienen el mismo peso al calcular la distancia entre diversos puntos en el espacio de los descriptores.

4. Finalmente, para eliminar los puntos aislados (que probablemente están mal clasificados), al resultado del algoritmo de K medias se le aplica un filtro de moda. La salida de este filtro es el valor más repetido en la vecindad del punto de aplicación, que en este caso comprendía 3 × 3 píxeles. Este filtro elimina píxeles aislados que están rodeados de otros pertenecientes a una clase diferente. Es similar a un filtro de mediana, aunque más rápido, y creemos que más adecuado en un caso como éste, en el que hay muy pocos niveles o clases.

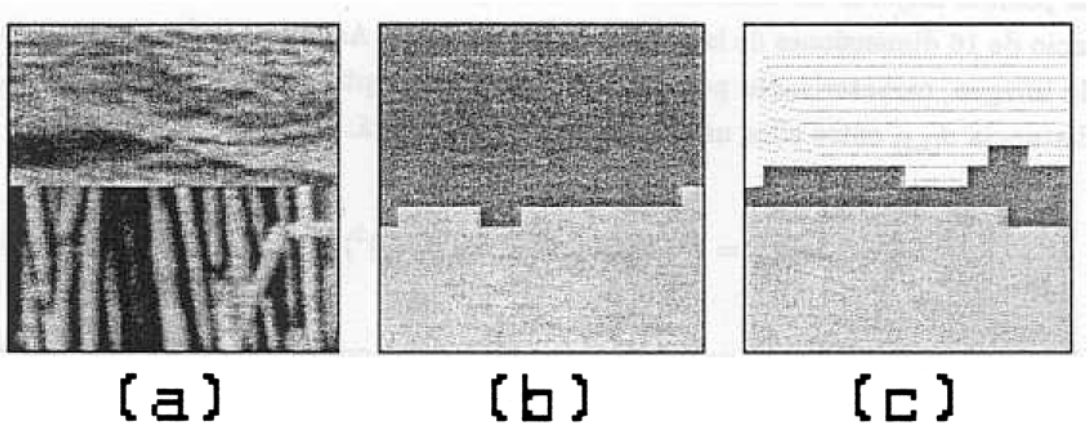


Figura 4.2: (a) Imagen conteniendo 2 texturas: mar (arriba) y paja (abajo). (b) Resultados de una segmentación con un algoritmo de K-medias suponiendo 2 regiones. (c) Lo mismo, pero ahora presuponiendo tres regiones en la imagen.

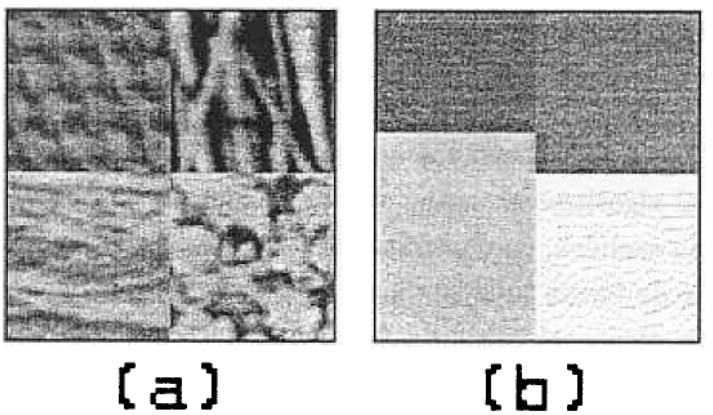


Figura 4.3: (a) Imagen conteniendo 4 texturas (de arriba abajo y de izquierda a derecha, tela de algodón, paja, mar y arena). (b) Resultados de una segmentación suponiendo 4 regiones.

Las Figs. 4.2b y 4.3b muestran los resultados obtenidos cuando el proceso anterior se ha aplicado a las imágenes de las Figs. 4.2a y 4.3a respectivamente. En estas imágenes, los niveles de gris no tienen un significado intrínseco. Queremos decir con esto que los píxeles con el mismo nivel de gris son los que han sido agrupados en una misma clase por el algoritmo, no habiendo por tanto ninguna relación entre píxeles con el mismo gris en las dos imágenes resultantes. Las imágenes de la segmentación presentan una apariencia de baja resolución, ya que sólamente se usaron 16 × 16 puntos por cada imagen. Los resultados son muy satisfactorios, ya que la segmentación llevada a cabo por el ordenador corresponde a la distribución de las diferentes texturas en las imágenes originales.

Cabe señalar que los errores aparecen únicamente a lo largo de las fronteras entre texturas, donde algunos puntos han sido adjudicados erroneamente a la clase contigua. Este problema de los límites entre diversas texturas en una imagen aparecerá y será discutido también en la siguiente Sección. La razón de esta anomalía es que cuando los canales de baja frecuencia se aplican cerca de las fronteras "ven" simultáneamente ambas texturas, debido a la gran extensión de su campo receptivo. Desde el punto de vista de estos canales de baja frecuencia, la frontera constituye una textura diferente, mezcla de las dos, lo que induce estos errores. No obstante, esta desventaja puede convertirse en todo lo contrario, sin más que permitir al algoritmo de segmentación que considere a las fronteras como clases diferentes. De esta forma se pueden detectar bordes entre texturas tal y como se ilustra en la Fig. 4.2c. La Fig. 4.2c muestra los resultados del proceso de segmentación de la Fig. 4.2a, pero esta vez, suponiendo que teníamos tres clases en vez de dos (Fig. 4.2b). Aparece claramente una tercera región correspondiente a la frontera entre las dos texturas.

4.3 Clasificación

La clasificación aparece en un nivel superior en el análisis de imágenes. Mientras que para la segmentación basta con dividir la imagen en regiones diferentes, la clasificación implica un reconocimiento. En el caso de texturas, la clasificación consistirá en asignar cada píxel de la imagen a una determinada textura. Para llevar a cabo esta clasificación nos basaremos de nuevo en la matriz de descriptores descrita en la Sección 4.1.

La función de un clasificador es asignar cada uno de los datos de entrada a una de las N posibles clases previamente definidas. En nuestro caso, la entrada será la

matriz de descriptores obtenida en cada punto, y las clases posibles, el conjunto de texturas con el que compondremos nuestras imágenes. Asignaremos dos píxeles a la misma clase en función de su similitud. Esta asignación se basa en una regla de decisión, la cual compara la entrada (la matriz de descriptores) con una cierta información a-priori sobre las diferentes texturas consideradas. Este conocimiento previo acerca de las diferentes clases suele ser algún tipo de información estadística, que depende de la regla de decisión concreta (algunos metodos requieren sólamente el valor medio de la matriz de descriptores para cada clase, otros precisan de un cierto número de matrices de descriptores de cada clase, etc). De todas formas, sea cual sea la regla de decisión elegida, siempre se requiere un entrenamiento o aprendizaje previo del algoritmo, con las distintas clases posibles, para adquirir dicha información. Las imágenes de entrenamiento serán texturas puras.

La principal elección en el diseño del clasificador es pues la regla de decisión a emplear. Los criterios más comunes son aquellos basados en la distancia mínima [104], la regla del vecino más proximo (NNR, next neighbour rule) [105], o la clasificación Bayesiana [103][104]. Hemos escogido ésta última dado que es la óptima desde un punto de vista teórico (también se probó una basada en la mínima distancia, pero fue descartada por sus peores resultados). Un clasificador bayesiano se basa en la regla de decisión siguiente: dado un punto $\mathbf{x}$ que debe ser clasificado, y sean $w_1 ... w_Q$ las Q posibles clases, $\mathbf{x}$ será asignado a la clase w_i que maximize la probabilidad de que $\mathbf{x}$ pertenezca a dicha clase, $p(w_i|\mathbf{x})$. A partir de la regla de Bayes, suponiendo funciones de probabilidad gaussianas, y en una situación en la que todas las clases son igualmente probables a priori, se demuestra que dicha probabilidad es [104]:

$$p(w_i|\mathbf{x}) \propto \frac{1}{\sqrt{|\Sigma_i|}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m}_i)^T \Sigma_i^{-1} (\mathbf{x}-\mathbf{m}_i)} , \tag{4.3}$$

donde m_i es la media de la matriz o conjunto de descriptores para la clase w_i, Σ_i la matriz de covarianza de dicha clase y $|\Sigma_i|$ su determinante.

Por lo tanto, el proceso de clasificación consta de dos diferentes etapas. Primeramente un entrenamiento previo, consistente en extraer m_i y Σ_i para cada clase considerada, y en segundo lugar, la clasificación propiamente dicha, donde se asignará cada píxel de una imagen de prueba a la correspondiente clase, usando la información adquirida en el entrenamiento.

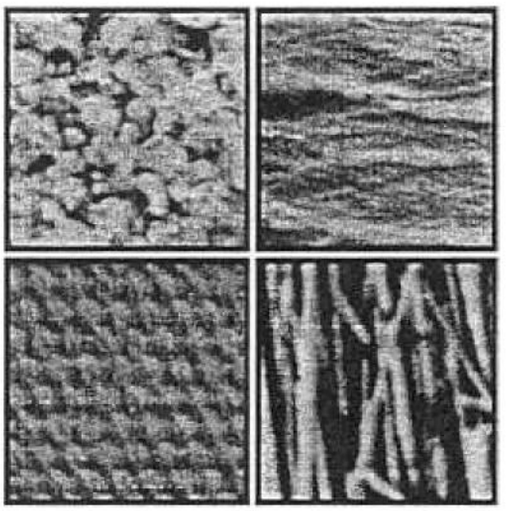


Figura 4.4: Texturas usadas como entrenamiento del clasificador de Bayes. De arriba abajo y de izquierda a derecha, arena, mar, tela de algodón y paja.

4.3.1 Entrenamiento

La fase previa de entrenamiento es fundamental, no sólo para obtener la información estadística de las distintas clases, sino también porque de esa información podremos extraer una estimación a priori de la capacidad de discriminación de nuestra matriz de descriptores. Con esta intención hemos elegido cuatro texturas del album de Brodatz [102] como imágenes de entrenamiento : mar [D37], tela de algodón [D77], paja [D15] y arena [D29]. Todas las imágenes empleadas eran de tamaño 128×128 píxeles, mostrandose en la Fig. 4.4. El esquema de Gabor se aplicó para obtener las matrices de descriptores en una red de muestreo con un espaciado de 8 píxeles en ambas direcciones. Posteriormente, calculamos la media y la matriz de covarianza de las 256 matrices resultantes. Una vez que este proceso se repite para cada una de las clases de entrenamiento, disponemos de los datos necesarios para realizar la clasificación bayesiana (ver Ec. (4.3)).

Antes de comenzar la clasificación propiamente dicha, podemos usar estos datos para obtener una estimación previa de la capacidad de nuestro conjunto de descriptores para esta tarea. Esto es especialmente útil en nuestro caso, dado que no tenemos nigún conocimiento previo al respecto. Se podría decir que un conjunto de descriptores es adecuado para una clasificación si es capaz de poner de manifiesto tanto las similitudes dentro de una clase, como las diferencias entre clases distintas. Una forma de medir esta capacidad es definir una "distancia" entre clases que diese una medida de cuanto se solapan entre ellas: cuanto menor es la distancia

entre dos clases, mayor es la probabilidad de error durante la clasificación. Este parámetro puede servir de comparación entre diversos métodos. Una de las distancias que han sido propuestas es la de Bhattacharyya [106][107][104], o B-distancia. La B-distancia (D_B) entre dos clases es una función de las distribuciones de probabilidad de ambas clases. Si suponemos distribuciones de probabilidad gaussianas [107][104], dicha distancia se puede expresar como :

$$D_B(w_i, w_j) = \frac{1}{2}(\mathbf{m_i} - \mathbf{m_j})^T(\frac{\Sigma_i + \Sigma_j}{2})^{-1}(\mathbf{m_i} - \mathbf{m_j}) + \frac{1}{2}\ln(\frac{\frac{1}{2}|\Sigma_i + \Sigma_j|}{\sqrt{|\Sigma_i| + |\Sigma_j|}}) \quad , \quad (4.4)$$

donde $\mathbf{m_i}, \mathbf{m_j}$ son las medias de los vectores descriptores de las dos clases y Σ_i, Σ_j sus matrices de covarianza. La B-distancia es una medida de cómo de similares o diferentes son las dos clases. Para dar una idea de sus valores y significado, diremos que en una clasificación donde las dos clases sean igual de probables a priori, la probabilidad de error está acotada por e^{-D_B}. Por ejemplo, si tenemos dos texturas con una B-distancia de 4, la probabilidad de error (en una clasificación en la que sólo intervinieran ellas dos) sería inferior a un 2% .

En la Tabla 4.1 se listan las B-distancias encontradas entre las texturas de entrenamiento de la Fig. 4.4. La B-distancia media es aproximadamente 8. Las texturas más "distantes" son la paja y la tela de algodón ($D_B = 13.1$), mientras que las más "próximas" son el mar y la arena ($D_B = 4.5$). De acuerdo con esta tabla, podemos esperar una buena clasificación (en [47] se puede encontrar una tabla de comparación entre las B-distancias obtenidas con otros métos con similares texturas).

Tabla 4.1: B-distancias entre pares de texturas procedentes del albúm de Brodartz usando la matriz 4×4 de descriptores de Gabor.

***	MAR	PAJA	ARENA	ALGODON
MAR	—			
PAJA	12.2	—		
ARENA	4.5	5.4	—	
ALGODON	6.6	13.1	6.4	—

La Tabla 4.2 muestra una comparación entre los resultados de aplicar varios esquemas de Gabor con un número distinto de canales de frecuencia y orientación.

Tabla 4.2: Comparación entre los resultados de diversos esquemas de Gabor de acuerdo a sus B-distancias interclases e intraclase.

Esquema de Gabor	Interclase	Intraclase	Variabilidad
4 Fr × 4 Or (Sólo módulo. Máscara 7 × 7)	7.1	1	7.1
4 Fr × 4 Or (Sólo módulo. Máscara 15 × 15)	7.6	0.7	10.8
4 Fr × 4 Or (Real e imag.)	7.5	1.2	6
4 Fr × 4 Or (Sólo real)	3.5	–	–
5 Fr × 4 Or (Sólo módulo)	10.3	1.2	8.6
4 Fr × 8 Or (Sólo módulo)	22.3	5.5	4.1

En la primera columna listamos una media de las B-distancias entre los seis pares posibles de texturas (la distancia interclase). La segunda columna es la distancia intraclase, es decir, la B-distancia entre diferentes zonas de una misma textura. Para que el clasificador reconociese diferentes partes de una misma textura como pertenecientes a la misma clase, esta distancia interna debería mantenerse baja. Esto es un aspecto muy importante a considerar, ya que no vale la pena tratar de incrementar la distancia interclase si la intraclase también aumenta en un factor similar. Un parámetro importante es pues el cociente entre la distancia externa y la interna, denominado variabilidad [104]. Este factor se muestra en la tercera columna. Aunque esta tabla debería considerarse sólamente como una estimación a priori de la capacidad para la clasificación, se pueden extraer algunas importantes conclusiones. Por una parte, se justifican las consideraciones expuestas en la sección 4.1 sobre las ventajas del uso del modulo de la matriz de descriptores frente a ambas componentes (en la fila tercera apenas mejora la distancia interclase e incluso disminuye la variabilidad) o frente al uso de s'olo uno de ellos (fila cuarta, dramática caída de los resultados usando el mismo número de descriptores). Asímismo, como ya se indicó en el Capítulo 3, el uso de una máscara mayor (15×15) aunque mejora ligeramente los resultados no compensa el multiplicar por 4 el coste de calcular la matriz de descriptores. Finalmente se han comparado esquemas que usan un mayor número de canales de orientación y frecuencia, en los que se aprecia que a pesar de elevar la distancia interclase (consecuencia lógica de un mayor número de descriptores) también elevan la intraclase, por lo que su variabilidad no mejora mucho (caso de 5 canales de frecuencia), o incluso empeora (esquema con 8 canales de orientación). En conclusión, no parece que el uso de un esquema más complicado vaya a mejorar los resultados significantemente, pudiendo incluso empeorarlos. Esta fué una razón más que nos confirmó en nuestro simple esquema 4 × 4.

A pesar de todo, estas estimaciones previas a partir de las B-distancias deben ser confirmadas con los resultados de una verdadera clasificación. La razón es que la B-distancia y el error en la clasificación sólo estan relacionados matemáticamente en el caso específico de una clasificación entre sólo dos clases (algunos resultados concernientes a teoría de clasificaciones multiclases pueden encontrarse en [108]). Además, dado que la B-distancia es una medida estadística, depende mucho del tipo de muestras tomadas en considearción, de su número, etc, por lo que las comparaciones con estimaciones de otros métodos son difíciles. Sin embargo, a pesar de todas estas limitaciones, este estudio preliminar muestra que cabe esperar buenos resultados en una clasificación, por lo que nuestra matriz de descriptores parece ser un buen conjunto de descriptores de la textura.

4.3.2 Asignación de píxeles

En este apartado se presentarán los resultados de la clasificación de un conjunto de imágenes conteniendo 1, 2 ó 4 texturas, tanto en porcentajes de píxeles correctamente clasificados, como gráficamente, mostrando la distribución espacial de los aciertos y errores. También se proponen diversas soluciones para el principal problema que, al igual que en el caso de la segmentación, es clasificar correctamente los puntos cerca de los límites entre texturas. Para todo ello hemos compuesto varias imágenes a partir de las texturas mostradas en Fig. 4.4. Es importante resaltar el hecho de que las texturas que van a ser clasificadas no son los mismos fragmentos usados en el entrenamiento; es decir, no estamos clasificando las mismas imágenes que se usaron para extraer la información a priori de las clases. Ejemplos de algunas de las imágenes test se muestran en la Fig. 4.5a (textura única: paja), Fig. 4.6a y 4.7a (dos texturas: mar–tela de algodón y paja–arena), y Fig. 4.8a (cuatro texturas: mar, paja, arena y tela de algodón).

A continuación se detalla el proceso completo que se ha realizado. Primeramente se aplica el esquema de Gabor a las imágenes test y se calcula la matriz de descriptores en los puntos a clasificar (en este caso todos los píxeles de la imagen). Luego, cada píxel se asigna a una de las cuatro clases a través de la regla de decisión en base a su matriz de descriptores. Esto equivale a calcular Ec. (4.3) para cada clase (usando sus valores de m_i y $\sum_i$) y entonces asignar el píxel a la clase que maximize $p(w_j|\mathbf{x})$. Este procedimiento se repite para cada píxel hasta completar la clasificación de toda la imagen.

Los resultados se sumarizan en la Tabla 4.3, donde se listan los porcentajes de clasificación correcta para las imágenes test empleadas. En la columna etiquetada

Tabla 4.3: Porcentajes de píxeles clasificados correctamente usando 4×4 descriptores en imágenes conteniendo varias texturas.

TIPO	TEXTURAS	% (cada clase)	% (global)	Valor Medio
1 textura	Mar	99	99	
	Paja	100	100	99
2 texturas	Mar—Paja	87—96	91	
	Mar—Arena	96—97	97	
	Mar—Algodón	99—64	81	
	Paja—Arena	97—93	95	88
	Paja—Algodón	97—64	81	
	Arena—Algodón	99—60	79	
4 texturas	Algodón—Paja—Mar—Arena	39—98—85—89	78	78

[% (cada clase)] se presenta el porcentaje de aciertos para cada una de las texturas individuales que componen una imagen. En la siguiente columna se ha calculado el porcentaje correcto en la clasificación de cada imagen, promediando los resultados de las texturas que la componen. Finalmente, en la tercera columna, se listan los valores medios para los casos correspondientes a imágenes compuestas de una, dos y cuatro texturas respectivamente.

En las Figs. 4.5b, 4.6b, 4.7b y 4.8b se puede apreciar la distribución espacial de los aciertos y errores corespondientes a las imágenes test antes citadas. En estas figuras la asignación a una textura determinada se ha representado como un nivel de gris. En una escala de claro a oscuro tenemos mar (blanco), paja (gris claro), arena (gris oscuro) tela de algodón (negro) respectivamente. Se puede apreciar que el error en la clasificación, siendo muy bajo en imágenes compuestas de una única textura, se va incrementando a medida que aumenta el número de texturas. En imágenes con una sóla textura el error de clasificación es aproximadamente del 1%. Imágenes con dos texturas presentan dos tipos de comportamiento que son representados en la Fig. 4.7b (paja/arena) y Fig. 4.6b (mar/algodón). El primer caso presenta una clasificación muy buena ($\simeq$ 95%). No hay problemas considerables a lo largo del borde entre texturas, aunque esa zona muestra un porcentaje mayor de píxeles mal clasificados. Sin embargo, en la Fig. 4.6b (mar/algodón) el mar está bien clasificado (97 %), pero un considerable número de errores aparece en la textura de tela de algodón (de la cual sólo un 64% de los píxeles están bien clasificados).

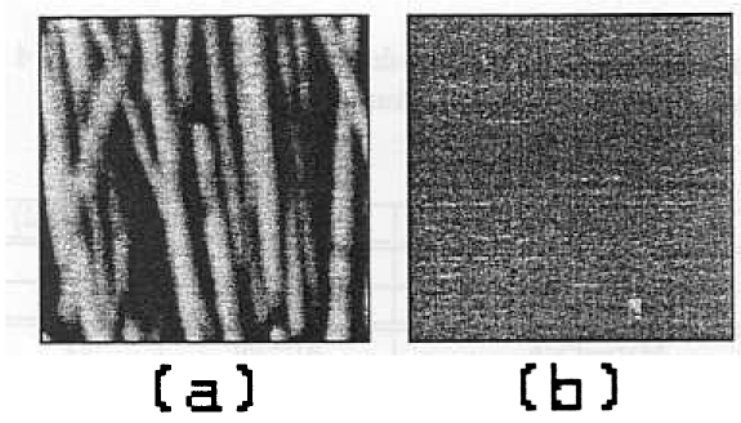


Figura 4.5: (a) Imagen conteniendo una textura: paja. (b) Clasificación bayesiana de (a) usando 4×4 descriptores.

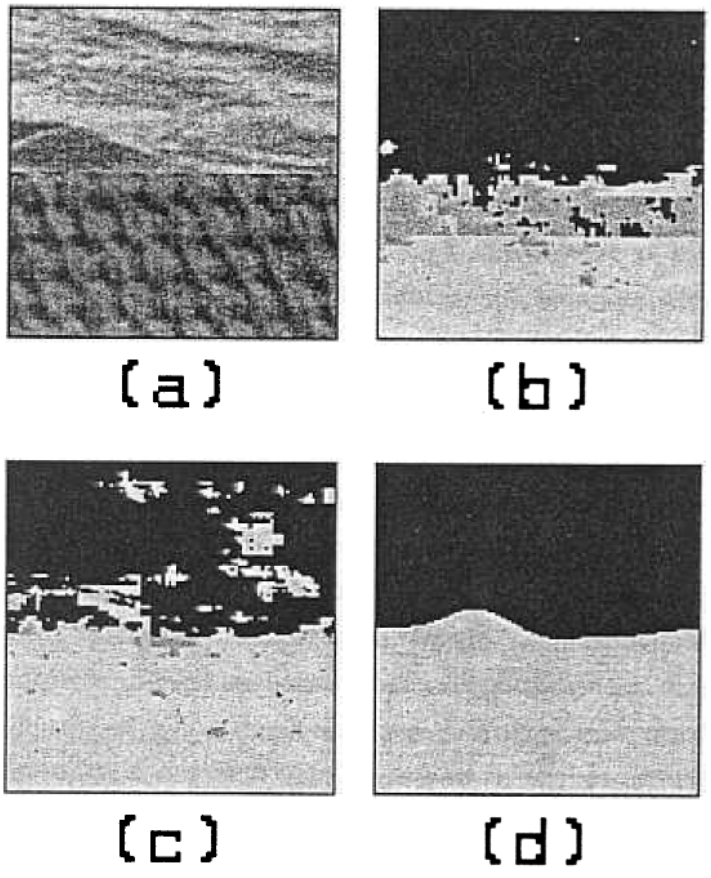


Figura 4.6: (a) Imagen conteniendo dos texturas: mar(arriba) y tela de algodón(abajo). (b) Clasificación bayesiana usando 4×4 descriptores. (c) Lo mismo, usando 3×4 descriptores. (d) Resultados después de procesar (c) con un filtro de moda.

Esto ocurre también en otras imágenes que contienen dicha textura, como se aprecia en la Tabla 4.3. Mientras que la media de clasificación correcta en otras imágenes con dos texturas es del 95%, este valor desciende a un 62% para la textura algodón. Este problema podría explicarse por el peculiar espectro de dicha textura, el cual está altamente concentrado en unos pocos puntos, debido a que es una textura artificial con una alta periodicidad. Ninguna de las otras texturas muestran tal característica de forma tan acusada. Por último, en la imagen con cuatro texturas (Fig. 4.8b) seguimos obteniendo un alto nivel de asignaciones correctas (78%), pero el problema de la clasificación de los puntos en las fronteras todavía persiste.

De la exposición anterior se deduce que el esquema de Gabor proporciona buenos descriptores de la textura, mientras que la mayoría de los problemas que empeoran la clasificación aparecen a lo largo de los límites entre dos texturas, como sucedía en el caso de la segmentación. La razón de esto es que la información a priori usada en la clasificación se obtuvo con texturas puras. Consecuentemente, la matriz descriptora de un punto cerca de un borde no se ajusta en realidad a ninguna de estas clases puras. El origen del problema es que los filtros de Gabor mezclan información de un área alrededor del punto de aplicación: en el caso de los filtros de frecuencia más baja, esta área puede ser del orden del 50×50 píxeles. Esto nos indica ya una posible manera de mejorar los resultados de forma sencilla, que consistiría simplemente en prescindir de la información de los canales de frecuencia más baja.

Con esta idea repetimos el proceso de clasificación usando una matriz de descriptores de dimensiones 3×4 (tres frecuencias y cuatro orientaciones). Los nuevos porcentajes de clasificación correcta se listan en la Tabla 4.4, y las Fig. 4.6c y 4.8c muestran los resultados de aplicar este método a las imágenes de las Fig. 4.6a y 4.8c respectivamente. Dado que se están usando menos descriptores, las clasificaciones de puntos internos (alejados de los bordes) empeoran. Sin embargo, la clasificación mejora notablemente en áreas que antes presentaban problemas, como por ejemplo la textura algodón, o las zonas cerca de los límites entre texturas.

Estas ventajas superan al inconveniente antes citado, ya que como se ve en la Fig. 4.8c la clasificación en los bordes para la imagen con cuatro texturas se mejora notablemente, y el problema de la mala clasificación del algodón se ha solucionado (ver también Fig. 4.6c, así como los porcentajes en Tabla 4.4). La clasificación en la imagen con 4 texturas se ha incrementado hasta un 84%, mientras que la de la textura algodón ha pasado de un 64% a un 86%.
Queremos resaltar que los resultados mostrados hasta ahora, en las figuras y en

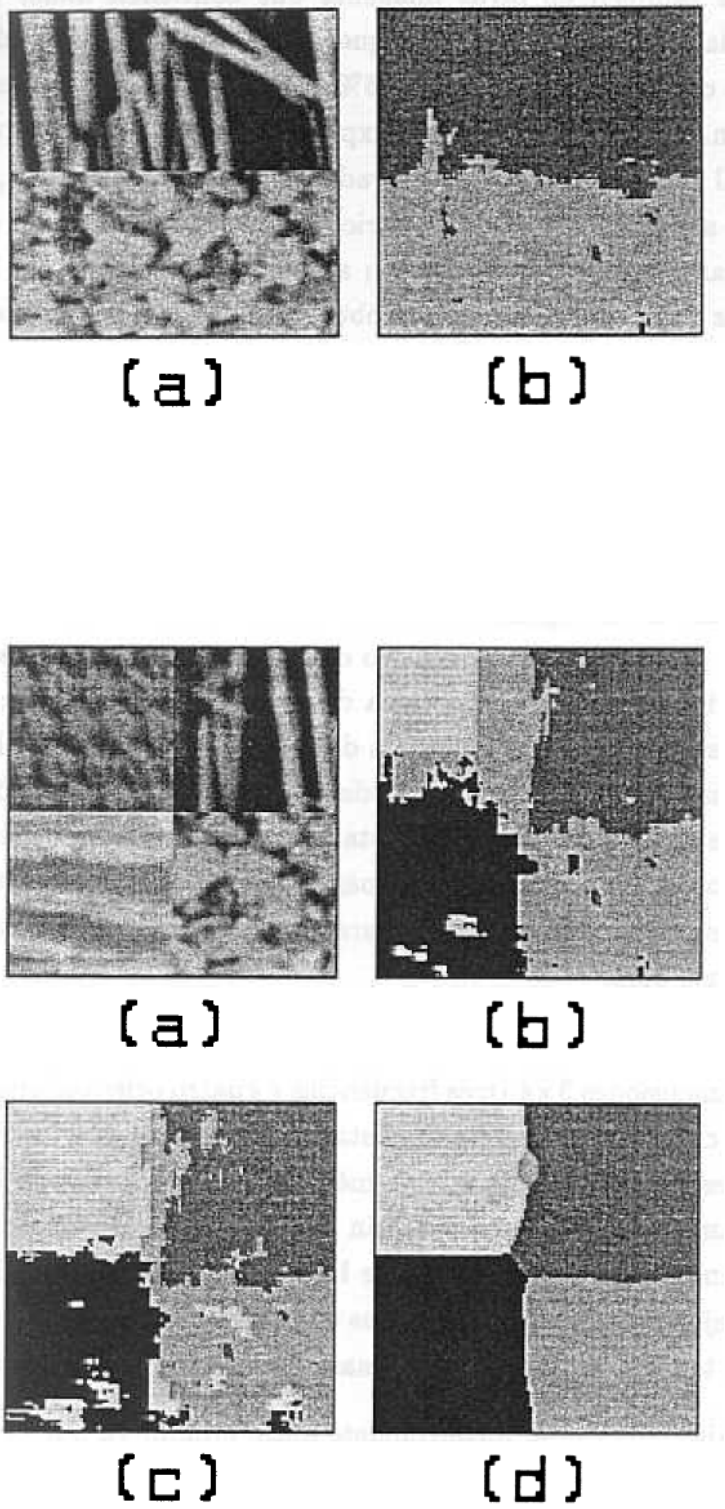


Figura 4.8: (a) Imagen conteniendo cuatro texturas (de arriba abajo y de izquierda a derecha: tela de algodón, paja, mar y arena. (b) Clasificación bayesiana usando 4×4 descriptores. (c) Lo mismo, usando 3×4 descriptores. (d) Resultados después de procesar (c) con un filtro de moda.

Tabla 4.4: Porcentajes de clasificación correcta usando 3 × 4 descriptores en imágenes conteniendo varias texturas.

TIPO	TEXTURAS	% (cada clase)	% (global)	Valor medio
1 textura	Mar	95	95	
	Paja	97	97	96
2 texturas	Mar—Paja	82—90	85	
	Mar—Arena	87—92	90	
	Mar—Algodón	83—85	84	
	Paja—Arena	96—82	89	86
	Paja—Algodón	97—82	89	
	Arena—Algodón	92—83	88	
4 texturas	Algodón—Paja—Mar—Arena	78—93—83—82	84	84

las tablas, son la salida directa del clasificador bayesiano, sin ningún procesado posterior. Sin embargo, lo usual es aplicar algoritmos adicionales para mejorar los resultados. Una solución sería el uso de algoritmos especificos que tomaran en cuenta el problema de las fronteras cuando se usan descriptores locales.

Un algoritmo de esas características ha sido propuesto recientemente [109], siendo un método general que se puede usar independientemente de los descriptores particulares que se hayan escogido, mejorando la calidad de la clasificación en las fronteras. Aquí no obstante, propondremos dos métodos muy sencillos que pueden mejorar significativamente los resultados obtenidos en la clasificación:

1. Se puede aplicar un filtro de moda para eliminar puntos aislados, como hicimos en el caso de la segmentación. Ahora empleamos un filtro de mayor tamaño (16 × 16) que además se aplica 5 veces sucesivas. En la Fig. 4.6d y 4.8d se muestran los resultados de este proceso cuando se aplica a las Fig 4.6c y 4.8c respectivamente. Las fronteras quedan ahora más claramente definidas, y por ejemplo, en el caso de la imagen con cuatro texturas, el porcentaje de aciertos en la clasificación ha aumentado hasta el 94%. El método presenta sin embargo una clara desventaja: los detalles en imágenes más complicadas podrían difuminarse. Por lo tanto, el tamaño del filtro debería ajustarse a la resolución requerida en la imagen que está siendo estudiada.

2. Un método fácil que da buenos resultados es llevar a cabo una segmentación

previa a la clasificación. Una vez que hemos realizado la segmentación, las diversas regiones podrían clasificarse a través de su matriz de descriptores media. Este enfoque presenta dos ventajas principales. En primer lugar se reduce enormemente el coste computacional, ya que sólo unas pocas regiones deberán ser clasificadas en lugar de todos los píxeles. Por otra parte, dado que lo que se usa es la matriz media, la influencia de los puntos cerca de las fronteras se reducirá notablemente. De hecho, este proceso se aplicó a las imágenes segmentadas en la sección anterior (Fig. 4.2b y 4.3b), y en todos los casos las regiones fueron correctamente identificadas. Consecuentemente, los resultados mostrados en las figuras de la sección 4.2 constituyen a la vez una segmentación y una clasificación.

Concluyendo, el esquema de Gabor proporciona resultados muy satisfactorios tanto en la segmentación como en la clasificación de texturas, como se ha comprobado usando métodos convencionales. Los resultados aquí presentados pueden ser fácilmente mejorados si se consideran procesados adicionales a posteriori. Es también posible ajustar o modificar el número de canales del esquema para mejorar las prestaciones en casos concretos.

Capítulo 5

Invariantes ante cambios de escala y/o rotaciones

En las situaciones de la vida real, las tareas de reconocimiento y clasificación que se han mostrado en el capítulo anterior presentan dificultades adicionales. En el mundo real nos encontraremos con texturas con diferentes escalas y con distintas orientaciones y a pesar de ello, un observador humano suele ser capaz de reconocerlas. Entre las numerosas capacidades que se deberían esperar de un sistema de visión está la de reconocer objetos a pesar de cambios de escala (dentro de un rango razonable) y orientación [110][111]. La búsqueda de invariantes ante cambios de escala y orientación es pues un problema fundamental en campos tales como el procesado de imágenes, inspección automática, visión artificial, prospección remota, etc. La importancia de estas invariancias estriba en la necesidad de reducir la alta dimensionalidad del mundo real, conteniendo un sinfín de posibilidades (por ejemplo, en cuanto a escala y orientaciones), a un sistema mucho menos complejo, que pueda ser manejado más fácilmente.

En este Capítulo se estudiará un posible enfoque del problema usando nuestro esquema de Gabor y la matriz de descriptores presentada en el Capítulo 4. Se mostrará cómo el esquema de Gabor permite predecir de forma muy sencilla la relación entre los datos obtenidos de una imagen modificada (rotada o ampliada) y los procedentes de la imagen original. Esto se pondrá de manifiesto en términos de cómo la matriz de descriptores cambia al rotar o escalar una imagen. Primeramente (Sección 5.1) discutiremos la rotación y en la siguiente Sección estudiaremos el caso más complejo de un cambio de escala. Finalmente, por su especial importancia, dedicaremos la Sección 5.3 a la aplicación de nuestro esquema al caso particular de las imágenes fractales.

5.1 Invariantes bajo rotación

La matriz de descriptores extraida de nuestro esquema de Gabor se compone de 4×4 elementos correspondientes a diferentes frecuencias (filas) y orientaciones (columnas). Dado que el esquema consta de 4 canales de orientación sintonizados a 0°, 45°, 90° y 135°, una rotación de 45° (o múltiplo) de la imagen original causará una permutación en la salida de los diferentes canales de orientación y por lo tanto, una permutación en las columnas de la matriz de descriptores. En la Fig. 5.1a se ilustra el efecto de una rotación de 45° en la matriz de descriptores. Cuando se rota la imagen, el valor que antes correspondía al canal horizontal aparecerá ahora en el canal correspondiente a 45°. Rotaciones sucesivas de 45° provocarán las correspondientes permutaciones de la matriz de descriptores.

$$\begin{pmatrix} a_1 & \mathbf{a_2} & a_3 & a_4 \\ b_1 & \mathbf{b_2} & b_3 & b_4 \\ c_1 & \mathbf{c_2} & c_3 & c_4 \\ d_1 & \mathbf{d_2} & d_3 & d_4 \end{pmatrix} \Longrightarrow \text{Rotación } 45° \Longrightarrow \begin{pmatrix} \mathbf{a_2} & a_3 & a_4 & a_1 \\ \mathbf{b_2} & b_3 & b_4 & a_2 \\ \mathbf{c_2} & c_3 & c_4 & a_3 \\ \mathbf{d_2} & d_3 & d_4 & a_4 \end{pmatrix}$$

(a)

$$\begin{pmatrix} \mathbf{a_1} & \mathbf{a_2} & \mathbf{a_3} & \mathbf{a_4} \\ b_1 & b_2 & b_3 & b_4 \\ c_1 & c_2 & c_3 & c_4 \\ d_1 & d_2 & d_3 & d_4 \end{pmatrix} \Longrightarrow \text{Ampliación } (1{:}2) \Longrightarrow \begin{pmatrix} ? & ? & ? & ? \\ \mathbf{a_1} & \mathbf{a_2} & \mathbf{a_3} & \mathbf{a_4} \\ b_1 & b_2 & b_3 & b_4 \\ c_1 & c_2 & c_3 & c_4 \end{pmatrix}$$

(b)

Figura 5.1: Efecto de rotaciones (arriba) y cambios de escala (abajo) de una imagen sobre la matriz de descriptores. Una rotación de 45° causa la permutación de una columna, mientras que una ampliación se refleja en un desplazamiento de las filas

Esto puede ser comprobado con texturas reales. Con este propósito, se digitalizaron varias texturas (mar, arena, paja y rafia[D84]) del álbum de Brodatz [102] y sus correspondientes versiones rotadas. Las texturas rotadas se obtuvieron girando 45° sucesivamente las fotografías originales antes de digitalizarlas con la cámara. Queremos resaltar que la rotación se hizo manualmente, con lo que el angulo girado es sólo aproximado. Además, el area digitalizada en cada imagen no es exactamente la misma. La Fig. 5.2a muestra cuatro versiones rotadas de la textura rafia. A estas cuatro imágenes se aplicó nuestro esquema de Gabor y se

extrajo la matriz de descriptores en diversos puntos, calculando la media para cada imagen. A continuación se realizó una permutación en cada matriz para corregir los efectos de la rotación: las matrices de imágenes con una rotación de 45° se permutaron una columna, las de 90° dos, y asi sucesivamente. La correlación existente entre los valores de la matriz de descriptores de la textura original (abscisas) y los de las matrices corregidas por la permutación (ordenadas) se muestra en la Fig. 5.2b. La mayoría de los puntos están cerca de la línea ideal $y = x$. Esto demuestra que la idea intuitiva original de que los efectos de rotar la imagen se pueden cancelar por una adecuada permutación de los canales de Gabor, es aplicable a casos reales. Estos resultados son tanto más interesantes si consideramos que las versiones rotadas de las texturas no se obtuvieron a partir de una simulación ideal en el ordenador, sino que son inexactas en dos sentidos: el angulo de giro es sío aproximadamente de 45°, y además no corresponden a la misma porción original de textura.

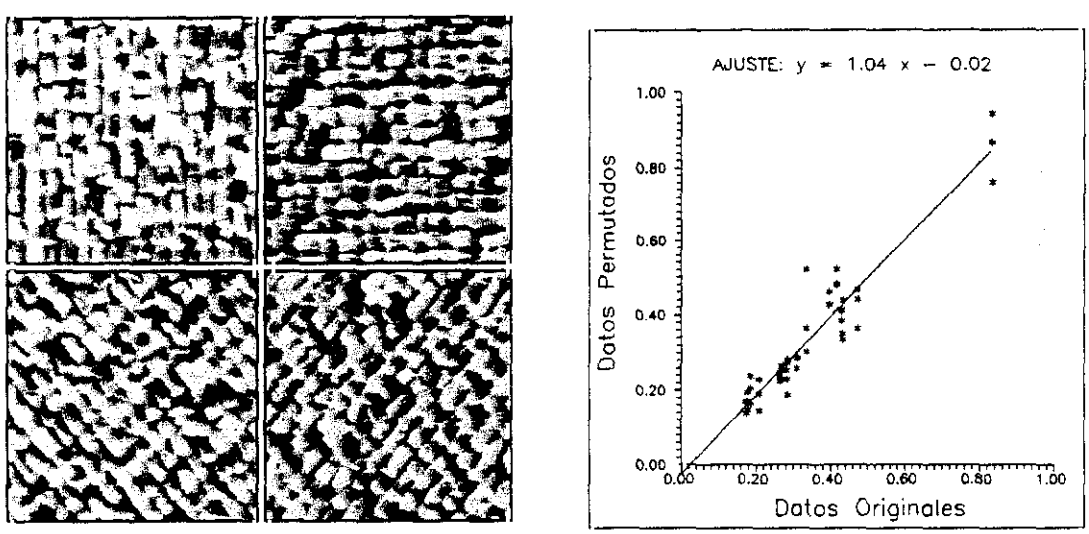


Figura 5.2: (a) Cuatro versiones rotadas de la textura rafia. (b) Correlación entre los descriptores de la textura no rotada (eje x) y los de sus versiones rotadas (eje y), tras haber sido corregidos con una adecuada permutación de columnas.

Una vez que hemos comprobado en la práctica el comportamiento de la matriz de descriptores ante rotaciones, el siguiente paso será estudiar su posible aplicación al reconocimiento de texturas rotadas. Debido al reducido número de canales de orientación (4) en nuestro esquema, la propiedad de la permutación de las columnas solo se cumple cuando la rotación es un multiplo de 45° (aunque esto podría mejorarse introduciendo más canales). A pesar de esto, vamos a mostrar cómo es posible el reconocimiento de texturas rotadas un ángulo cualquiera a través del uso

de descriptores invariantes ante rotación. En nuestro caso, cualquier parámetro extraible de la matriz de descriptores que no cambie ante una permutación de columnas será un invariante ante rotación. La elección más simple es reducir la matriz a un vector promediando las 4 orientaciones:

$$v_f = \sum_\theta M_{f,\theta} \ , \tag{5.1}$$

donde $M_{f,\theta}$ son los componentes de la matriz de descriptores (Ec. 4.1) y los v_f, el resultado de promediar las cuatro orientaciones, son a su vez los componentes de un nuevo vector **v**, que es invariante ante rotaciones. Este vector se relacionaría con la llamada piramide laplaciana [59], un conocido esquema piramidal que no incorpora selectividad en orientaciones.

De esta manera, una permutación de las columnas (es decir, una rotación de la imagen) no alteraría este nuevo vector de descriptores **v**. Con el propósito de comprobar la invarianza de estos nuevos descriptores escogimos 4 texturas del álbum de Brodatz (mar, paja, rafia, y arena) y digitalizamos cinco variaciones de cada una de ellas (la original, y rotaciones a 45°, 90° y 135°, además de una rotación con un angulo arbitrario, no múltiplo de 45°). Este conjunto de 20 imágenes rotadas se obtuvo con el mismo procedimiento que se explicó anteriormente. Una importante consideración es que debido al hecho de usar el módulo de los elementos de la matriz, nuestro esquema es ya invariante ante rotaciones de 180°. Para detectar rotaciones de este tipo tendríamos que tomar en cuenta también la fase de la matriz.

El siguiente paso fué repetir el mismo procedimiento de clasificación presentado en el capítulo anterior. Primeramente el esquema de Gabor se aplicó a las

Tabla 5.1: Comparación de las distancias interclases, intraclase y variabilidad entre versiones rotadas de varias texturas usando los descriptores invariantes ante rotación (Ec. 5.1).

Textura	Interclase	Intraclase	Variabilidad
ARENA	2.2	0.1	22
RAFIA	4.6	0.2	23
PAJA	2.7	0.1	27
ALGODON	7.7	0.3	25

20 imágenes (tamaño 128×128) obteniendose la matriz de descriptores en 16×16 puntos con un espaciado de 8 píxeles en ambas direcciones. De estas matrices se extrajeron los nuevos vectores descriptores invariantes ante rotación (según Ec. 5.1). A partir de estos datos se calcularon las B-distancias (Ec. 4.4) entre pares de texturas para los dos casos : distancia intraclase, entre versiones rotadas de la misma textura, y distancia interclase, entre diferentes texturas no rotadas. En la Tabla 5.1 se listan los resultados así obtenidos. En la primera columna se muestran la media de las distancias interclase entre la textura citada y todas las demás; en la segunda, la media de las distancias intraclase entre cada textura original y sus cuatro versiones rotadas. Finalmente en la última columna se lista el cociente entre las distancia interclase e intraclase, o variabilidad. Se aprecia que las B-distancias entre diferentes texturas son en general menores que las obtenidas cuando se usaban todos los descriptores (Tabla 4.1), lo que era de esperar, dado que ahora estamos usando sólo 4 descriptores en vez de 16. Esto nos indica que si no estamos interesados en el problema de reconocer texturas rotadas siempre será mejor usar la matriz de descriptores completa en la clasificación. La B-distancia media entre texturas diferentes es de 4.3, mientras que la distancia media entre versiones rotadas de la misma textura es 0.2. El punto importante aquí es que la distancia intraclase (incluso para el caso de la imagen rotada un ángulo arbitrario) es mucho menor que la distancia interclase entre diferentes texturas, o en otras palabras, que la variabilidad es alta, como se aprecia en la Tabla 5.1. Como se expuso en el Capítulo 4 cuanto mayor sea dicho parámetro, menor es la posibilidad de asignar una textura rotada a una clase diferente. Dado que los valores de la variabiliad son incluso mayores que los mostrados en la Tabla 4.2, esto parece asegurarnos una buena clasificación usando el nuevo vector de descriptores **v**. Esto se ha comprobando usando un clasificador bayesiano al igual que en el capítulo anterior. Los datos de entrenamiento son aquí las texturas originales, mientras que las rotadas constituyen las imágenes test que intentaremos clasificar. De esta forma, de las 16 texturas giradas, todas ellas excepto una (arena rotada 135°) fueron identificadas correctamente, incluso aquellas que habían sido rotadas un angulo no multiplo de 45°.

Los parámetros invariantes ante rotación que se han usado hasta ahora pueden identificar una textura rotada como tal, pero son incapaces (debido a su propia invarianza) de decirnos nada acerca de cuánto ha sido rotada. Si estamos interesados en obtener una estimación acerca del angulo de rotación, un simple método podría consistir en dos etapas:

1. Los parámetros invariantes v_j previamente definidos se usan para identificar

la textura.

2. Una vez que la textura ha sido identificada, podemos usar la matriz de descriptores completa (los 16 elementos) para conocer la orientación correcta. Esto se podría hacer fácilmente comparando la matriz de descriptores de la imagen test con sucesivas permutaciones de la matriz de la textura sin rotación. El número de permutaciones necesarias para reducir al mínimo las diferencias entre ambas matrices nos dará una estimación aproximada (con una precisión de 45°) de la rotación de la imagen test. Usando una interpolación entre los valores de las dos matrices que estén más próximas podríamos obtener una mayor resolución del angulo de rotación.

5.2 Invariantes ante cambios de escala

Un cambio de escala de la imagen modifica la matriz de descriptores en una forma similar al efecto de una rotación, causando una redistribución de sus elementos. Cuando una imagen se aumenta en un factor 2, hay un desplazamiento en las filas de la matriz de descriptores tal como se muestra en la Fig. 5.1b. Puede apreciarse que hay una fila nueva correspondiente a la frecuencia más alta, mientras que la fila de frecuencia más baja de la vieja imagen desaparece. La razón de este cambio radica en el hecho de que en nuestro esquema piramidal los sucesivos niveles de resolución se obtienen reduciendo la imagen a la mitad. Por lo tanto, si comenzamos con una imagen ampliada (con un factor 2), después de la primera reducción, los resultados serán los mismos que con la original: de aquí el corrimiento vertical en la matriz. Con un factor de ampliación 4 tendríamos un corrimiento de dos filas. El efecto de aplicar el esquema a una imagen reducida sería un desplazamiento de filas en la dirección opuesta.

Podemos comprobar fácilmente esta propiedad con ejemplos reales, de la misma forma que hicimos con las rotaciones. Para ello, se han usado tres texturas (arena, rafia y tela de algodón), digitalizandose dos versiones de cada una, con un factor de escala entre ellas de 2, cambiando la posición de la cámara. Al igual que en el caso de las rotaciones, el factor de escala no se controló de forma muy precisa y tampoco nos preocupamos de obtener la misma zona de la textura en ambas imágenes. En la Fig. 5.3a podemos ver estos tres pares de texturas. De forma similar a como hicimos en el caso de la rotación, se calculó la media de la matriz de descriptores para cada imagen, se desplazaron las filas de las matrices correspondientes a las versiones escaladas, y finalmente se compararon dichas matrices

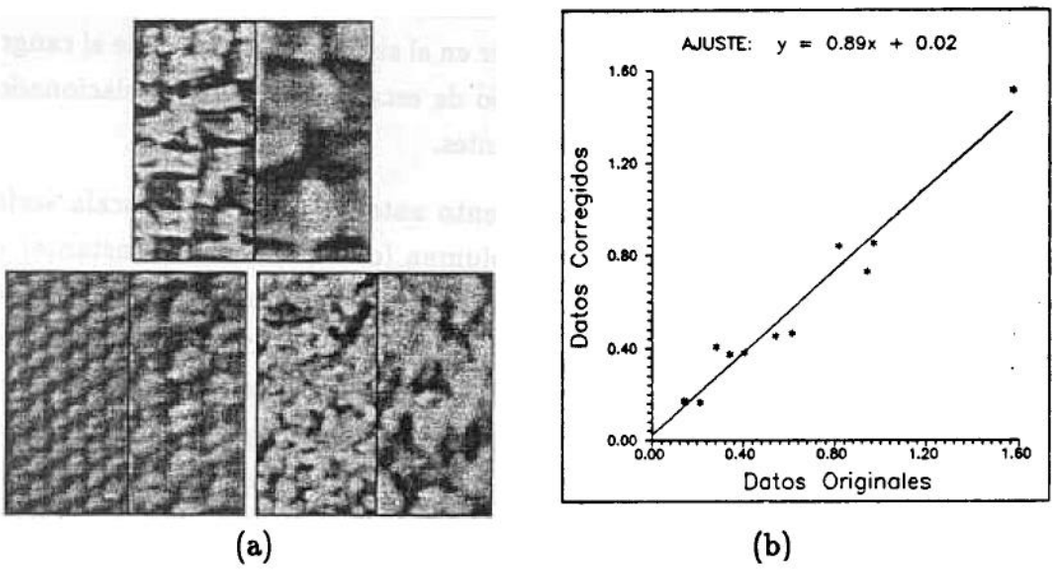


Figura 5.3: (a) Tres texturas (en el sentido de las agujas del reloj, rafia, arena y algodón) y sus versiones ampliadas en un factor 2. (b) Correlación entre los valores de los descriptores correspondientes a las dos versiones de la textura rafia tras haber sido corregidos con un adecuado desplazamiento de filas.

con las de las texturas no escaladas. Ahora la comparación esta limitada a 12 de los 16 descriptores, dado que en el proceso de cambio de escala se pierde una fila. En la Fig. 5.3b se muestra la correlación de dichos descriptores para el caso de la textura rafia: en las abscisas tenemos los valores de la imagen original y en las ordenadas los de la imagen escalada. Se aprecia que, como en el caso de la rotación, la correlación es muy buena.

A pesar de la similitud de los efectos de una rotación y un cambio de escala sobre la matriz de descriptores, es mucho más dificil encontrar parámetros invariantes y desarrollar un algoritmo de reconocimiento práctico para este segundo caso. El principal obstaculo es que aquí lo que tenemos no es una permutación sino un desplazamiento. Por lo tanto, incluso en el mejor caso, un cuarto de la información (una fila de un total de cuatro) se pierde. El problema se acrecienta cuando tenemos cambios de escala mayores (una ampliación o reducción por un factor 4 hace que la mitad de la información original se pierda). Esta situación no deja de ser realista, ya que nuestro esquema está basado en un modelo simple del sistema visual, e incluso un observador humano podría fracasar en el reconocimiento de una textura si ésta se aumenta lo suficientemente. Incluso con factores 2 ó 4 podemos tener dificultades identificando una textura escalada (ver Fig. 5.3a). En nuestro esquema, la principal limitación viene determinada por el número de canales, ya

que un mayor número de canales (filas) disminuiría la importancia de la pérdida de alguno de ellos. Algo similar podría suceder en el sistema visual, donde el rango de reconocimiento de texturas con un cambio de escala podría estar relacionado con el numero de canales de frecuencia presentes.

Una posible solución para el reconocimiento ante un cambio de escala sería tratar de ajustar los descriptores de cada columna (con orientación constante) a una curva determinada. De esta forma la textura se caracterizaría a través de los parámetros de dicha curva, en vez de a traves de los descriptores originales. Así, aunque algunos de los puntos varíen al cambiar la escala, se mantendría el ajuste original, y la textura sería identificable. De hecho, el espectro de potencias de las texturas e imágenes naturales tiende a presentar un decaimiento exponencial o lorentziano [95][96]. Sin embargo, este enfoque presenta dos principales dificultades. Primeramente, es muy difícil ajustar una curva de forma fiable con sólo cuatro puntos. Además, como discutiremos con mayor detalle en el apartado siguiente, el decaimiento del espectro de potencias de texturas naturales no siempre acontece de forma suave. De hecho el decaimiento lorentziano o exponencial es sólo una tendencia que aparece cuando se promedian numerosos casos, mientras que ejemplos particulares pueden presentar grandes diferencias frente a ese comportamiento medio. Esto hace que el ajuste de las columnas de la matriz de descriptores sea ineficaz. En el siguiente apartado veremos que este enfoque del problema sólo es posible cuando consideramos tipos especiales de texturas, tales como los llamados fractales brownianos, cuyos espectros presentan decaimientos regulares que siguen una fórmula simple.

5.3 Dimensión Fractal y Funciones de Gabor

Recientemente, se han aplicado diversos parámetros extraidos de la geometría fractal al procesado de imágen, principalmente en el caso de análisis [89][90][112] y síntesis [91] de texturas. En este apartado mostraremos cómo el esquema de Gabor es aplicable también a la clasificación de imágenes fractales, así como a la determinación de su dimensión fractal. También discutiremos la posibilidad de aplicar métodos similares al estudio de texturas naturales.

Un fractal es, por definición, un conjunto para el cual su dimensión Hausdorff–Besicovich (dimensión fractal) excede su dimensión topológica [113]. En el caso de imágenes, la dimensión topológica de una imagen $z = I(x, y)$ es de 2, mientras que su dimensión fractal puede variar entre 2 y 3. Intuitivamente la dimensión fractal

de una imagen se relaciona con la apreciación visual de su rugosidad [89][90]. Sin embargo, si pretendemos estudiar imágenes a través de su dimensión fractal, es preciso aclarar previamente varios aspectos. En primer lugar, los ordenadores trabajan con imágenes discretas, cuya dimensión topológica es por definición nula, ya que dichas funciones son sólo no nulas en un conjunto discreto de puntos; por lo tanto, sólo podremos hablar de propiedades fractales dentro de un rango de resolución determinado. En segundo lugar, mientras que las tres dimensiones de un fractal real son equivalentes (todas ellas son dimensiones espaciales), una imagen es una superficie $z = f(x, y)$, donde la componente z (nivel de gris) es diferente de las coordinadas espaciales x y y. De aquí en adelante, será conveniente recordar estas limitaciones cuando hablemos de la dimensión fractal de las imágenes.

En este apartado nos vamos a centrar en los llamados fractales brownianos, que pueden ser generados digitalmente por el método propuesto por Voss [114]. Estos fractales tienen la propiedad de que su espectro de Fourier decae de acuerdo con la regla siguiente:

$$|H(f)| \propto f^{-m} \ , \tag{5.2}$$

donde $|H(f)|$ es el modulo del espectro, f la frecuencia espacial, y m el parámetro del decaimiento, que está relacionado con la dimensión fractal [113][89].

En la Fig. 5,4a se muestran cuatro imágenes de fractales brownianas generadas siguiendo el método de Voss, con dimensiones fractales 2.1, 2.3, 2.6 y 2.9 respectivamente. Se aprecia la relación directa entre la dimensión fractal y el aspecto rugoso como se indicó anteriormente. Sobre estas imágenes aplicaremos los métodos de clasificación y segmentación del Capítulo 4 y posteriormente mostraremos cómo el esquema de Gabor puede ser utilizado para estimar su dimensión fractal.

5.3.1 Segmentación y clasificación

En la Fig. 5.4b podemos ver los resultados de una segmentación de la Fig. 5.4a (compuesta de 4 fractales difierdo únicamente en su dimensión fractal) usando el mismo método (algoritmo de K medias) descrito en el Capítulo 4.

Los resultados son buenos, habiendose diferenciado claramente las cuatro áreas de la imagen con distinta dimensión fractal, aunque sigue presentandose un problema en las fronteras, principalmente entre las dimensiones 2.1 y 2.3. También realizamos una clasificación bayesiana de dicha imagen.

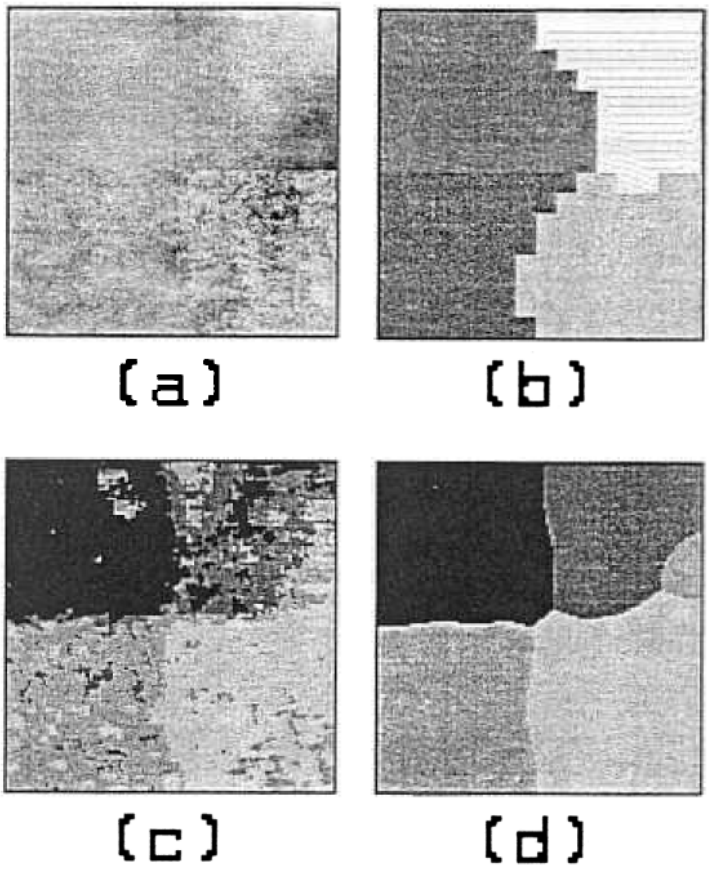


Figura 5.4: (a) Imagen conteniendo cuatro texturas fractales de dimensión (de arriba abajo y de izquierda a derecha) 2.1, 2.3, 2.6 y 2.9. (b) Segmentación de (a) suponiendo cuatro clases. (c) Clasificación bayesiana de (a) usando la matriz de descriptores 4×4. (d) Resultados de procesar (c) con un filtro de moda.

Para ello se generaron 4 imágenes fractales puras (compuestas de una única dimensión fractal) para ser usadas como datos de entrenamiento, a partir de las cuales se extrajeron los datos estadísticos necesarios para la clasificación. Los resultados directos de dicha clasificación se muestran en la Fig. 5.4c. El porcentaje de píxeles correctamente clasificados es del 76%. Si a estos resultados aplicamos un postprocesado, consistente en el filtro de moda descrito en el capítulo anterior, este porcentaje se incrementa hasta un 94%, como puede apreciarse en la Fig. 5.4d. Se ve que a pesar de que las imágenes fractales son muy parecidas entre sí, los resultados del esquema de Gabor cuando se aplica a este tipo de imágenes son muy similares a los obtenidos con texturas cualesquiera. La mayor diferencia es que el problemas de píxeles mal clasificados en los bordes es menor en el caso de fractales. Esto podría explicarse por las diferencias entre las imágenes usadas entonces y las características particulares de los fractales brownianos. Los errores

de clasificación a lo largo de los límites se debían a la aplicación del mismo filtro de Gabor simultameamente a dos texturas muy diferentes. Dado que estas imágenes fractales son mucho más parecidas entre sí, este problema se hace menos patente. Por el contrario, los puntos internos se clasifican mejor en el caso de texturas de Brodatz, mientras que el método de generación de las imágenes fractales, por su naturaleza aleatoria, favorece la variabilidad dentro de la misma clase. Esto provoca mayores errores en la clasificación de puntos aislados.

5.3.2 Estimación de la dimensión fractal

Acabamos de mostrar que el esquema de Gabor es capaz de discriminar entre imágenes con diferentes dimensiones fractales. Por lo tanto, cabría pensar en la posibilidad de extraer algún parámetro a partir de la matriz de descriptores que se relacione con la dimensión fractal. El enfoque más sencillo y directo consiste en aprovechar el hecho de que un fractal browniano tiene un espectro que sigue un decaimiento exponencial, expresado en la Ec. (5.2). Este decaimiento debe reflejarse en la matriz de descriptores, dado que ésta no es más que un muestreo particular del dominio de Fourier. De hecho, Pentland [89] propuso un método basado en el espectro local de Fourier para estimar la dimensión fractal. Dado que la transformación de Gabor (Ec. (2.3)) no es sino un análisis de Fourier local [115] empleando una ventana gaussiana, es fácil encontrar la relación entre ambos métodos.

El método desarrollado se basa en el hecho de que, para fractales brownianos, si representamos en un gráfico con ejes logaritmicos las respuestas de los cuatro canales de frecuencia frente a sus frecuencias centrales, éstas deben caer a lo largo de una línea recta de acuerdo con la Ec. (5.2). Para comprobar esto, se aplicó el esquema de Gabor a cuatro imágenes (128×128) consistentes cada una de ellas en un fractal de una determinada dimensión. Las matrices de descriptores 4×4 se calcularon en 16×16 puntos a lo largo de estas imágenes, y se obtuvo una matriz promedio para cada una de ellas. Dado que nuestros fractales son isótropos, la matriz se redujo a un vector promediando en las orientaciones según Ec. (5.1), obteniendo los descriptores invariantes ante rotación descritos en la Sección 5.1. De esta forma obtenemos cuatro valores que constituyen un muestreo del módulo del espectro de Fourier (promediado en orientaciones) en las frecuencias $\frac{1}{4}, \frac{1}{8}, \frac{1}{16}$ y $\frac{1}{32}$ ciclos/píxel respectivamente.

Para verificar el decaimiento exponencial hemos dibujado estos valores con ambos ejes logarítmicos. De esta forma, la Ec. (5.2) se convierte en:

$$log(|H(f)|) = K - m \ \log \ f \ , \tag{5.3}$$

donde K es una constante y m el parámetro de decaimiento exponencial. A continuación, los cuatro valores del vector de descriptores se ajustan a esta recta teórica usando mínimos cuadrados. La bondad de dicho ajuste nos indicará si la hipótesis inicial fué correcta, es decir, si el decaimiento exponencial del espectro se refleja o no en la matriz de descriptores. Si el ajuste fuese bueno, la pendiente m nos permitiría estimar la dimensión fractal. La Fig. 5.5 muestra los resultados del ajuste para tres ejemplos correspondientes a dimensiones fractales 2.1, 2.3 y 2.6 respectivamente. En todos los casos el ajuste es bueno, con un parámetro de regresión linear r que varía entre 0.999 (D=2.1) y 0.97 (D=2.6).

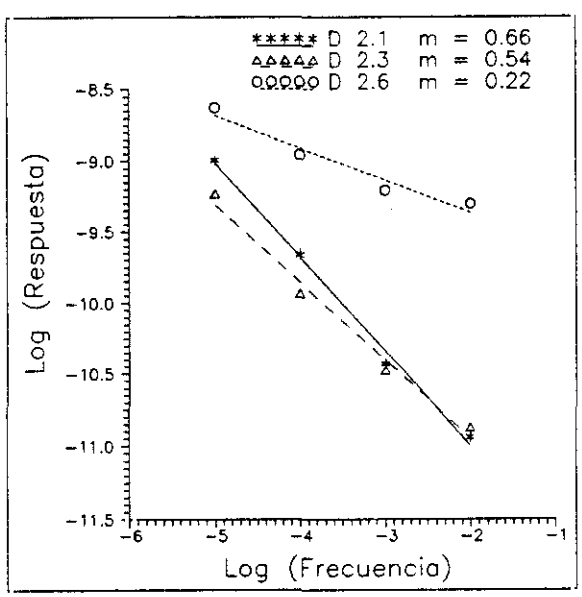


Figura 5.5: Respuesta de las células de Gabor (promediadas en orientaciones) frente a la frecuencia de sus canales (ejes logarítmicos), para tres imágenes fractales de dimensiones 2.1, 2.3 y 2.6.

Hemos observado que el ajuste empeora al aumentar la dimensión fractal. El parámetro r pasa de 0.999 (dimensión 2.1) a 0.97 (dimensión 2.6) para bajar hasta 0.6 para una dimensión de 2.9. Esto es probablemente debido a que tratar con fractales en imágenes discretas se hace más difícil para dimensiones fractales altas, por las razones antes aducidas. Basicamente, lo que ocurre es que las imágenes con una alta dimensión fractal son más ruidosas, y los problemas debidos al submuestreo (dado que el espectro es más ancho) pueden empezar a ser notorios. De hecho, dimensiones fractales altas van a estar siempre afectadas por problemas de

"aliasing". De cualquier forma el ajuste es suficientemente bueno si la dimensión fractal se mantiene por debajo de 2.7 ($r <= 0.94$).

Una vez que hemos comprobado que la idea original es correcta en la práctica, podemos usarla para establecer una calibración del esquema. Esta se llevó a cabo de la siguiente forma: se generaron diferentes imágenes fractales con dimensiones comprendidas entre 2.1 y 2.9, a las que se aplicó el método anteriormente descrito, obteniendose el parámetro m para cada caso. Los resultados se muestran en el gráfico de la Fig. 5.6, donde la pendiente m se representa en función de la dimensión fractal D. Se observa una clara relación lineal entre ambos parámetros, que coincide con las previsiones de la teoría [113].

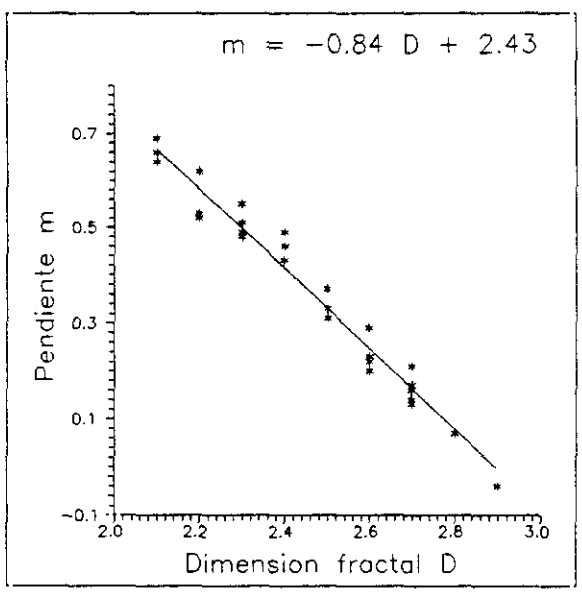


Figura 5.6: Gráfico de la pendiente m (extraida de la matriz de descriptores como muestra la Fig. 5.5) frente a la dimensión fractal D. A partir de esta figura calcularse la dimensión fractal de imágenes, usando la línea del ajuste como calibración del método.

Como resultado, este método puede aplicarse para medir la dimensión fractal de imágenes o texturas usando los descriptores de Gabor. La ecuación de la recta de ajuste de la Fig. 5.6 constituiría la calibración del esquema, a partir de la cual podemos obtener D una vez obtenida m a partir de la matriz de descriptores. Esta relación entre D y m viene dada por:

$$D = 2.88 - 1.19m \ , \tag{5.4}$$

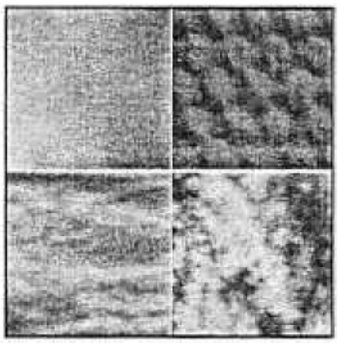
que no es más que el resultado de despejar D en la ecuación del ajuste de la gráfica de la Fig. 5.6.

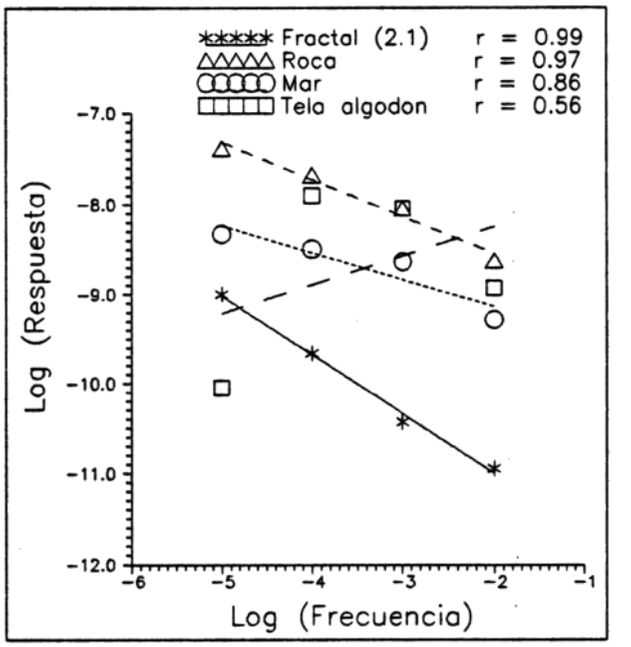
Se aprecia en el gráfico que para algunos valores de las abscisas (D) se presentan varios valores de m en las ordenadas. Esto es debido a que se generaron varias imágenes fractales para cada dimensión con objeto de comprobar la repetibilidad del método. Por lo tanto, la dispersión de los diferentes valores de m para una D dada proporciona una indicación del error. Como se puede comprobar directamente sobre la gráfica, estas variaciones de m dan un incertidumbre a la hora de conocer D del orden de $\pm$ 0.05.

5.3.3 Dimensión fractal de texturas naturales

Sería interesante estudiar si este método basado en la estimación de la dimensión fractal podría usarse también para clasificar cualquier tipo de textura. Sin embargo, como ya ha sido puesto de manifiesto por otros autores [90], hemos observado que la dimensión fractal por sí sóla no es un buen descriptor para la mayoría de las texturas incluidas en el álbum de Brodatz. La principal razón de esto es que la mayoría de esas texturas no son fractales, y no presentan un decaimiento exponencial. Con objeto de verificar este hecho, hemos comparado el ajuste a un decaimiento exponencial para un fractal puro (D=2.1) y otras tres texturas: roca [D2], mar y tela de algodón, que se muestran en la Fig. 5.7a. En la Fig. 5.7b se muestran el resultado de los ajustes correspondientes a los cuatro ejemplos escogidos (equivalente a la Fig. 5.5, sólo que ahora no consideramos únicamente fractales). Se aprecia que la textura roca, cuyo parecido visual con una imagen fractal es alto, muestra un ajuste mejor (r = 0.97) que el de las otras texturas. Por otra parte, no es posible ajustar los valores de la textura algodón a una línea recta (r = 0.55). Esta textura es artificial, presentando una alta periodicidad, por lo que no se parece a un fractal browniano. En estos ejemplos el parametro de la regresión lineal r se podía considerar un índice de "fractalidad" : cuanto más próximo a la unidad más similar a un fractal browniano es la textura considerada.

En conclusión, el esquema de Gabor es también aplicable a la clasificación de imágenes fractales de la misma forma que lo fué para texturas naturales. Además puede ser usado para estimar la dimensión fractal con una buena precisión. Sin embargo, el concepto de dimensión fractal sólo tiene sentido cuando estamos tratando con imágenes fractales o con texturas que se asemejen a ellos. El grado de fractalidad parece guardar una cierta relación con lo naturales que sean las texturas. Así, texturas naturales inanimadas (rocas, nubes, paisajes, líneas costeras) presentan





(a) (b)

Figura 5.7: (a) Tres texturas de Brodatz (tela de algodón, mar y roca) y un fractal (D=2.1). (b) Ajuste de las respuestas de las células de Gabor a las frecuencias de sus canales para las cuatro texturas mostradas en (a). El parámetro de error del ajuste podría considerarse un "indice fractal" de las imágenes.

un alto grado de fractalidad. Por el contrario, texturas biológicas con periodicidades y sobre todo aquellas artificiales (como el caso de la tela de algodón antes considerado) estan lejos de seguir una geometría fractal.

Parte III

Capítulo 6

Procesado primario de la información visual: caso foveal

En el esquema de Gabor presentado en los Capítulos 2 y 3 y en sus posteriores aplicaciones (Capítulos 4 y 5), usamos algunas de las características que aparecen en el sistema visual. El uso de canales localizados en el espacio de Fourier, cada uno llevando información de frecuencia y orientación determinada, su implementación en el dominio espacial, el empleo de funciones de Gabor para modelar los canales, etc, son todos ellos aspectos que nos recuerdan la base fisiológica de tal esquema. Sin embargo, es claro que hay aspectos de dicho esquema que no concuerdan con lo que sabemos del proceso visual. El uso de una red rectangular no es un buen modelo para el muestreo que se lleva a cabo en la fóvea, dada la disposición cuasihexagonal de los fotoreceptores. Más importante es el hecho, a menudo subestimado, de que aunque se tiende a ver a la fóvea como una región de características homogéneas, la densidad de fotoreceptores decae al alejarnos del eje óptico, cayendo aproximadamente un factor 4 para una excentricidad de sólo 1° [14]. Debido a esto, cualquier esquema que considere un muestreo uniforme será un pobre modelo incluso para la fóvea central (foveola). Otros factores no considerados en el esquema previamente propuesto, y que deberían ser tomados en cuenta en un modelo realista, son los efectos de la óptica del ojo (caracterizados por la función de transferencia de modulación o MTF), así como el hecho de que el muestreo se realize no por receptores puntuales ideales, sino por conos con una apertura finita.

En este capítulo desarrollamos un modelo más realista del procesado primario que se lleva a cabo en el sistema visual considerando los factores antes mencionados. En los últimos años han aparecido importantes contribuciones que permiten una cuantificación de tales efectos desde muy diversos campos: fisiología [10][13][14]

[30][29], psicofísica [15]–[18][31][32], óptica [6][116], etc. Estos serán los datos que usaremos en nuestro modelo, que estará restringido a la fóvea, entendiendo por tal el área central del campo visual que se extiende hasta una excentricidad de 0.5 mm, o lo que es lo mismo (según datos recientes de Curcio et al. [13][14] sobre el ojo humano) subtendiendo un angulo de $\pm$ 1.7°.

En este Capítulo iremos presentando sucesivamente las diversas etapas que componen nuestro modelo, empezando con el efecto de la óptica del ojo (6.1), el muestreo por la red de fotoreceptores (6.2), y el efecto de la apertura finita de los conos (6.3). No se ha entrado en detalles sobre las distintas interconexiones de los niveles intermedios desde la retina hasta el cortex (células ganglionares, cuerpo geniculado lateral, etc), sino que se han modelado los campos receptivos de células simples corticales usando funciones de Gabor (Sección 6.4). Finalmente en la Sección 6.5 presentamos algunas aplicaciones del modelo, reproduciendo algunos fenómenos observados experimentalmente en el sistema visual humano.

6.1 MTF

En un modelo simplificado como el nuestro la óptica del ojo puede ser resumida en la MTF del sistema. Datos sobre la MTF del ojo en todo el campo visual han sido suministrados recientemente por Navarro et al. [6] (a partir de métodos objetivos de doble paso). Un ajuste propuesto [6] para la MTF media del ojo humano a distintas excentricidades dependiendo sólo de unos pocos parámetros es :

$$MTF(u) = (1 - C)\exp(-Au) + C\exp(-Bu) \ , \tag{6.1}$$

donde u es la frecuencia en ciclos/grado y A, B y C dependen de la excentricidad ϵ (en grados) de la forma siguiente:

$C = c_1 - c_2\epsilon$
$A = a_1 \exp(a_2\epsilon)$
$B = b_1 \exp(b_2\epsilon)$

donde los valores de las constantes de ajuste a_1, a_2, b_1, b_2, c_1 y c_2 son 0.174, 0.039, 0.036, 0.017, 0.215 y 0.00224 respectivamente.

En el área considerada, la MTF apenas varía, como ya se había mostrado [116] tras mediciones directas en el centro de la fóvea y a un grado de excentricidad. La Fig. 6.1 muestra la MTF calculada según la Ec. (6.1) para $\epsilon = 0$ y ϵ=2°.

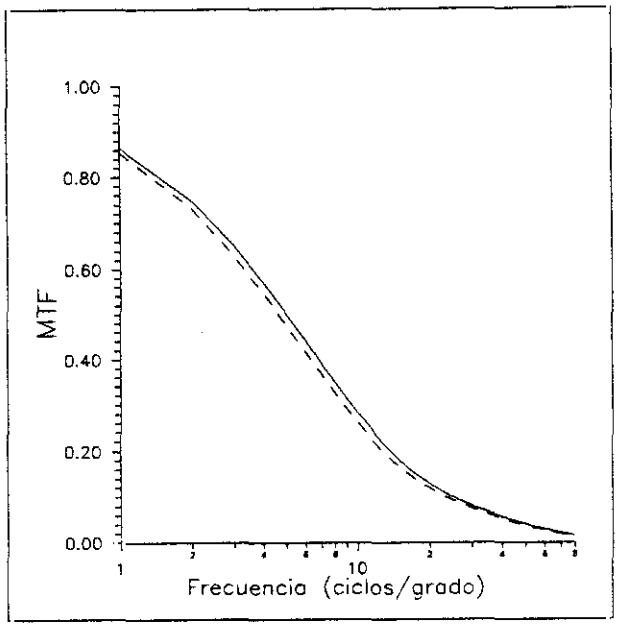


Figura 6.1: Comparación entre la MTF del ojo a 0° (línea continua) y a 2° (línea discontinua) de excentricidad [6].

Como se aprecia sólo hay una ligerísima caida de la curva. Podríamos pues considerar la MTF invariante en ese rango. Sin embargo, aunque las diferencias son escasas, hemos optado por una implementación más realista con un filtrado espacialmente variante, aplicando diferentes filtros para simular la MTF según nos vamos alejando del centro de la fóvea.

6.2 Distribución de fotoreceptores

En la región que estamos modelando, la fóvea, apenas existen bastones, así que no es ninguna simplificación el ocuparnos sólo de los conos. En los últimos años ha habido un intenso debate sobre qué nivel en la retina impone los límites del muestreo: conos o células ganglionares. La cuestión no está completamente, pero hay un acuerdo general [15] [31] de que, al *menos en lo que respecta a la fóvea*, son los conos los que limitan la resolución espacial. La situación cambia en la periferia, donde hay un 'pooling' (es decir, una suma de impulsos, promediando sobre una cierta extensión espacial) de las señales de los conos a las células ganglionares. En esa zona serían dichas células las que limitasen la resolución [11][31]. Otros autores muestran experimentos [32] que señalan hacia otro filtrado paso-bajo posterior a las células ganglionares para altas excentricidades. En este modelo suponemos que

en la zona considerada son los conos los que limitan la resolución espacial, lo cual es completamente realista en la fóvea.

Los estudios sobre la distribución de conos en la retina se remontan al trabajo clásico de Ostemberg [9], pero los datos más actuales y fiables de la fóvea humana provienen de Curcio et al. [14], y son los que usaremos en este modelo. Para ello hemos promediado sus datos sobre densidades de conos en el área 0–0.8mm para ambos meridianos, a un lado y otro del eje óptico (nuestro modelo será isotrópico). Los resultados son los que se muestran en la Tabla 6.1.

Tabla 6.1: Datos de Curcio et al. [14] sobre densidad de conos en la fóvea humana en función de la excentricidad

Excentricidad (grados)	0.0	0.35	0.69	1.04	1.38	1.73	2.07	2.41
Densidad (10^3 mm^{-2})	200	112	73	53	44	35	30	25

A partir de estos datos hemos comprobado que una curva del tipo :

$$D = \frac{D_0}{1 + k\epsilon} \ , \tag{6.2}$$

donde D_0 es la densidad a excentricidad nula, k una constante de ajuste y ϵ la excentricidad en grados, se ajusta muy bien a la densidad de conos D en el área considerada.

Si tomamos $D_0 = 235210$ conos/mm^2, y $k = 3.32$ grados^{-1}, el parámetro de error en el ajuste lineal de la inversa de la densidad con la excentricidad es 0.9994. La bondad de ese ajuste puede apreciarse en la Fig. 6.2.

Se ve que el ajuste sólo falla a excentricidades muy bajas. Esto no es un obstaculo importante, ya que en los datos experimentales [14] se aprecia que aunque para excentricidades superiores a unos 0.7° la densidad de conos es muy constante entre diversos sujetos, ésta varía grandemente en el centro de la fóvea. Curcio muestra sujetos con densidades máximas entre 100000 y 300000 conos/mm^2, por lo que nuestro ajuste puede considerarse realista. Una vez que tenemos una expresión que modela la distribución de la densidad de conos, podemos calcular su espaciado. Para un empaquetamiento hexagonal, dada una densidad D, la distancia entre conos d vendrá dada por :

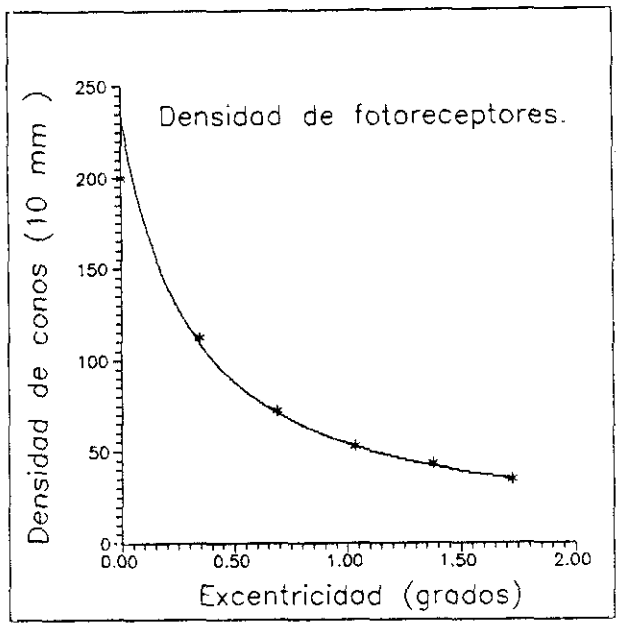


Figura 6.2: Ajuste (línea continua) de la Ec. (6.2) a los datos (*) de densidad de fotoreceptores de Curcio et al. [14].

$$d = \sqrt{\frac{2}{\sqrt{3}D}} \ . \tag{6.3}$$

De esta forma, substituyendo el valor de D (Ec. (6.2)) en Ec. (6.3) obtenemos:

$$d = \sqrt{\frac{2}{\sqrt{3}D}} = \sqrt{\frac{2(1+kr)}{\sqrt{3}D_0}} = d_0\sqrt{1+kr} \ , \tag{6.4}$$

donde d_0 es la distancia entre conos a excentricidad ϵ=0°.

En la Fig. 6.3 se muestra gráficamente el espaciado entre conos entre 0° y 1.7° de excentricidad, variando desde 2.2 (centro) hasta casi 6 micras ($\epsilon = 1.7°$).

Hasta ahora, la información que hemos considerado es únicamente unidimensional y nos dice cual tiene que ser el espaciado entre puntos de muestreo para una excentricidad determinada. Si nuestro problema fuera 1D, la construcción de una red de muestreo acorde a estos datos sería trivial. Pero el muestreo en el sistema visual es bidimensional, así que el problema ahora es diseñar una red de muestreo bidimensional acorde con los datos de espaciado de que disponemos.

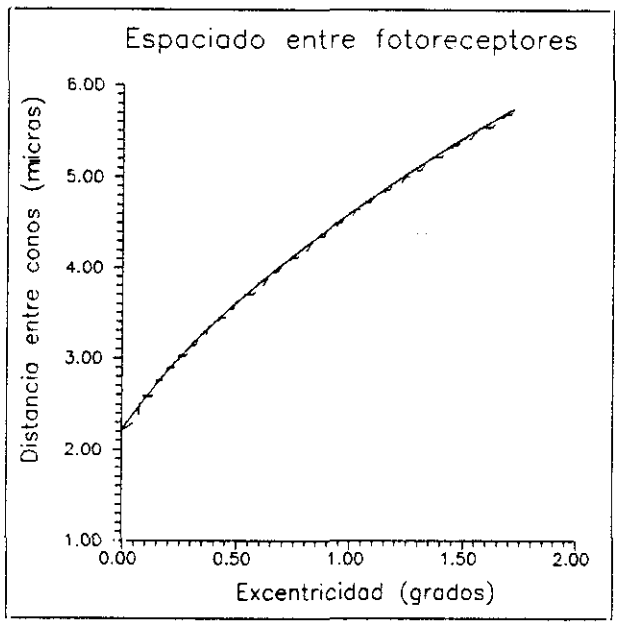
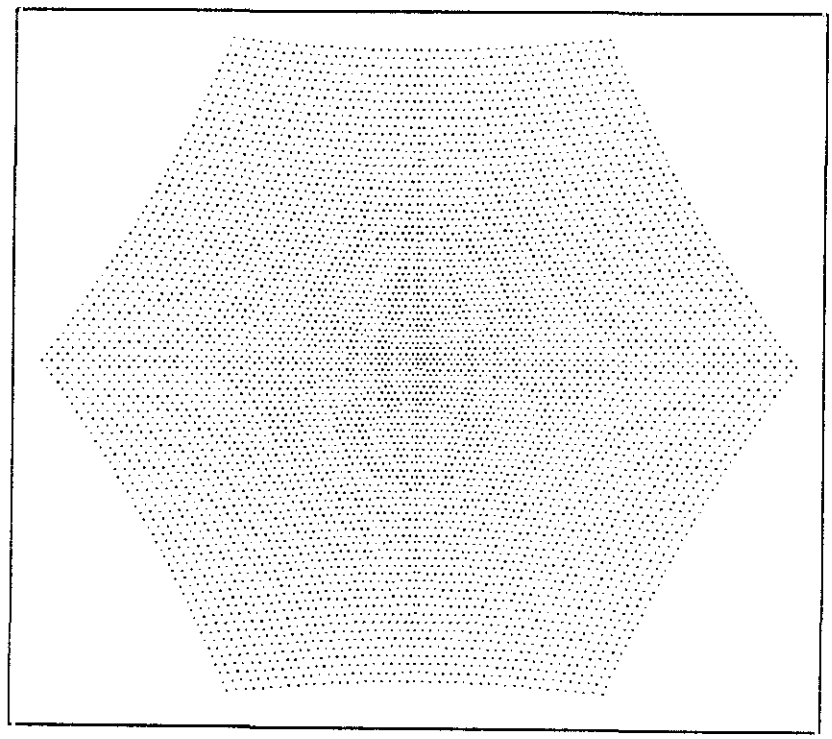


Figura 6.3: Variación del espaciado entre conos con la excentricidad según nuestro ajuste a los datos de Curcio et al. [14] (línea continua). La línea discontinua marca el espaciado de la red de muestreo usada.

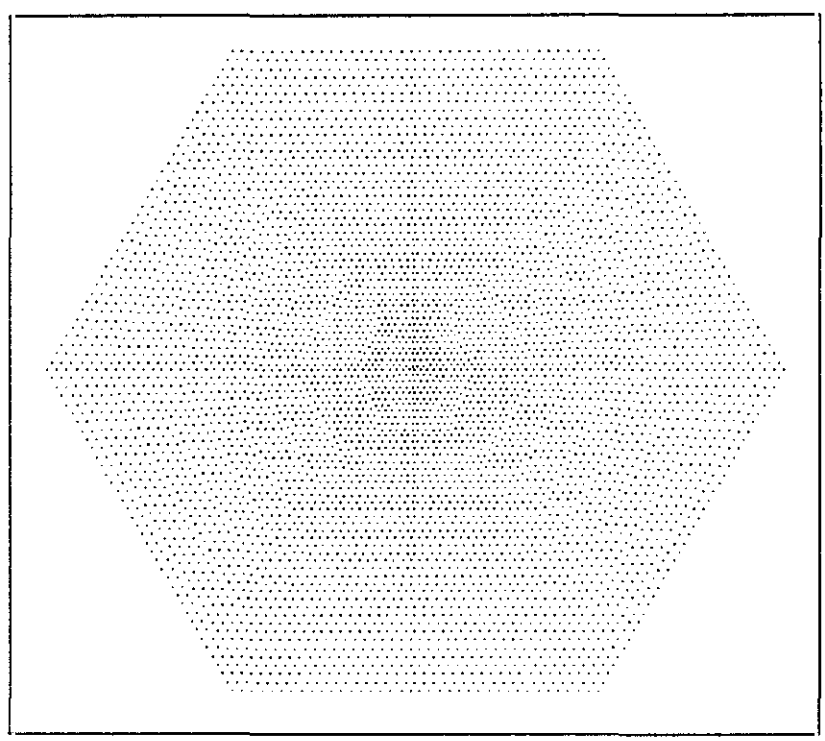
Hay dos casos extremos entre los que podríamos optar como aproximaciones al caso real:

- **Red regular deformada:** la idea más sencilla sería partir de una red hexagonal perfecta y deformarla radialmente para adaptarla al espaciado variante requerido por la Ec. (6.4). Las ventajas de este enfoque son claras, ya que una red de ese tipo es fácilmente direccionable. Dado un punto es posible decir inmediatamente cuales son sus vecinos (al igual que en una red rectangular a la derecha del punto (i, j) se encuentra el $(i, j + 1)$). Esto es importante porque a la hora de aplicar unos filtros, es fundamental conocer los vecinos de cada punto. Sin embargo, como se aprecia en ls Fig. 6.4a, es imposible deformar una red hexagonal de esa forma y mantener una estructura hexagonal entre los vecinos. Se puede observar en dicha figura cómo el espaciado es distinto según nos movamos tangencial o radialmente, incluso a pesar de que el área mostrada corresponde solamente a una excentricidad de $\pm$ 0.40° en la retina. Conforme aumenta la excentricidad este tipo de problemas se acentúa, lo que hace de este tipo de red una mala elección.

- **Red aleatoria:** lo ideal en cuanto a realismo sería la generación de una red totalmente aleatoria, por "crecimiento", con un programa de empaque-



(a)



(b)

Figura 6.4: Mosaicos de fotoreceptores obtenidos a partir de una deformación radial de una red hexagonal regular (a) y a partir de un crecimiento en espiral (b).

tamiento de esferas, cuyos radios vayan aumentando con su distancia al centro, de acuerdo con la Ec. (6.4) . Tal red ha sido propuesta por Ahumada y Poirson [117], aunque no en el marco de un modelo completo. Es claro que una red de estas características sería muy realista, pero presenta dos inconvenientes principales:

- Aunque el algoritmo es sencillo, el cálculo de un mosaico de las dimensiones requeridas en este modelo (del orden de unos 70000 conos) requeriría un alto coste computacional (aunque con la ventaja de que sólo tendría que calcularse una vez).
- Tal red no sería direccionable, por lo que obtener un mapa de los vecinos de cada punto para poder aplicar filtros supondría de nuevo una alta carga para el ordenador.

El modelo que hemos adoptado es el resultado de un compromiso entre estos dos casos extremos. Consiste en la red de muestreo cuyo aspecto puede apreciarse en la Fig. 6.4b, y que ha sido diseñada especificamente para este trabajo. La red se ha construído con un procedimiento de crecimiento en espiral. Una vez construida cada capa hexagonal, la siguiente se construye a la distancia indicada por el espaciado para esa excentricidad, pero ese espaciado se modifica ligeramente si es necesario, para que en cada lado de la capa hexagonal entren un número entero de puntos de muestreo con el mismo espaciado.

El espaciado es pues el mismo en una dirección u otra para una excentricidad dada. De esta forma vamos generando una red que mantiene unas caracteristicas fuertemente hexagonales, pero con un espaciado variable. En la Fig. 6.3 se compara el espaciado real [14] y el de esta red. Asímismo, en la Fig. 6.5 se muestra una comparación entre el número de conos dentro de una excentricidad dada para nuestra red y el esperado según nuestro ajuste de los datos de Curcio et al. [14] En ambos casos puede comprobarse que la red elegida para el modelo reproduce adecuadamente los datos experimentales sobre espaciado y densidad de fotoreceptores.

El mayor problema de esta red, en cuanto a su realismo, es que todavía es fuertemente regular. Un ejemplo de esto es que sigue manteniendo los ejes de una red hexagonal regular (en una red crecida aleatoriamente esos ejes se mantendrían unas pocas capas, cambiarían luego, etc). Este inconveniente es a su vez una ventaja cuando se trata de indexar la red para la aplicación posterior de filtros. Aunque la red no es inmediatamente direccionable, es mucho más fácil calcular los

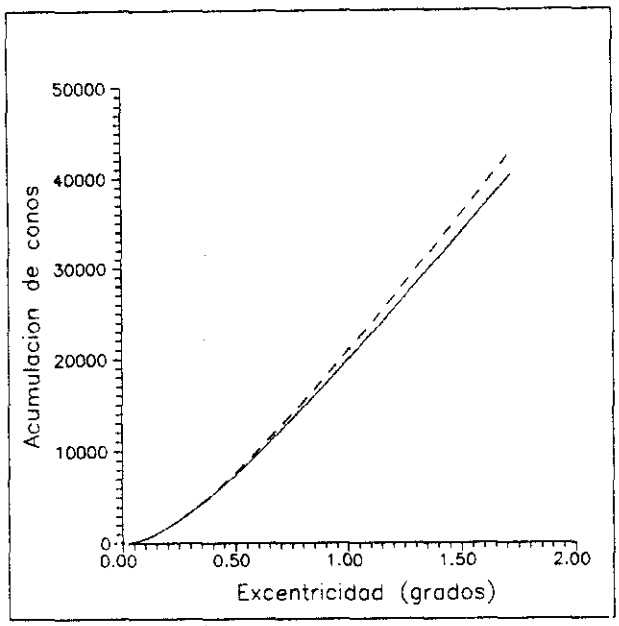


Figura 6.5: Número de conos dentro de un radio dado según el ajuste a los datos de Curcio et al. [14] (línea continua), y de acuerdo al mosaico implementado (línea discontinua)

vecinos para un punto dado en este tipo de red, que en una completamente aleatoria. Recientemente se ha establecido un debate sobre la importancia del desorden de los conos en el muestreo [22] y en particular sobre su efecto en la prevención del aliasing [19]–[21][118]. La idea original de que el desorden era crítico para prevenir el aliasing [19] no ha prevalecido [20][21]. Nosotros creemos que el desorden de los receptores tiene el efecto de disimular los fenómenos de aliasing, pero no de prevenirlos (el aliasing sigue ahí, pero su energía está más repartida, con un menor contraste y visibilidad). Los fenómenos debidos al aliasing deberán aparecer también en nuestro modelo aunque probablemente de forma ligeramente distinta (patrones de aliasing más regulares). Por otra parte, estudios estadísticos sobre la distribución de conos en fóvea tanto en monos [118][11] como en humanos [12], han mostrado que el mosaico foveal es más regular de lo que se pensó inicialmente [19]. Concretamente, es muy regular en el centro de la fovea (zona considerada en este modelo), disminuyendo dicha regularidad con la excentricidad. Sin embargo, una objección a nuestra red es que dichos estudios muestran que la regularidad en orientaciones no es tan alta, mientras que el mosaico escogido presenta una gran regularidad en orientaciones.

6.3 Apertura de los conos

Los posibles efectos de la apertura de los conos han sido discutidos en varias referencias [119][20], pero el consenso general (comprobado por nosotros en este modelo) es que el tamaño de los conos y su consiguiente efecto paso-bajo es prácticamente despreciable frente al filtrado que supone la MTF del sistema óptico del ojo. Para cuantificar esta apreciación se hizo una comparación entre dos muestreos, uno de ellos considerando la apertura finita de los conos, y el otro usando fotoreceptores puntuales. La relación señal/ruido entre la diferencia de ambas y una de ellas era del orden de 40 dB (para la imagen de la Fig. 6.10a). Sin embargo, en casos donde se ha eliminado el efecto de la óptica del ojo, la apertura de los conos sí que fué fundamental para reducir los posibles efectos de aliasing [20] (ver apartado 6.5.2). Por esta razón se ha tenido en cuenta la influencia de la apertura finita de los conos. Adoptaremos la simplificación de suponer que la eficacia de un cuanto de luz es la misma si cae dentro de un cierto radio crítico, es decir, se usará un un perfil cilíndrico. En la zona central de la fóvea el diametro de los conos es del orden de un 80 % de la distancia entre centros [14], habiendose usado esa relación.

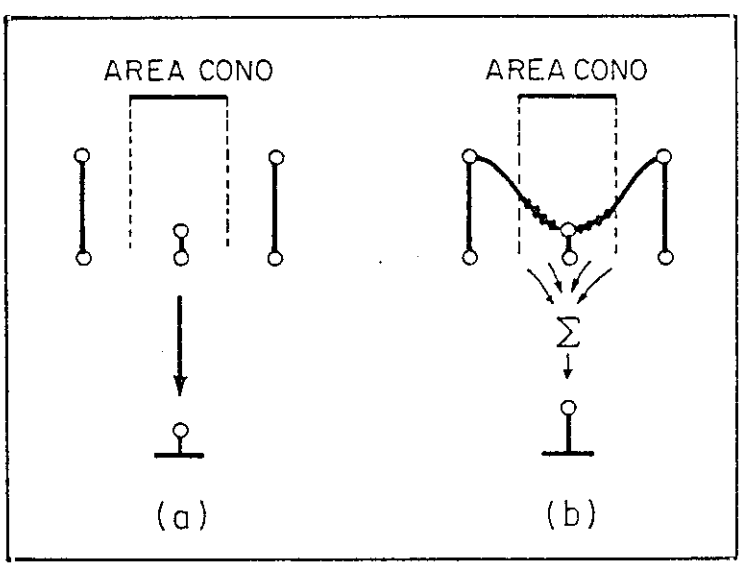


Figura 6.6: Situación en la que tenemos pocos puntos muestreados dentro del área de un cono (a). En este caso, es más exacto hacer un sobremuestreo de la zona y promediar los nuevos puntos así obtenidos (b).

Dado que nosotros partimos de una imagen no continua, sino muestreada previamente, tendremos en general pocos puntos muestreados en el área de un cono.

Para evitar esto, previamente hacemos un sobremuestreo de la zona (ver Fig. 6.6) alrededor de del cono, interpolando los datos conocidos para posteriormente sumar dichos valores sobre el área de dicho cono. Concretamente, para calcular la respuesta de cada cono se promediaron alrededor de 50 valores calculados dentro de la apertura considerada para dicho cono. Dichos valores se obtuvieron por interpolación a partir de las muestras de la imagen original.

6.4 Campos receptivos

Una vez muestreada espacialmente la información, hay un preprocesado previo en la retina antes de llegar al córtex [24]. El resultado final es una excitación de las células ganglionares, cuyos axones forman el nervio óptico que conduce la información hasta el cerebro. Los campos receptivos de dichas células se han modelado mediante diferencias de gaussianas [26], no habiendo evidencias de que sean selectivas a orientaciones. No obstante, este es un estadio intermedio, siendo más interesante modelar lo que ocurre finalmente en el área V1 de la corteza visual. Las células simples en este área son selectivas a orientaciones y frecuencias, habiendose modelado sus campos receptivos por funciones de Gabor [38][39][51][41].

La aplicación de las funciones de Gabor se realiza aquí de una forma muy similar al caso de un muestreo uniforme (Capítulo 3). En este caso, la implementación se hará en el dominio espacial, con las ventajas ya comentadas en el Capítulo 3, y la añadida de que la implementación en Fourier es ahora mucho más complicada debido al muestreo hexagonal espacialmente variante. Hay algoritmos especiales para llevar a cabo FFTs con una geometría hexagonal [8], pero en nuestro modelo no podemos usarlos, ya que precisan un muestreo uniforme. Por tanto, al no ser posible calcular transformadas de Fourier rápidas, la implementación en el dominio de Fourier se hace ordenes de magnitud más lenta que en el dominio espacial.

La implementación de esta parte del modelo presenta muchas analogías con la descrita en el Capítulo 3. Primeramente generamos los filtros de frecuencia más alta ($\simeq$ 38 ciclos/grado) y al igual que hicimos entonces, diseñamos máscaras de pequeño tamaño que den una respuesta en frecuencias lo más parecida a la deseada, pero esta vez con un muestreo hexagonal. El tamaño elegido para las máscaras es de 61 muestras, es decir un filtro hexagonal con 4 anillos ($1 + 6 + 12 + 18 + 24$). Un ejemplo de los filtros usados se muestra en la Fig. 6.7a, en ambos dominios, espacial y de Fourier. Son filtros paso-banda cuya función de transferencia es 1 para la frequencia y orientación a la que están sintonizados y

		-0.003	-0.016	-0.005	0.014	0.008		
	0.016	0.004	-0.029	-0.022	0.013	0.016		
-0.001	0.031	0.028	-0.032	-0.043	0.002	0.018		
-0.016	-0.016	0.036	0.059	-0.012	-0.052	-0.013	0.013	
0.005	-0.020	-0.039	0.019	0.075	0.019	-0.039	-0.020	0.005
0.013	-0.013	-0.052	-0.012	0.059	0.036	-0.016	-0.016	
	0.018	0.002	-0.043	-0.032	0.028	0.031	-0.001	
		0.016	0.013	-0.022	-0.029	0.004	0.016	
			0.008	0.014	-0.005	-0.016	-0.003	

(a)

		0.000	0.026	0.000		
	0.026	0.111	0.111	0.026		
0.000	0.111	0.178	0.111	0.000		
	0.026	0.111	0.111	0.026		
		0.000	0.026	0.000		

(b)

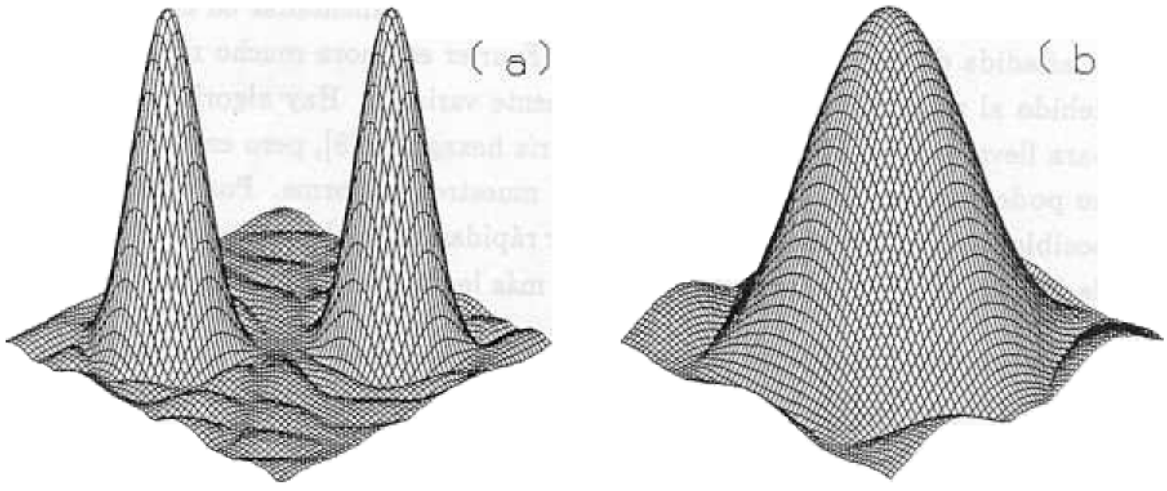


Figura 6.7: Máscaras hexagonales empleadas en la implementación espacial (arriba) y su respuesta en frecuencia (abajo), correspondientes a una función de Gabor (a) y al filtro paso-bajo (b).

que carecen de componente continua (suma de sus muestras = 0). Para aplicar los filtros a la imagen muestreada debemos conocer los vecinos de cada cono. Para ahorrar tiempo, estos vecinos se han calculado previamente para todos los puntos de la red y esa información se guarda en un fichero aparte, requiriendose cuando se filtra la imagen. Este fichero representaría de alguna manera las conexiones del "cableado" interno entre las células de nuestro modelo de retina. Por supuesto, al no ser una red hexagonal perfecta, los vecinos encontrados no estarán siempre formando hexágonos perfectos, pero eso es algo esperable, que ocurre tambien en la retina.

En cuanto a los canales de frecuencia más baja, utilizamos el método piramidal descrito en el Capítulo 3. Se reduce la imagen, previo filtrado con un filtro pasobajo (también hexagonal pero con 19 muestras, ver Fig. 6.7b), y se aplica de nuevo el *mismo conjunto de filtros*. La *razón de este es proceso es puramente práctica* (disminuir el coste computacional), dado que no hay evidencias para suponer que un proceso similar pueda suceder en la retina, donde lo que tenemos son campos receptivos de diversos tamaños. Hay documentados [27][25] dos tipos de células ganglionares en la retina de los primates: magnocelulares (M) y parvocelulares (P). Las células P son alrededor del 80 % del total y llevan información de alta frecuencia. En nuestro modelo, estas células P serían las que llevarían la información del canal de frecuencia más alta, mientras que las células M, que como su nombre indica tienen campos receptivo más grandes, llevarían información de frecuencias más bajas. Hasta ahora la evidencia fisiológica se limita a células M y P, pero no se descarta la existencia de otros tipos que, por su escasez, serían más difíciles de detectar. En nuestro modelo la cantidad de información en el canal de alta frecuencia es proporcional al número de muestras de la imagen original, N. Para el canal de frecuencia inmediatamente inferior es $N/4$, puesto que la imagen se reduce en un factor 4. Así se ve que la cantidad de "células ganglionares" necesarias para los distintos canales de frecuencia es proporcional a: N, $N/4$, $N/16$, etc. Consecuentemente, en nuestro modelo la información de altas frecuencias representa aproximadamente un 80 % del total, lo que se ajusta bien al porcentaje de células P, cuyas características comparten: campo receptivo pequeño e información de altas frecuencias.

Por otra parte, el hecho de aplicar los mismos filtros a una red cuyo espaciado no es constante, provoca el interesante resultado de que el mismo canal (mismo filtro) estará sintonizado a frecuencias diferentes al aumentar la excentricidad. En la Fig. 6.8 se ve cómo un filtro que está sintonizado a una frecuencia de unos 40 ciclos/grado en el centro de la fóvea, al irnos a una excentricidad de 1° ha

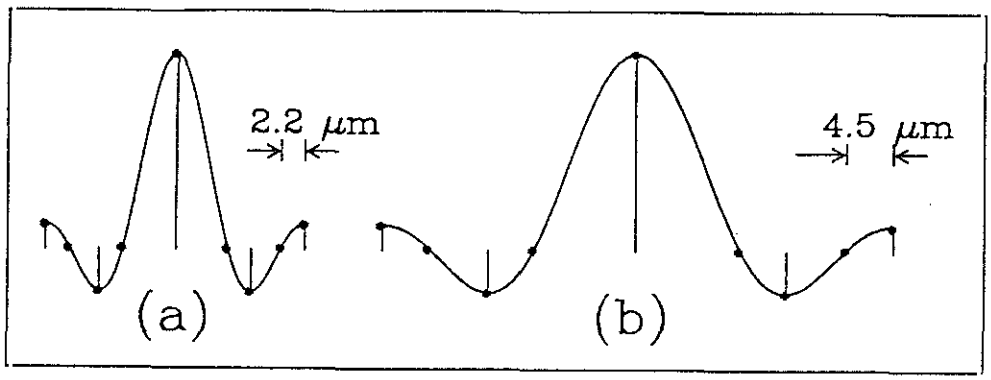


Figura 6.8: El mismo filtro, aplicado en el centro de la fóvea (a) y a una excentricidad de 1° (b), estará sintonizado a frecuencias de 38 y 20 ciclos/grado respectivamente debido a la variación en el espaciado de los conos.

desplazado su pico a unos 20 ciclos/grado. De esta forma se reproduce el conocido efecto de la disminución en la agudeza visual al aumentar la excentricidad, así como la diferente apreciación de las frecuencias espaciales al movernos hacia la periferia [41][121]. Esto se relaciona con el concepto de escala local propuesto por Rovamo et al. [122], que postula que el procesado espacial de las imágenes es homogeneo a traves del campo visual, excepto por un cambio de escala, o factor de magnificación cortical [123]. En nuestro caso el proceso es espacialmente invariante (mismo filtro) y la diferencia viene dada por un cambio de escala en el espaciado de las muestras. En la siguiente sección se muestran simulaciones de la función de sensibilidad al contraste (CSF) de nuestro modelo, que confirman estas expectativas.

Otro aspecto importante para hacer el modelo más realista es asignar distintos pesos a los diferentes canales de frecuencia. Si los pesos de todos los canales son iguales, la descomposión en canales se comportaría aproximadamente como un filtro paso-todo (una vez sumados los distintos canales), con las diferencias comentadas en la Sección 3.2. Esto, sumado al efecto paso-bajo de la MTF óptica haría que el resultado final global fuese un filtrado paso-bajo de la imagen. Ahora bien, es bien conocido el hecho de que ilusiones visuales tales como las bandas de Mach (Fig. 6.9a) son causados por una mala respuesta a bajas frecuencias en el sistema visual. En la Fig. 6.9b se muestra un corte de la imagen recuperada tras aplicar la MTF y la descomposición en canales a la Fig. 6.9a cuando los pesos son iguales a la unidad para todos los canales. Se observa un filtrado paso-bajo que no explicaría la aparición de las bandas de Mach. Sin embargo, si asignamos los pesos 5, 6, 5, 2.5 y 1 a los distintos canales desde la frecuencia más alta a la más baja y reconstruimos de nuevo la imagen (Fig. 6.9c), obtenemos el corte mostrado

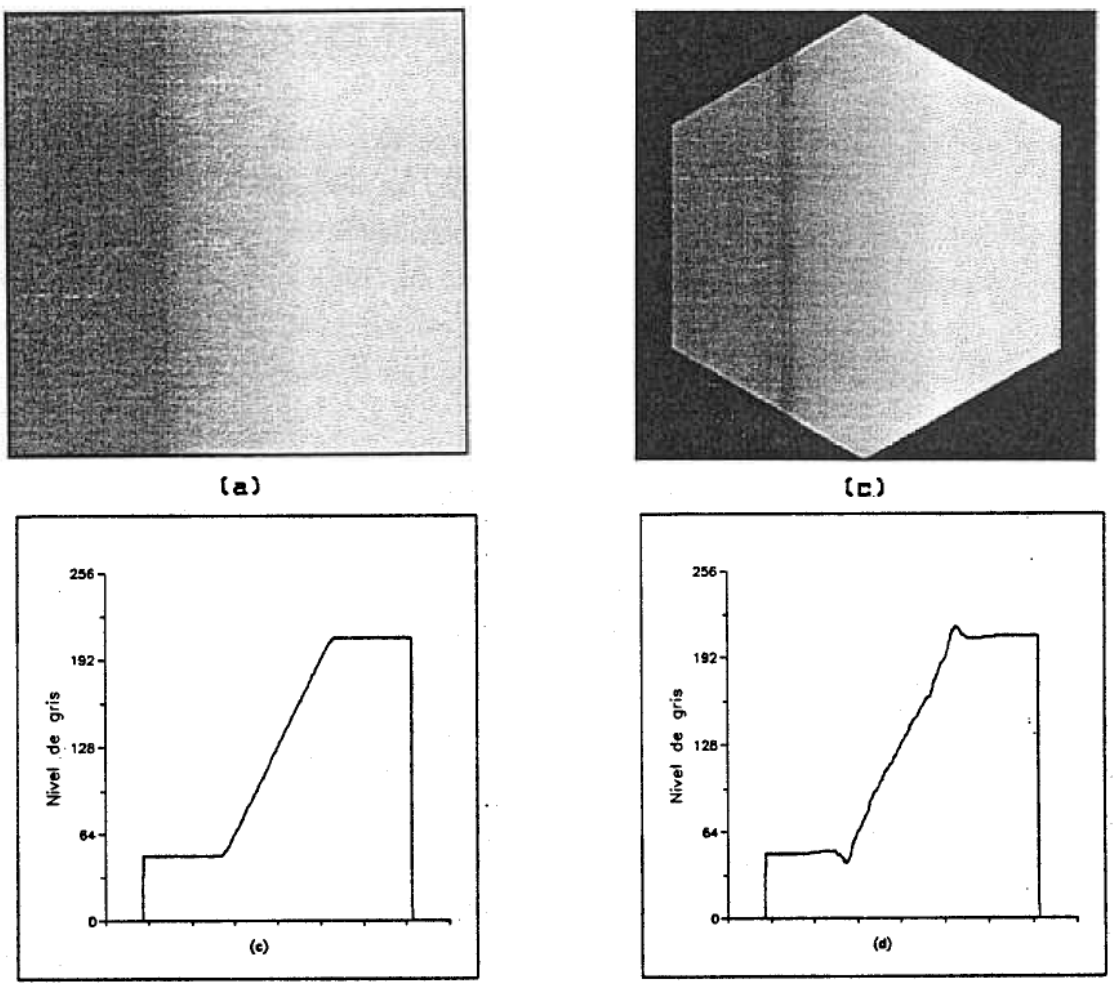


Figura 6.9: (a) Ilusión de bandas de Mach. (b) Perfil horizontal de (a). (c) Recuperación de (a) tras pasar por nuestra simulación del sistema visual. (d) Perfil de (c), mostrando cuantitativamente la ilusión visual.

en la Fig. 6.9d. En él se aprecia que lo que antes era sólo una ilusión, ahora se ha traducido a algo cuantificable. Los pesos usados son los que se calcularán posteriormente en la Sección 6.5.3, al ajustar la CSF del modelo. Una posibilidad mas de este esquema, que no hemos utilizado aquí, sería utilizar también pesos distintos para los diferentes canales de orientación, ya que el sistema visual posee una sensibilidad más desarrollada en los ejes horizontal y vertical que a 45° (efecto oblicuo).

6.5 Ejemplos de aplicación del modelo

En esta sección mostraremos primeramente el proceso completo de la aplicación de nuestro modelo a una imagen concreta, ilustrando las diferentes etapas. Posteriormente se estudian dos casos donde tratamos de reproducir los datos experimentales que se conocen acerca del fenómeno del aliasing en visión humana, y en segundo lugar, la capacidad de detectar frecuencias espaciales (cuantificada en la función de sensibilidad al contraste, CSF) y su variación con la excentricidad.

6.5.1 Aplicación del proceso completo

En este ejemplo partimos de la imagen que se muestra en la Fig. 6.10a, de 1024 × 1024 píxeles y 256 niveles de gris con un muestreo rectangular . El espaciado entre píxeles corresponde a la mínima distancia entre conos en la fóvea, es decir unas 2.2 micras según nuestro modelo (de esta forma se puede muestrear sin pérdida de información en la parte central del mosaico fóveal). Por lo tanto, la imagen completa subtiende aproximadamente $\pm$ 4° de campo visual. Esto sería equivalente a observar dicha imagen en un monitor de alta resolución de 20 pulgadas desde una distancia de unos 2 metros. Sobre esta imagen aplicamos la MTF, obteniendo el resultado mostrado en Fig. 6.10b. Por razones computacionales sólo se ha calculado el efecto de la MTF en la zona donde posteriormente se aplicarán los filtros de Gabor, lo que ocasiona la ventana circular que aparece en dicha figura. Desde el punto de vista de nuestro modelo ambas son todavía imágenes continuas, la primera correspondiente al mundo exterior, y la segunda la proyección de aquella sobre la retina. A continuación muestreamos la imagen con el mosaico de conos y posteriormente aplicamos los filtros de Gabor, descomponiendo la imagen en los diferentes canales, todo ello de acuerdo a lo explicado en la Sección anterior.

De acuerdo con nuestro modelo, el conjunto de señales obtenidas tras estos pasos es similar al que se recibe en el córtex visual ante un estímulo externo semejante. A partir de esta información, en las capas más profundas del cerebro, tienen lugar procesos de análisis e interpretación. Sin embargo, esto involucra procesados de orden superior que están fuera del objeto del modelo aquí presentado, dado que éste sólo cubre las etapas primarias (desde la retina hasta el córtex). Por lo tanto, para poder estudiar el comportamiento del sistema visual ante un determinado estímulo, la alternativa que nos queda es tratar de llevar a cabo la transformación inversa, reconstruir la imagen a partir de la suma de sus canales. A partir de las pérdidas de información en la imagen recuperada trataremos de predecir o justi-

(a)

(b)

Figura 6.10: Imagen original 1024×1024, correspondiente a 8° de campo visual (a), y su proyección sobre la retina (b), tras aplicarle la MTF.

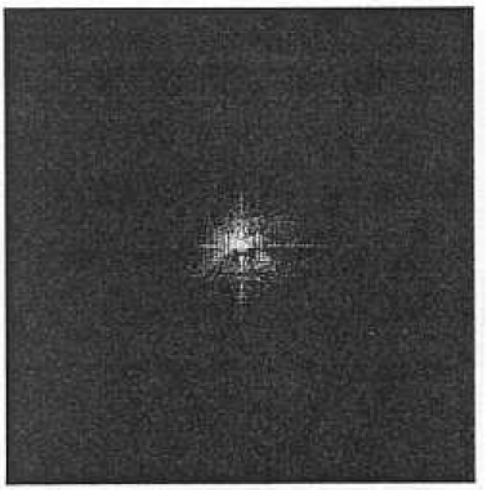


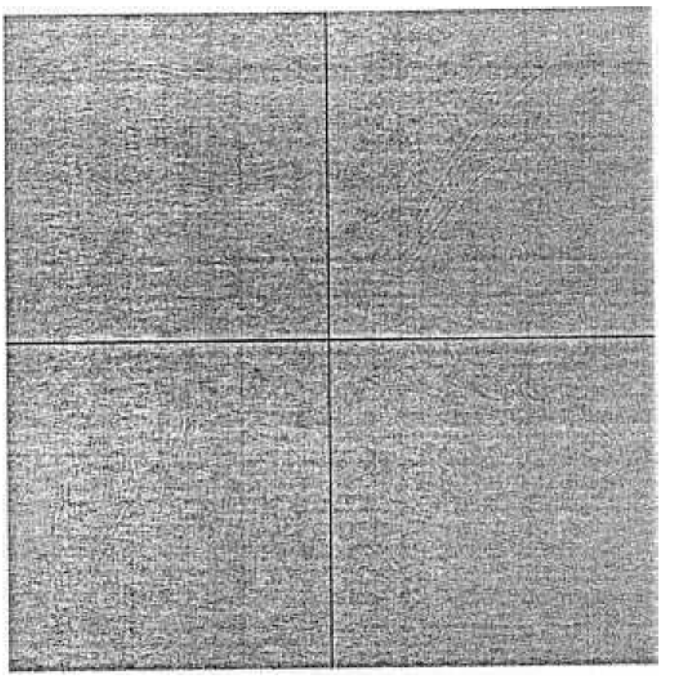
Figura 6.11: Muestreo de la Fig. 6.10b con la red de fotoreceptores.

ficar posibles limitaciones del sistema visual (de forma análoga a como se hizo en el apartado anterior con las bandas de Mach).

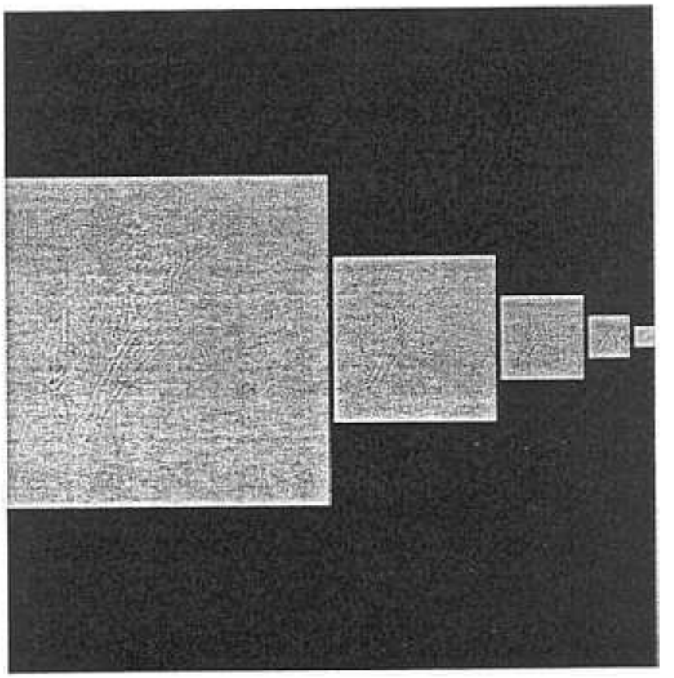
Sin embargo, tal proceso de reconstrucción es algo ficticio, que no ocurre realmente en el cerebro, por lo que no es extraño que nos encontremos con diversas dificultades. Por ejemplo, si se han introducido ganancias distintas para los diferentes canales, como se comentó en el apartado anterior, al intentar sumar de nuevo éstos, es muy posible que el rango de niveles de la imagen recuperada no se encuentre ya entre 0 y 256. Asímismo, la presentación de una imagen muestreada se observaría demasiado oscura (Fig. 6.11), por lo que para una mejor visualización, en algunos de los siguientes resultados se han rellenado los huecos correspondientes a lugares sin conos con una interpolación de los valores de los conos más próximos. Estos aspectos han de tenerse en cuenta cuando se examinen las imágenes que se muestran a continuación.

En la Fig. 6.11 se muestra el muestreo de la imagen de la retina (Fig. 6.10b) con el mosaico de la Fig. 6.4b, sobre el cual aplicamos los filtros de Gabor. En la Fig. 6.12a se muestra la descomposición de la imagen en sus cuatro canales de orientación para la frecuencia más alta.

Aplicando el proceso piramidal previamente descrito se obtiene una descomposición similar para cada canal de frecuencia. La recuperación de la imagen es análoga al proceso descrito en el Capítulo 3. Primeramente sumamos para cada



(a)



(b)

Figura 6.12: (a) Salidas de los cuatro canales de orientación de la frecuencia más alta para la Fig. 6.9a. (b) Suma de las cuatro orientaciones para cada canal de frecuencia y residuo de baja frecuencia.

frecuencia todos sus canales de orientación, llegando al resultado mostrado en Fig. 6.12b, que es algo muy similar a la pirámide laplaciana [59]. Sobre los canales así obtenidos aplicamos las ganancias (5,6,5,2.5,1) citadas en el apartado anterior. Finalmente los sumamos, previa interpolación de los de frecuencia más baja, para conseguir la imagen recuperada que se muestra en la Fig. 6.13, antes (a) y después (b) de interpolar las zonas entre conos. El tamaño de la imagen recuperada corresponde a la zona considerada en este modelo, unos 3° de diámetro. Se observa una mejora en cuanto al contraste respecto a la imagen con la MTF, ya las ganancias utilizadas suponen una potenciación de las altas frecuencias. De hecho, el aumento de contraste es mayor que el observado, ya que, al mostrar la recuperación, ha habido que reducir su rango dinámico. De esta forma, el filtrado que se lleva a cabo compensa el efecto paso-bajo de la MTF inicial. Este proceso recuerda a la aplicación de un filtro inverso para restaurar la degradación de una imagen (en este caso, la degradación por la óptica del ojo).

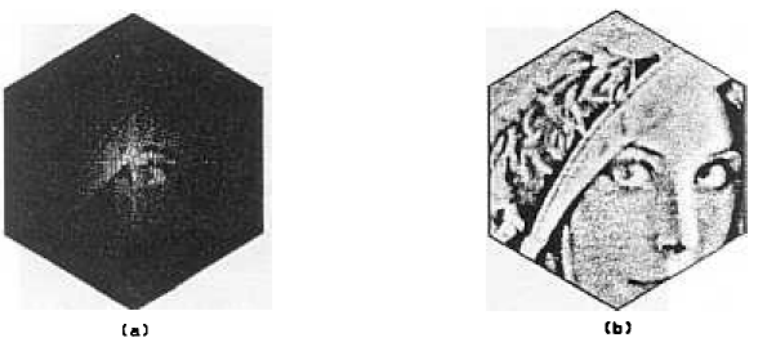


Figura 6.13: (a) Recuperación de la imagen (en los puntos donde se muestreó) obtenida sumando todos los canales de la Fig. 6.12b. (b) Interpolación de (a).

6.5.2 Aliasing

Uno de los efectos documentados recientemente [15][16][18] [32] ha sido la posibilidad de observar efectos de aliasing en el sistema visual, bien evitando la óptica del ojo por métodos interferométricos, o buscando en zonas donde la frecuencia de corte de la MTF sea mayor que la frecuencia de Nyquist de la red de muestreo. En la Fig. 6.14 se muestran algunos resultados cuando presentamos al sistema miras sinusoidales de alta frecuencia.

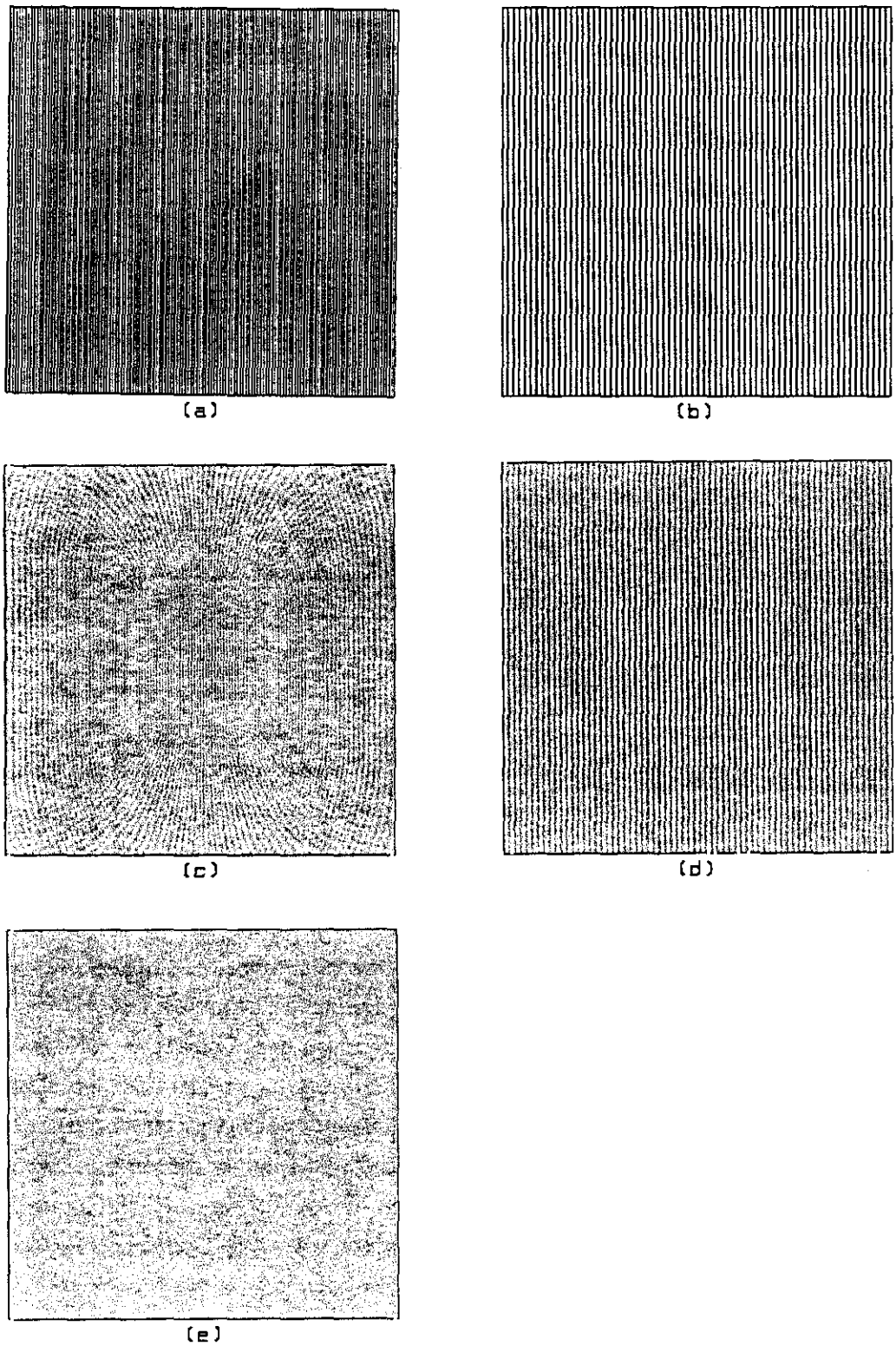
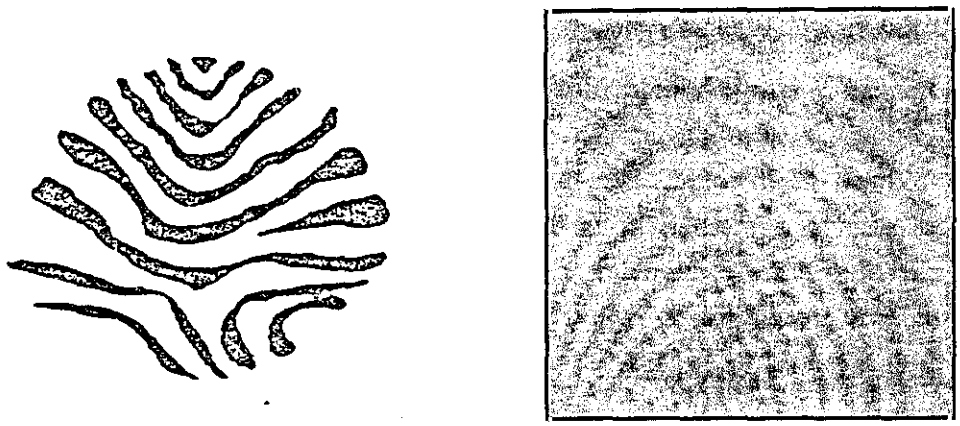


Figura 6.14: Miras sinusoidales de 38 (a) y 19 (b) ciclos/grado y su muestreo por el mosaico de fotoreceptores ((c) y (d) respectivamente). Si previamente se aplica la MTF, los efectos de aliasing en (c) se reducen notablemente (e).

El tamaño de las imágenes presentadas es de unos 4° de lado. En la parte superior se ven las miras sinusoidales usadas, cuyas frecuencias (desde el punto de vista de nuestro modelo) son 38 y 19 ciclos/grado (Fig. 6.14a y b respectivamente). Si suponemos que usamos un sistema interferométrico para proyectarlas sobre la retina, no sufren degradación debido a la MTF, y entonces, tras muestrearlas con nuestra red hexagonal, ofrecen el aspecto que se observa en la parte media (Fig. 6.14c y d). En este caso, dado que se trata de frecuencias muy altas a las que además no se ha aplicado previamente la MTF, el efecto de la apertura de los conos no es despreciable, y en consecuencia, ha sido tenido en cuenta. Se ve que la frecuencia más baja apenas presenta aliasing si exceptuamos algunos, poco notables, efectos a excentricidades altas, del orden de unos 2°, lo que se debe al mayor espaciado de los puntos de muestreo en ese área. Por el contrario, la frecuencia de 38 ciclos/grado, excepto en su parte más central, muestar numerosos fenómenos de aliasing. Para muestrear correctamente tal frecuencia se necesitaría un espaciado de fotoreceptores del orden de 3.5 micras. Como se mostró en la Fig. 6.3, tal espaciado es superado a una excentricidad de unos 0.5°, que coincide con aquella a partir de la cual el aliasing para dicha frecuencia se hace notorio en la Fig. 6.14c. Podemos fijarnos en varios aspectos distintos del aliasing. Por una parte, algunos de los patrones de Moiré que se observan (Fig. 6.15a) cuando se usan frecuencias muy altas (del orden de 80 ciclos/grado) son similares a las 'zebra patterns' (Fig. 6.15b) observados experimentalmente por Williams [15]. Estos patrones son diferentes para cada sujeto y dependen del mosaico concreto (como huellas dactilares). La diferencia es que nuestros patrones son mucho más regulares, debido a las características de nuestra red, que como ya se comentó es más regular que un mosaico real. En las diagonales se observa claramente que los patrones de aliasing de baja frecuencia se presentan en orientaciones diferentes a la original. Este fenómeno ha sido también descrito por Coletta y Williams [18] con el nombre de aliasing de orientación. Finalmente, en los laterales observamos la aparición de franjas de aliasing de frecuencia variante (no porque varie el estímulo, sino por la variación de la red de muestreo) que conservan la orientación del estímulo original. En conclusión, podemos ver que nuestro modelo reproduce algunos de los diversos fenómenos de aliasing descritos en la literatura, aunque los patrones son más regulares y localizados debido a la mayor regularidad de la red empleada.

Si consideramos el efecto de la MTF en el caso de la frecuencia más alta, (ahora la sinusoide sería un test presentado externamente), se obtiene el resultado mostrado en la parte inferior (e) de la Fig. 6.14. La percepción del aliasing desaparece en la parte central. De hecho, el posible aliasing sigue ahí, ya que



(a) (b)

Figura 6.15: (a) Reproducción de los patrones de aliasing observados por Williams [15]. (b) Patrones de Moiré que aparecen en nuestro modelo al introducir frecuencias de unos 80 ciclos/grado, evitando la MTF.

la MTF sólo reduce el contraste. Sin embargo, para una frecuencia tan alta, el contraste es tan bajo que el aliasing no se aprecia. Persiste sin embargo un cierto aliasing en la zona más exterior (a partir de unos 2°). Esto es debido a que la calidad óptica del ojo en esa área es virtualmente idéntica a la del centro de la fovea, mientras que la calidad de muestreo empeora notablemente. Esto indicaría que en principio sería posible observar aliasing en condiciones de visión normales (sin recurrir a métodos interferométricos), ya que 40 ciclos/grados no es una frecuencia que no pueda pasar a través de la óptica del ojo. Una de las razones por la que no nos encontramos con este tipo de fenómenos en la vida cotidiana (aunque sí se han inducido en experimentos "ad hoc", pero a excentricidades mayores [32]) es posiblemente porque los espectros de escenas naturales tienden a decaer exponencialmente siendo poca la energía a altas frecuencias que podría provocar aliasing. Además, el modelo considerado es estático y monocular, mientras que en la vida real, los rápidos movimientos del ojo y la visión binocular podrían contribuir a eliminar el aliasing.

6.5.3 Reproducción de la CSF

En este apartado vamos a cuantificar la respuesta de nuestro modelo a frecuencias espaciales puras, y su variación con la excentricidad. En el sistema visual, esta respuesta viene dada por la función de sensibilidad al contraste (CSF). El valor de la CSF para una frecuencia es el inverso del contraste umbral requerido para la detección de dicha frecuencia. Está claro que la capacidad de detectar si una frecuencia esta presente o no, vendrá dada por el número de células que se excitan y su nivel de excitación. Por debajo de una excitación mínima el cerebro descarta tal información como posible ruido y no apreciamos nada.

A partir de estas consideraciones hemos pensado en calcular la "CSF" de nuestro modelo, y ver si su forma, así como su variación con la excentricidad, se asemejan a los resultados experimentales. Para ello deberemos definir un parámetro de respuesta global o excitación del sistema visual a un estímulo exterior.

Para calcular la respuesta global del sistema a partir de la de sus canales, Quick [124] propuso un método, adoptado en varios modelos [42][125][126], consistente en una simple suma vectorial. De esta forma la respuesta del sistema sería la norma (no necesariamente euclídea) del vector compuesto por las respuestas de los diferentes canales:

$$R = (\sum_{f,\theta} |R_{f,\theta}|^Q)^{\frac{1}{Q}} \ , \tag{6.5}$$

donde $R_{f,\theta}$ es la repuesta de un canal, que puede expresarse de forma análoga a partir de todos los elementos que lo componen. De esta forma:

$$R_{f,\theta} = (\sum_{x,y,p} |R_{f,\theta,x,y,p}|^P)^{\frac{1}{P}} \ . \tag{6.6}$$

En nuestro caso vamos a utilizar una ligera variante de las expresiones anteriores. En primer lugar, se han usado valores de P y Q iguales a la unidad, aunque algunos autores se inclinan hacia valores mayores que potencian la respuesta de aquellos mecanismos que están mejor sintonizados con el estímulo original. Nuestra razón para hacer ésto es que, como se explicará posteriormente, nuestro procedimiento para calcular la CSF no es exactamente el análogo a un experimento psicofísico, ya que sería inviable por razones de tiempo de cálculo. En vez de eso usamos un proceso en esencia similar, pero para que ambos sean cuantitativamente equivalentes es preciso que la respuesta del sistema sea basicamente lineal, lo que

se consigue con $P = Q = 1$. En segundo lugar, dado que nuestro esquema cuenta con un diferente número de células N_f en cada canal de frecuencia, normalizamos previamente la respuesta de cada canal por N_f. Una justificación de esta normalización es que la sensibilidad de una célula es proporcional al área de su campo receptivo [25] y por lo tanto, inversamente proporcional a N_f en nuestro modelo. La normalización es una forma de dar más peso a las células de canales de baja frecuencia, con campos receptivos mayores, y por lo tanto, más fiables.

Por lo tanto, la expresión utilizada en el modelo para la respuesta o excitación global del sistema E, a partir de las respuestas individuales de todos sus elementos, $R_{f,\theta,x,y}$, es:

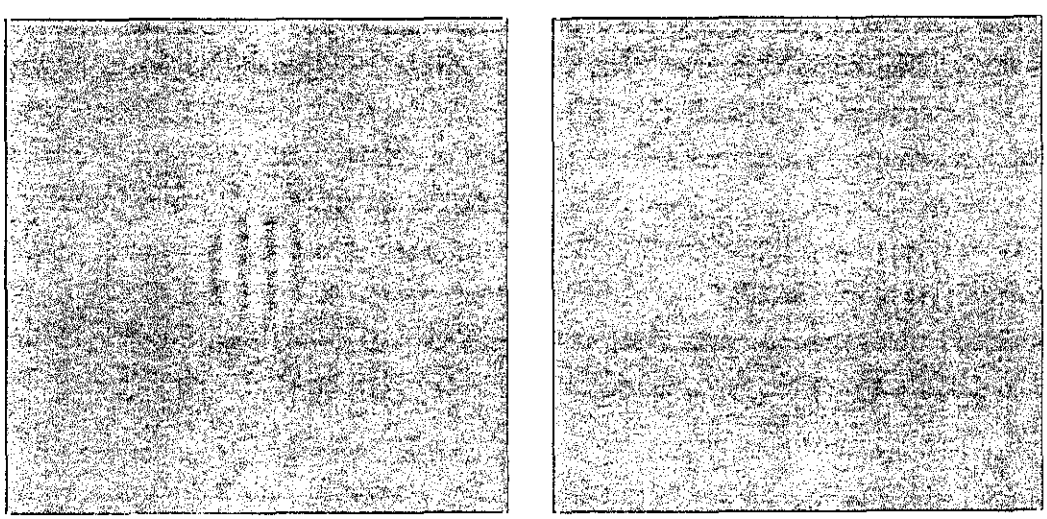
$$E = \sum_{f,\theta} |\frac{1}{N_f}(\sum_{x,y} |R_{f,\theta,x,y}|)| \ . \tag{6.7}$$

A partir de aquí, la forma correcta de proceder sería escoger una CSF típica y obtener a partir de ella el contraste umbral mínimo percibible y la frecuencia a la cual se obtiene tal umbral. Se generaría un estímulo con dichas características y se pasaría a través del modelo, calculandose la excitación producida a partir de la Ec. (6.7). Dicha excitación sería la mínima detectable por el sistema. A partir de ahí se presentarían estímulos de diferentes frecuencias y se iría reduciendo su contraste hasta que la excitación final cayera por debajo de la mínima. Entonces, según nuestro modelo, dicha frecuencia dejaría de ser observable y el inverso del contraste alcanzado constituiría la CSF. Sin embargo este método, que es análogo al empleado al medir la CSF del sistema visual, sería demasiado tedioso, pues al ordenador le cuesta mucho más tiempo simular el paso del estímulo por el sistema visual y decidir sobre su visibilidad que al observador humano decidir si lo ve o no.

Hemos escogido otro método más rápido que creemos esencialmente equivalente. Lo que hacemos es generar un conjunto de estímulos con frecuencias comprendidas entre 1.2 y 38 ciclos/grado y con un contraste común a todas ellas, c=0.3. A continuación aplicamos nuestro modelo a cada uno de estos estímulos, calculandose en todos los casos nuestro parámetro de excitación E, según Ec. (6.7). Los ganancias para cada canal han sido las mismas que se han empleado anteriormente (empezando por el canal de frecuencia más alta) 5, 6, 5, 2.5 y 1. Es decir, en este caso, en vez de ver qué contrastes nos dan la excitación mínima, tenemos un único contraste, calculandose las diferentes excitaciones que produce para las diferentes frecuencias. Estos valores serán los que usaremos para dibujar nuestra CSF. Es claro que si una frecuencia da una excitación alta podremos bajar

su contraste mucho sin que la excitación baje por debajo del nivel de ruido (lo que indica un contraste umbral bajo, y consecuentemente una CSF alta). Por el contrario, si una frecuencia provoca una excitación pequeña, a poco que bajemos su contraste llegaremos al mínimo percibible, con lo cual dicha frecuencia tendrá un contraste umbral alto y una CSF baja. Se ve claramente la relación directa entre la CSF como inversa del contraste umbral (psicofísica) y la CSF como excitación (fisiológica). Esta relación, así como la básica linealidad de nuestro modelo justifican que consideremos ambos métodos equivalentes para la estimación de la CSF.

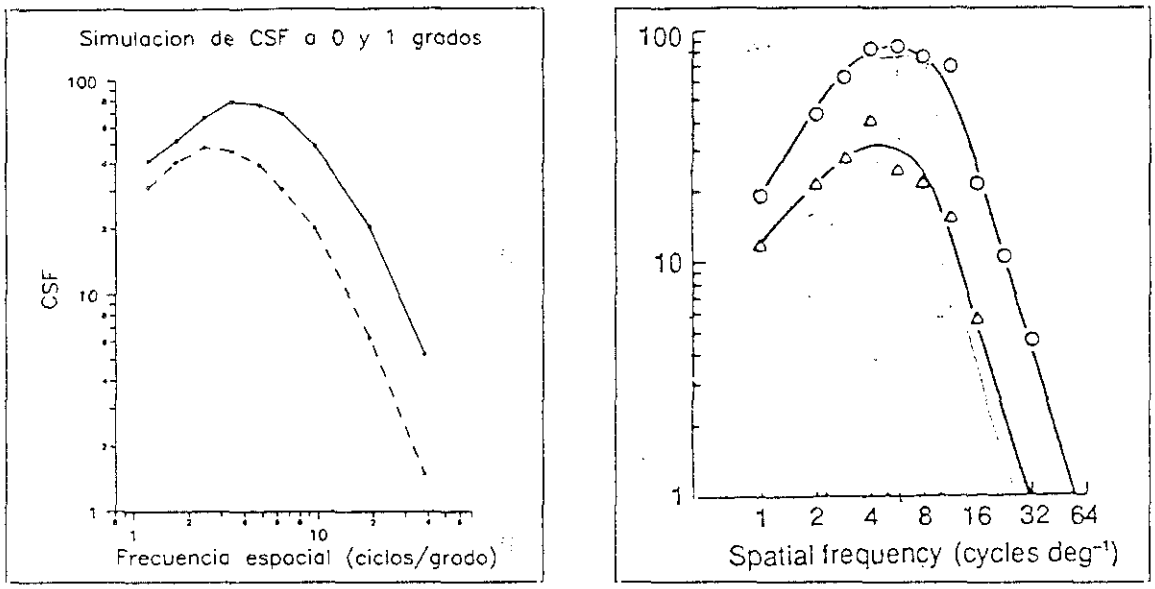
Primeramente calcularemos la CSF en el centro de la fóvea. Para ello trabajaremos con estímulos consistentes en una frecuencia espacial pura con 1° de diámetro. En la Fig. 6.16a se muestra uno de estos estímulos, correspondiente a una frecuencia de 4.8 ciclos/grado. La orientación de las franjas es vertical y se ha mantenido constante en todos los estímulos presentados. Así pues, cuando a partir de ahora hablemos de la CSF, nos estaremos refiriendo a un corte de la CSF bidimensional. Sin embargo nuestro modelo incorpora posibilidades direccionales por lo que podría calcularse la CSF bidimensional.



(a) (b)

Figura 6.16: Estímulos circulares de 1° de diámetro usados en la determinación de la CSF del modelo: (a) frecuencia de 4.8 ciclos/grado en el centro de la fóvea; (b) frecuencia de 9.6 ciclos/grado a una excentricidad de 1°.

En la Fig. 6.17a se presenta una CSF (línea continua), calculada de la forma antes explicada, en coordenadas logarítmicas. En el eje de ordenadas, en vez de representar directamente los valores de nuestro parámetro E (que es arbitrario), éstos se han escalado de forma que su máximo coincidiera con el de una CSF típica.



(a) (b)

Figura 6.17: (a) CSFs obtenidas con nuestro modelo para ϵ=0° (línea continua) y ϵ=1° (línea discontinua). (b) CSFs obtenidas experimentalmente por Rovamo et al. [122] a ϵ=0° (arriba) y ϵ=1.5° (abajo).

Se observa que la curva presenta el aspecto característico de una CSF real, con un máximo entre 4 y 6 ciclos/grado, y una caída hacia ambos lados, siendo ésta más acusada en las frecuencias altas.

Podría argumentarse en este punto que dado que hemos dispuesto como parámetros libres de las ganancias de cada canal, no ha sido muy difícil hacer que los resultados se ajusten a una CSF. Sin embargo la elección de las ganacias de los canales (púramente empírica y no exhaustiva por otra parte) sólo afecta al ajuste fino de la posición del máximo de la curva. Se puede comprobar que el uso de otras distintas, o incluso la eliminación de todas ellas (=1 para todos los canales) no altera la forma funcional de la CSF encontrada, con un pico en frecuencias medias y caidas a ambos lados. En segundo lugar, y como prueba más convincente de que el resultado anterior no sólo es producto de la elección de las ganancias, a continuación estudiaremos la evolución de la CSF al alejarnos del centro de la fóvea, manteniendo las mismas ganancias usadas en el caso anterior. De esta forma se insiste en la idea de mantener la invarianza espacial del proceso, al mantener constantes dichas ganancias con la excentricidad. Se repitió el proceso anterior, pero esta vez con diferentes estímulos situados a una excentricidad de 1° en el meridiano horizontal, como el que se muestra en la Fig. 6.16b corre-

spondiene a una frecuencia de 9.6 ciclos/grado. En la Fig. 6.17a se muestran los resultados obtenidos (línea discontinua), con la misma escala que la usada en el caso de excentricidad nula. En la Fig. 6.17b se presentan para su comparación una reproducción de resultados experimentales del trabajo de Rovamo et al. [122], correspondientes a CSFs medidas a 0 y 1.5° usando estimulos de tamaño similar al nuestro. Se observa la gran analogía entre ambas figuras, mostrando un similar factor de decaimiento. Asímismo, la simulación con nuestro modelo reproduce el desplazamiento del máximo de la CSF hacia frecuencias más bajas al aumentar la excentricidad. También se reproduce la corvergencia de ambas curvas hacia frecuencias bajas, y su caída, casi en paralelo para altas frecuencias.

Dado que la CSF caracteriza la visión espacial globalmente, el hecho de que nuestro modelo la reproduzca con un buen ajuste, sugiere que éste captura las características fundamentales del procesado primario de la información en el sistema visual humano.

Capítulo 7

Aprendizaje de campos receptivos en el córtex [1]

En los capítulos precedentes supusimos que ciertas células de la corteza visual presentan campos receptivos similares a funciones de Gabor. Consecuentemente, hemos usado este tipo de funciones como base de modelos del sistema visual (Caps. 2 y 6) y de diversas aplicaciones (Caps. 3, 4 y 5). Al hacer esto, nos hemos basado en los distintos trabajos [38][39][51][41] que justifican tal hipótesis.

Por otra parte, como se comentó en la introducción, en el córtex visual se puede encontrar una clara organización en lo que se conocen como módulos corticales [33]-[35]. Dichos módulos agrupan células que, respondiendo a distintas frecuencias y orientaciones, tienen sus campo receptivo en la misma zona del campo visual. Diversos estudios han mostrado que en un principio las conexiones entre el cuerpo geniculado lateral (CGL) y el córtex muestran una mayor difusión [127], mostrando éste último una menor organización. Sólo durante un proceso posterior, culminado después del nacimiento, aparece la organización típica de los adultos. Estas evidencias anatómicas se ven refrendadas por estudios psicofísicos [128] que muestran que la habilidad de distinguir entre orientaciones es muy pobre en recien nacidos, lo que indica que los campos receptivos correspondientes no están completamente formados.

Si la red de conexiones finalmente establecida en el cerebro tuviese que estar codificada genéticamente, la cantidad de información necesaria seria enorme. Consiguientemente, deben existir procesos que den lugar a la autoorganización que podemos apreciar en el córtex. En este Capítulo proponemos un posible mecan-

[1] Los resultados aquí presentados son el resultado de una colaboración con A.J. Ahumada, del **NASA Ames Research Center**, iniciada durante una estancia de 2 meses en 1991-1992.

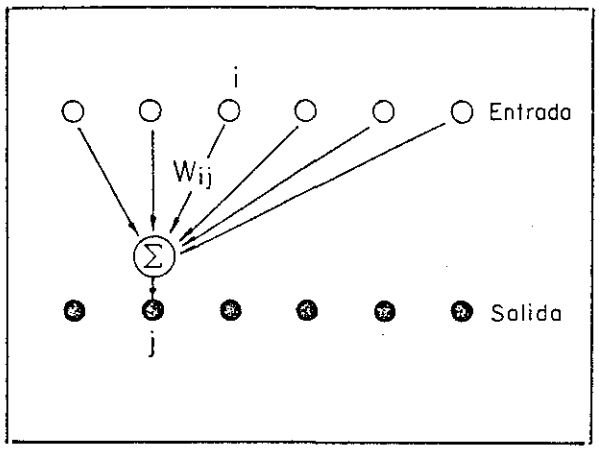


Figura 7.1: Red neuronal lineal de un solo nivel. Cada elemento del nivel de salida (j) es la combinación lineal de la entrada (i), pesada con unos coeficientes W_{ij}

ismo para explicar cómo podrían formarse durante el desarrollo campos receptivos similares a los de las células simples de la corteza visual, organizandose en "columnas" que comparten la misma posición espacial. El modelo está basado en una red neuronal, y no precisa de hipótesis ni restricciones previas sobre la organización deseada o la forma de los campos receptivos a obtener.

Antes de entrar en materia, dado que se trata de un concepto nuevo en esta presentación, daremos a continuación una muy breve introducción a las redes neuronales. Una revisión del tema puede encontrarse en [69], mientras que en [70] se da una una introducción a su teoría matématica. Una red neuronal es un ensamblado de muchas unidades sencillas, operando en paralelo, y con una alta conectividad. El término neuronal indica sólamente el origen de la inspiración de este tipo de estructuras, ya que en general estas redes no trabajan con modelos de neuronas reales. Nuestro caso es particularmente simple (ver Fig. 7.1), ya que las operaciones realizadas van a ser todas lineales, y sólo contaremos con dos niveles: la entrada y la salida. De esta forma, cada respuesta en el nivel de salida y_i será combinación lineal de las entradas x_j, pesadas con unos ciertos coeficientes $w_{j,i}$:

$$y_i = \sum_j W_{ji} \cdot x_i \quad \text{ó, en notación matricial,} \quad \mathbf{y} = \mathbf{W} \cdot \mathbf{x} \,, (7.1)$$

donde $\mathbf{x}$ es el vector de entrada, $\mathbf{y}$ el de salida, y $\mathbf{W}$ la matriz de los pesos o interconexiones de la red.

Por otra parte, la otra característica esencial de una red neuronal es su capacidad para aprender. Debemos pues dotar a nuestra red de un algoritmo de

aprendizaje, capaz de modificar los pesos progresivamente durante una fase de entrenamiento, hasta que se alcance el equilibrio. A partir de ese momento los pesos permanecerán estables y , si el entrenamiento ha sido satisfactorio, la respuesta de la red para una entrada determinada será la deseada. Lo que entendemos por respuesta deseada dependerá de la aplicación que estemos pensando para nuestra red. Podría ser que las distintas células de la salida se activaran dependiendo del tipo de entrada (clasificador), o reconstruyeran alguna imagen (previamente "aprendida") en base a una entrada incompleta (memoria), etc.

7.1 Antecedentes y planteamiento

Siguiendo el trabajo de von der Malsburg [129] diversos autores han propuesto que el desarrollo de células selectivas a frecuencias y orientaciones dadas, con campos receptivos aproximadamente lineales, podría explicarse mediante principios asociativos hebbianos [75][76]. El reciente modelo de Sanger [78] basado en estos supuestos resulta especialmente interesante a causa de la elegancia matemática de su resultado, originalmente encontrado por Oja [77]. Todos estos modelos se engloban en lo que se conoce como algoritmos de aprendizaje no supervisado para redes neuronales. En los aprendizajes supervisados partimos de un conjunto de entradas de entrenamiento, así como de las respuestas que idealmente esperaríamos de ellas, variando los pesos de la red en función de lo que se parece la respuesta obtenida a la deseada. Por el contrario, en un aprendizaje no supervisado no hay un profesor, es decir, no hay una actuación desde el exterior que diga cómo y cuáles deberían ser las salidas a una entrada dada. La red debe descubrir por sí misma regularidades, correlaciones o clases en los datos de entrada y codificar estas características en sus salidas. De esta forma, la red debe lograr un cierto grado de autoorganización.

Con este tipo de algoritmos una red neuronal puede aprender a reconocer algunas caraterísticas significativas de las imágenes que le presentamos. En particular, el modelo de Sanger [78] permite desarrollar "columnas" de campos receptivos que comparten una misma posición en el campo visual. Algunos de estos campos receptivos aprendidos muestran parecidos con detectores de barras y bordes, con una orientación específica, estando localizados en el dominio de Fourier. Este tipo de campos receptivos presenta numerosas similitudes con los detectados en células simples del córtex visual primario [39][41]. Dado que este modelo usa técnicas de aprendizaje hebbiano, a partir de ahora lo denominaremos algoritmo hebbiano generalizado (AHG).

El propósito de partida del algoritmo consiste en reducir el número de muestras de la señal. Para ello se dispone de una red con un número de salidas inferior al de entradas, tratando de encontrar los pesos que minimizen la pérdida de información. La solución es una red que lleve a cabo un análisis de componentes principales, lo que en teoría de comunicación se conoce como la transformada de Karhunen-Loeve. Partiendo de N entradas y M salidas ($M <= N$), se trata de encontrar M vectores (los pesos de la red) que definan un subespacio que comprenda el máximo posible de la varianza presente en los datos de entrada. En general dichos vectores serán los M vectores propios de la matriz de covarianza ($N \times N$) de la entrada con los mayores autovalores y hacia ellos converge la red propuesta por Sanger. Según la teoría [78], dichos vectores propios se ordenarán por orden decreciente de sus autovalores. Sin embargo, dado que muchos vectores propios están degenerados, *columnas con una diferente posición espacial en el campo visual no tienen por qué* desarrollar los mismos campos receptivos en el mismo orden. Asímismo se pueden desarrollar combinaciones lineales de vectores propios degenerados con diferentes *coeficientes en las diferentes posiciones espaciales. Todo esto provocará la falta de* invarianza ante translación de las columnas de campos receptivos así generadas.

Por otra parte, Maloney & Ahumada [71] propusieron un algoritmo para calibrar un *sencillo sistema visual lineal. La idea era reconstruir una imagen a partir* de un muestreo donde no se conoce exactamente la posición de los fotoreceptores. Dada una red de muestreo distorsionada, el algoritmo calcula unos pesos $w_{i,j}$ para la red que corrigen la imagen muestreada irregularmente, de forma que la salida de la red es la misma que si hubiesemos usado una red regular desde el principio. La red neuronal "aprendía" la posición de los receptores de muestreo y compensaba los efectos de sus posibles desplazamientos de la posición regular. Dicho de otra forma, lo que el algoritmo hace es construir en las distintas posiciones de la "retina" campos receptivos equivalentes a una función delta compensando los efectos de las posibles distorsiones de la red de muestreo. De ahora en adelante, llamaremos a este procedimiento algoritmo de invarianza de translación (AIT). El AIT en sí mismo sólo intenta que todos los campos receptivos sean equivalentes; la razón por la que todos terminan siendo funciones delta es que uno de ellos es forzado a serlo, al haberse conectado su salida a una única entrada. A partir de esto, es claro que el AIT puede ser usado para entrenar la red neuronal de forma que ésta propague un campo receptivo arbitrario. La respuesta de la red en una determinada posición será la que hubieramos obtenido de haber muestreado la imagen regularmente y haber aplicado el campo receptivo elegido a ese muestreo regular. En particular podemos usar el AIT para generar un nivel de células, todas ellas con el mismo campo receptivo, sintonizado a una orientación y frecuencia

dadas.

La mayor limitación de este método es que para que el AIT propague un determinado campo receptivo (sea una función delta, de Gabor, o cualquier otra) tenemos que fijar el perfil deseado para al menos una salida de la red. Eso equivaldría a conferir un status especial a una determinada célula (la que tiene su campo receptivo prefijado, sin posibilidad de modificación durante el proceso de aprendizaje), lo que no parece realista desde el punto de vista biológico.

Con objeto de obtener un modelo más realista, en este Capítulo proponemos una red neuronal que combinará ambos algoritmos (AIT y AHG), eliminando así las objeciones que se plantean a ambos al actuar por separado. Esto se hizo esperando que fuesen compatibles y que el AHG desarrollara una "columna" de campos receptivos similares a los encontrados en la corteza visual. El AIT por su parte se encargaría de asegurar que campos receptivos similares aparecieran en el mismo nivel en columnas con diferentes posiciones en el campo visual, corrigiendo además los posibles problemas derivados de las irregularidades en la red de muestreo.

7.2 Modelo del sistema visual

Trabajaremos con un modelo muy simple de sistema visual, cuya representación esquemática puede verse en la Fig. 7.2. El nivel de entrada consistirá en una red de fotoreceptores, irregular o no, que muestreará las imágenes del exterior. Las imágenes de entrenamiento usadas en este trabajo son series finitas de Fourier compuestas de un limitado número de frecuencias espaciales con amplitudes aleatorias (con una distribución gaussiana de dichos numeros aleatorios). Después de generar estas imágenes aplicamos un filtro paso-bajo gaussiano de desviación estándar $\sigma = 2$. El uso de un número limitado de frecuencias espaciales y el filtrado paso-bajo posterior pueden justificarse como un modelo simple de las características espectrales de las imágenes naturales (con poca energía en las altas frecuencias) y el desenfoque producido por la óptica del sistema visual respectivamente.

Se usará un modelo sencillo, con conexiones lineales entre las entradas (muestras de las imágenes) y las salidas de la red neuronal. El tamaño de red de muestreo será de N fotoreceptores (aunque se presentarán resultados del caso bidimensional, toda la argumentación se llevará a cabo para el caso unidimensional, en atención a la claridad de exposisión). La salida consistirá en varios niveles (N_{RF}) de N células cada uno. Cada nivel está compuesto de campos receptivos de un determinado

tipo. Por lo tanto, para cada posición de salida tendremos N_{RF} diferentes campos receptivos. Esta organización se puede apreciar gráficamente en la Fig. 7.2.

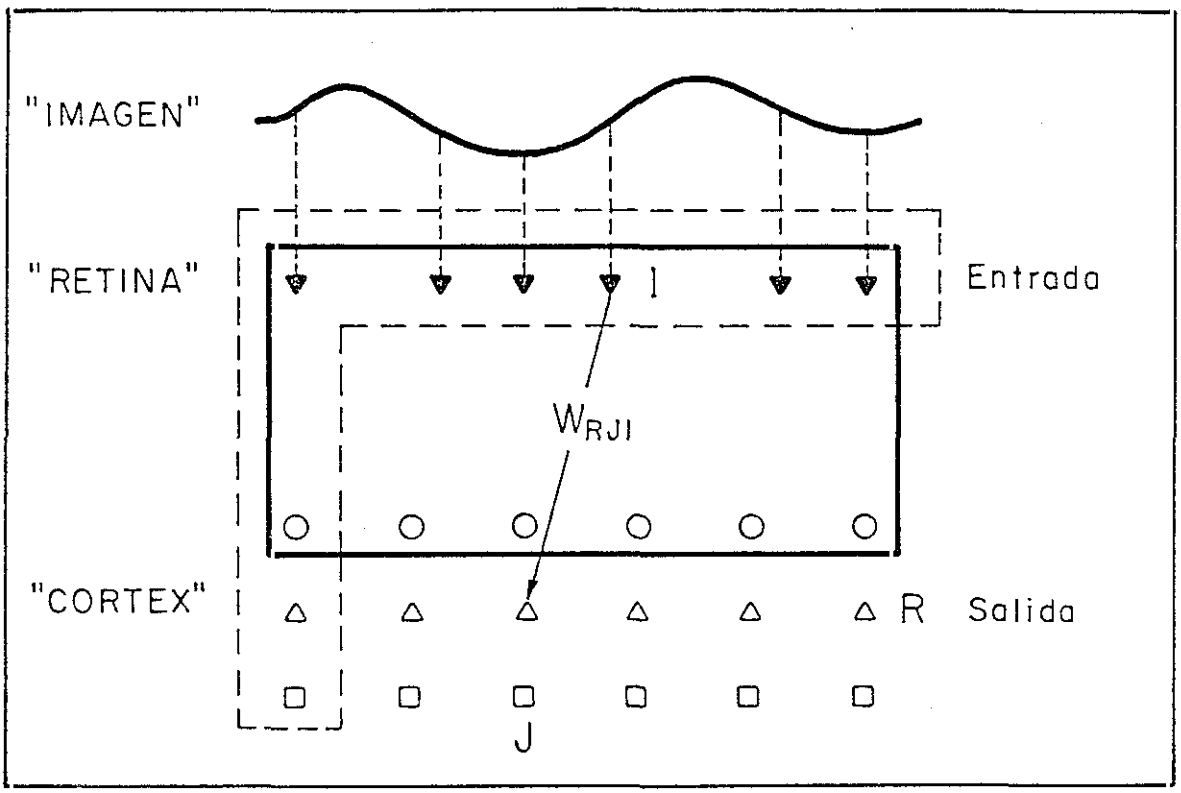


Figura 7.2: Modelo simple del sistema visual basado en una red neuronal. La imagen es muestreada por los N fotoreceptores, en una red irregular, que se encuentran linealmente conectados con el córtex. Moviendonos verticalmente en nuestro "córtex" tenemos columnas de N_{rf} campos receptivos que comparten la misma posición en el campo visual. La línea contínua marca el dominio de aplicación del AIT y la discontínua la del AHG.

De esta forma, los pesos conectando las entradas (x_i), y las salidas $(y_{r,j})$ tendrán tres subíndices, de manera que:

$$y_{r,j} = \sum_i W_{r,j,i} \cdot x_i \ , \tag{7.2}$$

donde r indica cada uno de los N_{RF} posibles campos receptivos para una posición dada; i y j marcan la posición espacial de entrada y de salida respectivamente (de 0 a $N-1$).

En particular, si fijamos una "célula" especificando tanto el tipo de campo receptivo (r_0) como su posición espacial j_0, el conjunto de pesos que comparten

dichas etiquetas, $W_{r_0,j_0,[0...N-1]}$, es lo que denominamos campo receptivo de dicha célula.

7.3 Procedimiento de entrenamiento de la red neuronal

En este modelo, como se apuntó anteriormente, se aplica el AIT independientemente para cada tipo de campo receptivo (r fijo) modificando los pesos $W_{r,[...],[...]}$. Igualmente el AHG operará independientemente en cada posición espacial de salida (j fija) actuando sobre los coeficientes $W_{[...],j,[...]}$. El dominio de aplicación de cada algoritmo, así como sus entradas y salidas, se muestra gráficamente en la Fig. 7.2. Primero describiremos someramente cómo trabajan ambos algoritmos. Para una explicación más detallada ver [71]–[74] (AIT) ó [78][70][77] (AGH).

7.3.1 La componente del AHG

La regla de aprendizaje del AGH de Sanger incrementa los pesos de la red neuronal en cada iteración según la ecuación:

$$\Delta W_{r,j,i} = \gamma y_{r,j}(x_i - \sum_{k<=r} W_{k,r,i}\, y_{k,r}) \ , \tag{7.3}$$

donde γ es el parámetro de aprendizaje del proceso. Este incremento ΔW se aplica sucesivamente a todas las posiciones de salida, $j = 0 \ldots N-1$.

Para el primer nivel de la columna de campos receptivos ($r = 0$), los pesos se incrementan en función de la correlación entre las entradas y las salidas. De esta forma, las entradas más frecuentes provocarán las salidas más altas en ese nivel. Cuando los pesos van convergiendo hacia sus valores definitivos, es como si las componentes de la entrada ya capturadas por los niveles inferiores fueran restadas a dicha entrada, no siendo "vistos" por los niveles superiores. De esta forma, dichos niveles deben capturar características diferentes a las ya obtenidas por niveles inferiores. La consecuencia de esto es que los campos receptivos generados al final del proceso dentro de una misma columna (misma posición espacial) son ortogonales entre sí.

7.3.2 La componente del AIT

EL AIT intenta ajustar lo que ve después de un "movimiento del ojo" con lo que esperaría encontrar. Esta imagen "esperada" se construye a partir de la última imagen muestreada antes de la translación. Conseguirá este ajuste sólo si es capaz de aprender la posición de los fotoreceptores, lo que equivale a calibrar el sistema. Para conseguir esto, por cada movimiento simulado del ojo, el sistema calcula la imagen esperada y la compara con la real. Esto da un término de error que se usa para corregir los pesos empleando un algoritmo de Widrow-Hoff modificado. El algoritmo de Widrow-Hoff [130][131] intenta que las futuras salidas sean más parecidas a las deseadas modificando los pesos con un incremento del tipo:

$$\Delta W_{r,j,i} = \lambda(y'_{r,j} - y_{r,j})x_i \ , \tag{7.4}$$

donde $y'_{r,out} - y_{r,out}$ es la diferencia entre la salida deseada (usualmente proporcionada externamente en los casos de aprendizaje supervisado) y la salida real, siendo λ el parámetro de aprendizaje. En el AIT, sin embargo, $y'_{r,out}$ no es proporcionada exteriormente, sino que es la imagen que el sistema espera encontrarse tras el movimiento del ojo, calculada (con una translación y posterior muestreo) a partir de la anterior.

7.3.3 El proceso global

El procedimiento empleado para el entrenamiento de nuestra red neuronal es el siguiente: una imagen de entrenamiento es generada, presentada al sistema y muestreada por la red de fotoreceptores. Estas muestras serán los datos de entrada para tanto el AIT como el AHG. El AIT se aplica en primer lugar. Dado que tenemos en nuestro modelo diferentes tipos de campos receptivos que queremos propagar, el AIT se aplica independientemente a cada nivel r (ver Fig. 7.2) de acuerdo con Ec. (7.4). Tras esto, el AHG se aplica a cada una de las posiciones de salida de la red j, modificando sus pesos según Ec. (7.3). Despues de aplicar ambos algoritmos se produce un "movimiento del ojo" (translación) arbitrario, la imagen se muestrea otra vez, y se aplican los algoritmos de nuevo. Esto se repite un número determinado de veces (usualmente de 10 a 50), para posteriormente generar una nueva imagen y empezar de nuevo.

El parámetro de aprendizaje del AIT es constante ($\lambda \simeq 0.5$), mientras que el del AHG va siendo reducido progresivamente de acuerdo con la siguiente regla:

cada cierto número de iteraciones (normalmente cada 10 imágenes) promediamos el error en el AHG, y si este error va disminuyendo, suponemos que el parámetro de aprendizaje es todavía adecuado, manteniendolo. En caso contrario se reduce por un factor constante ($\simeq 0.75$).

El proceso continúa hasta que los errores de ambos algoritmos bajan por debajo de un cierto límite previamente definido. Para los resultados aquí presentados estos límites han sido de 0.001 (lo que significa que para entonces los coeficientes de la red se modificaban solo un 0.1%). El número de imágenes de entrenamiento usadas para llegar a ese límite era habitualmente de unas 500–1000.

7.4 Resultados

7.4.1 Muestreo regular

Empezaremos entrenando la red usando un muestreo regular de los fotoreceptores, esperando que tal simplificación nos proporcione algunos resultados fáciles de interpretar. En este caso particular deberíamos obtener un conjunto de "columnas" de campos receptivos en cada posición espacial de la "retina". Estas columnas deberían mostrar una invarianza ante translación, de forma que los mismos campos receptivos aparezcan en todas las columnas y ordenados en los mismos niveles.

Esto es lo que sucede, como se aprecia en la Fig. 7.3, donde se muestran algunos casos cuando el tamaño de la red de muestreo es de 7×7 (49 fotoreceptores). En un primer caso (7.3a) se establecieron tres niveles de campos receptivos, cinco en el caso 7.3b, y finalmente siete en los resultados mostrados en la Fig. 7.3c. Solamente se muestra una columna de los perfiles obtenidos en cada caso para una posición dada, ya que todas las demás son similares como esperabamos. Los perfiles obtenidos corresponden a un filtro paso-bajo (que aparece siempre y en primer lugar) y a una serie de filtros paso-banda con distintas orientaciones. El primero es siempre el filtro paso-bajo, ya que la teoría nos indica que el primer autovector corresponderá al patrón de entrada más habitual, y éste es siempre la componente continua de la entrada. En cuanto a los restantes filtros, el orden de aparición no es único, debido a que todos ellos corresponden a autovectores degenerados. Su orden de aparición es pues arbitrario, dado que las imágenes de entrenamiento son generadas aleatoriamente. Casi todos los perfiles muestran orientaciones verticales, horizontales, y a 45 grados. La única excepción es el cuarto filtro de la Fig. 7.3b, y aun éste parece ser una combinación lineal de uno vertical y

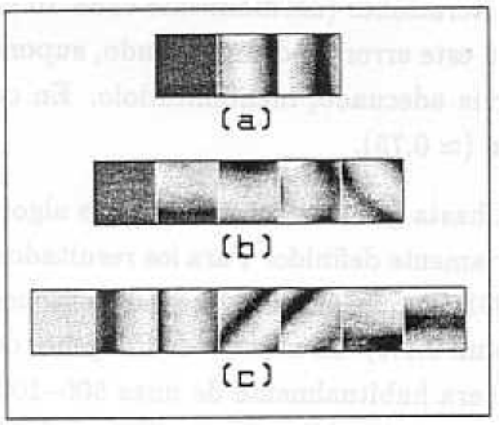


Figura 7.3: Campos receptivos generados por la red con un mosaico de 7×7 fotoreceptores y 3 (a), 5 (b) ó 7 (c) niveles de campos receptivos (N_{rf}).

otro oblicuo. También vemos que la mayoria de estos campos receptivos (aunque no todos) se presentan por pares, compartiendo la misma frecuencia espacial y orientación, pero con un desplazamiento de 90 grados entre sus fases, es decir, se encuentran en cuadratura de fase. Todos ellos son ortogonales entre sí. Como se ve, algunas de estas características se asemejan a las propiedades de los campos receptivos en el córtex visual [39][51].

La respuesta de una célula de nuestro "córtex" simulado a una imagen presentada, es decir, el producto interno entre la imagen y su campo receptivo, capturará una característica particular del espectro de frecuencias de dicha imagen. La invarianza ante translación es fácil de comprobar en este caso ya que al ser el muestreo uniforme, basta ver si los perfiles generados en las distintas posiciones son iguales, como así sucede (el grado de similaridad entre ellos depende de los limites impuestos al error del AIT). Dada esta invarianza espacial, el conjunto de respuestas de un nivel de nuestra red neuronal en todas las posiciones espaciales constituirá pues la convolución de la imagen de entrada con un filtro localizado en el dominio de frecuencias espaciales. Esto es lo que se había definido como un canal del sistema visual. Se puede así mostrar que en el caso sencillo de un muestreo regular, este modelo de red neuronal genera estructuras que aunque sencillas, son similares a las de los diferentes canales de frecuencia y orientación encontrados en el sistema visual.

7.4.2 Muestreo irregular

Aplicamos ahora este modelo de red neuronal al caso más complicado (e interesante) de un muestreo irregular (y cuya irregularidad desconoce la propia red). En este caso los campos receptivos que vayamos a obtener pueden ser muy diferentes de los del caso regular. Por una parte, será más difícil determinar a simple vista si los perfiles obtenidos corresponden con los esperados, ya que corresponderán a un muestreo irregular de los campos receptivos. Además, apreciar la esperada invarianza ante translación no será obvio en este caso, porque los pesos se adaptan para compensar las distorsiones de la matriz de muestreo, y dado que desde cada posición se ve un desorden diferente de los alrededores, no habrá equivalencia entre ellas.

Todas estas consideraciones pueden ser apreciadas gráficamente en la Fig. 7.4. La Fig 7.4a muestra los perfiles obtenidos con una matriz de muestreo regular 5×5 y 5 niveles de diferentes campos receptivos. Estos resultados son similares a los del apartado anterior (Fig. 7.3). Ahora, con la misma estructura en la red, repetimos el proceso con una matriz de muestreo desordenada. El desorden impuesto en las coordenadas (x, y) de la red fue :

$$\begin{pmatrix} (0,0) & (0,1) & (0,-1) & (1,0) & (1,1) \\ (0,1) & (1,0) & (0,0) & (-1,1) & (0,0) \\ (-1,1) & (0,1) & (-1,-1) & (1,0) & (0,1) \\ (1,0) & (1,1) & (0,0) & (1,-1) & (0,-1) \\ (0,1) & (0,0) & (-1,-1) & (1,1) & (1,0) \end{pmatrix},$$

donde cada "unidad de desorden" equivale a una translación del un cuarto de la distancia entre receptores de la red regular. Con esta distorsión, el conjunto de campos receptivos obtenidos en la posición de salida $(0,0)$ se muestra en la Fig 7.4b. Nótese que ahora se hace necesario especificar la posición de salida, dado que los perfiles calculados por la red varían de un punto a otro.

A continuación mostraremos que lo que obtenemos cuando operamos con esta nueva red (el producto interno entre estos nuevos vectores propios y las muestras no regulares de la imagen) es similar a lo que conseguíamos en el caso regular (algo que reproducía los canales de frecuencia y orientación del sistema visual). Partamos de los dos diferentes tipos de entradas que tenemos en ambos casos: $\mathbf{x} = \{x_i\}$, que es el vector de entrada cuando muestreamos con la matriz regular e $\mathbf{y} = \{y_i\}$, el vector de las muestras no regulares.

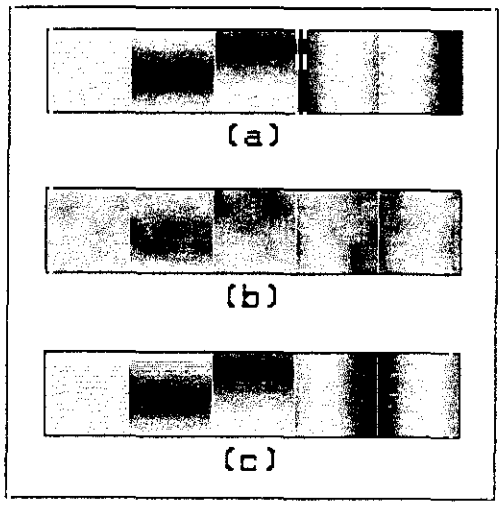


Figura 7.4: (a) Resultados generados por la red con 5×5 fotoreceptores y 5 niveles de campos receptivos en el caso regular. (b) Misma configuración de red, pero con una red de muestreo irregular. (c) Campos receptivos de (b) corregidos con la matriz T^{-1}.

Si aceptamos ciertas restriciones acerca de las posibles imágenes de entrada (las que se comentaron en la sección 7.2 sobre sus limitaciones en el ancho de banda) que hacen converger al AIT, podremos recuperar las muestras regulares a partir de las irregulares a través de una matriz de conversión T (ésta es exactamente la matriz calculada por el algoritmo de Maloney & Ahumada [71]). Por lo tanto :

$$\mathbf{x} = T \cdot \mathbf{y} \ . \tag{7.5}$$

Entonces, siendo Q_x la matriz de correlación de la entrada correspondiente al muestreo regular, x, y Q_y la del irregular, y, se cumple que:

$$Q_x = < x_i x_j > = < \mathbf{x} \cdot \mathbf{x}^T > = < T \cdot \mathbf{y} (T \cdot \mathbf{y})^T > = T \cdot < (\mathbf{y} \cdot \mathbf{y}^T > \cdot T^T = T \cdot Q_y \cdot T^T \ , \tag{7.6}$$

donde $< \ldots >$ significa el valor esperado.

Sabemos que con el AHG estamos calculando los primeros autovectores de esta matriz Q_x (nuestros campos receptivos). Llamamos C_x a la matriz cuyas filas son los vectores principales de Q_x. Resultados elementales de diagonalización de matrices nos dicen que para matrices simétricas como Q_x or Q_y se cumple que:

$$\begin{aligned} A_x &= C_x \cdot Q_x \cdot C_x{}^T \\ A_y &= C_y \cdot Q_y \cdot C_y{}^T \ , \end{aligned} \tag{7.7}$$

donde A_x y A_y son matrices diagonales.

Dado que A_x y A_y son ambas diagonales, podemos igualar una con otra a través de otra matriz diagonal F, de forma que $A_y = F \cdot A_x \cdot F^T$, y entonces combinando Ec. (7.6) y (7.7) obtenemos que:

$$A_y = F{\cdot}(C_x{\cdot}Q_x{\cdot}C_x{}^T){\cdot}F^T = F{\cdot}(C_x{\cdot}(T{\cdot}Q_y{\cdot}T^T){\cdot}C_x{}^T){\cdot}F^T = (F{\cdot}C_x{\cdot}T){\cdot}Q_y{\cdot}(F \cdot C_x \cdot T)^T \ , \tag{7.8}$$

y nos basta pues igualar las Ec. (7.8) y (7.6) para relacionar C_y con C_x :

$$C_y = F \cdot C_x \cdot T. \tag{7.9}$$

Podemos ver ahora que la salida de la red en el caso de un muestreo regular se puede expresar como el producto interno entre cada campo receptivo y la entrada x,

$$C_x \cdot \mathbf{x} \ , \tag{7.10}$$

ya que las filas de C_x son los autovectores de Q_x (los campos receptivos generados por el AHG). Ahora bien, en el caso de un muestreo irregular tenemos :

$$C_y \cdot \mathbf{y} = (F \cdot C_x \cdot T) \cdot \mathbf{y} = F \cdot C_x \cdot \mathbf{x} = F(Cx \cdot \mathbf{x}) \ . \tag{7.11}$$

Puesto que F es diagonal, la única diferencia entre ambos casos es un factor para cada campo receptivo. Además, el factor va a ser el mismo para el caso de valores propios degenerados, lo que significa que la corrección afectará por igual a aquellos pares de campos receptivos que diferían solamente en un desplazamiento de su fase.

Para comprobar finalmente lo expuesto anteriormente, hemos corregido los campos receptivos obtenidos en la Fig. 7.4b (C_y), con la correspondiente matriz T. Calculando $C_y \cdot T^{-1}$ deberiamos obtener algo similar a los campos receptivos del caso regular, C_x, (excepto por un factor), ya que según la Ec. (7.9), $C_y \cdot T^{-1} =$

$F \cdot C_x$. De hecho, esto es lo que sucede, como se muestra en la Fig. 7.4c, donde se ve que una vez "corregidos", los campos receptivos del caso irregular (C_y) son análogos a los del regular (Fig. 7.4a). En la figura los factores de la matriz F no se aprecian puesto que los valores se han expandido entre 0 y 256 para mejorar su visibilidad. Estos factores se ponen de manifiesto si calculamos el producto interno entre los cinco diferentes campos receptivos (una vez corregidos) correspondientes a la posición de salida $(0,0)$. Cada fila de la siguiente matriz está compuesta por el producto interno entre un campo receptivo y todos los demás:

$$\begin{pmatrix} 0.694 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.588 & -0.001 & 0.000 & 0.000 \\ 0.000 & -0.001 & 0.588 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.674 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.675 \end{pmatrix},$$

Se puede apreciar que la matriz es practicamente diagonal, lo que indica la ortogonalidad de los campos receptivos. También queda de manifiesto cómo campos receptivos que sólo difieren en una fase (filas 2 y 3 y filas 3 y 4) se ven alterados por el mismo factor. Con esto terminamos de mostrar que lo que la red neuronal está calculando, en el caso de un muestreo irregular, es lo que nosotros pretendíamos: la generación "espontánea" de unos ciertos coeficientes que hacen que la salida sea lo que hubiera sido si hubiesemos aplicado ciertos campos receptivos (similares a los que aparecen en el córtex visual) a una imagen muestreada regularmente.

En conclusión, se ve que ambos algoritmos, siendo capaces de converger simultáneamente, generan campos receptivos similares a los de células simples del córtex visual, sin imposiciones previas sobre su forma. Además, se establece una organización en distintos niveles, cada uno de los cuales está compuesto de aquellos campos receptivos que compartiendo una misma posición en el dominio de frecuencias, poseen una distinta posición espacial. Dado que una organización similar se presenta en el córtex visual, esquemas de este tipo podrían explicar el desarrollo de campos receptivos en esa zona del cerebro como resultado del aprendizaje.

Este modelo de red incorpora también (a través del AIT) la posibilidad de modificarse (aprender) ante cambios en la posición de los fotoreceptores. Esto es importante, puesto que se sabe que dichos fotoreceptores cambian su posición durante las primeras etapas del desarrollo [68]. Por consiguiente es necesario un sistema lo suficientemente flexible como para adaptarse a dichas variaciones, lo que también se conseguiría con un esquema como el aquí presentado.

Capítulo 8

Conclusiones

De los resultados presentados en esta memoria se extraen las siguientes conclusiones:

1. Se ha propuesto un esquema de representación de imágenes basado en funciones de Gabor como un entorno multipropósito, al facilitar una base común para aplicaciones tales como análisis, procesado, codificación, etc. El esquema consta de cuatro canales de frecuencia (dispuestos en octavas) y cuatro de orientaciones. Los diversos parámetros que caracterizan el esquema se han elegido en base a su similitud con los datos conocidos del sistema visual humano.

   Para el cálculo rápido de esta transformación de Gabor (TG) se han puesto a punto dos implementaciones alternativas en ambos dominios. Se ha prestado especial interes a la implementación en el dominio espacial, donde el uso de máscaras de tamaño 7×7 y la introducción de un esquema piramidal han reducido el coste computacional, haciendolo similar al de las implementaciones más convencionales en el dominio de Fourier. Además, una implementación espacial resulta especialmente ventajosa en ciertas aplicaciones, tales como el análisis de puntos o zonas aisladas de una imagen, o en casos donde se parte de un muestreo no uniforme.

2. A pesar de que la TG no es exacta, se ha optimizado la recuperación de la imagen original mediante la asignación de diferentes ganancias a los distintos canales de frecuencia. De esta forma se propone una transformación inversa "cuasicompleta", con la que se obtienen recuperaciones de la imagen comparables, bajo criterios visuales, con las de otras transformaciones exactas.

En particular, se ha mostrado que en casos de pérdidas parciales de información la TG presenta una notable robustez, con efectos espúreos menores que los de una transformación exacta como la Córtex (TC). Dado que dicha transformación se ha usado con éxito en aplicaciones de codificación de imágenes, esto nos permite suponer que esquemas de codificación basados en la transformada de Gabor pueden alcanzar similares prestaciones.

3. Con objeto de comprobar el carácter multipropósito de nuestro modelo se ha estudiado la posibilidad de usar directamente las salidas del esquema de Gabor como descriptores de la textura (mas una no-linealidad, consistente en extraer el módulo del par complejo formado por los canales simétricos y antisimétricos). Dichos descriptores forman una matriz 4×4 cuyos componentes serán las respuestas de los 4×4 canales del esquema de Gabor en el entorno del punto considerado.

4. Dichos descriptores se han aplicado a dos tareas de análisis de imágenes:

   - **Segmentación:** usando un algoritmo convencional de K-medias se han obtenido excelentes resultados, siendo capaz el algoritmo de dividir las imágenes en las diferentes texturas que las componían.
   - **Clasificación:** se ha aplicado un algoritmo de Bayes, más un filtrado de moda, para clasificar los píxeles de una imagen según la textura a la que pertenecieran. Los porcentajes de clasificación correcta obtenidos oscilan entre un 78 y un 99 %.

   Estos buenos resultados, a pesar de usar métodos convencionales poco sofisticados, ponen de manifiesto que nuestra matriz 4×4 basada en el esquema de Gabor constituye una buena descripción de las texturas estudiadas.

5. Usando la misma matriz de descriptores, se ha estudiado el problema de reconocimiento invariante de texturas ante:

   - **Rotaciones:** se ha comprobado que una rotación se refleja en una permutación de las columnas de la matriz, proponiendose un vector de parámetros invariantes ante rotación. Con este nuevo vector de descriptores se clasificaron correctamente imágenes que contenían texturas rotadas un ángulo arbitrario.
   - **Cambios de escala:** análogamente, un cambio de escala provoca un desplazamiento de las filas de la matriz. Sin embargo, la pérdida de información que se produce en un esquema con pocos canales, como el

aquí propuesto, imposibilita la construcción de invariantes ante cambios de escala, excepto en el caso de imágenes con un decaimiento regular en su espectro (como ocurre en fractales).

6. Se ha comprobado que nuestra matriz de descriptores es tambien util para la clasificación de fractales brownianos en base a su dimensión fractal. A partir de este hecho se ha propuesto un método para estimar el grado de fractalidad de una textura. Para el caso de fractales se presenta una calibración que permite determinar fácilmente su dimensión fractal a partir de la matriz de descriptores. Usando este método se ha comprobado que el grado de fractalidad de las imágenes naturales es alto, mientras que el de texturas artificiales es bajo.

7. Se presenta un modelo realista de las primeras etapas del sistema visual restringido a la fóvea, que considera los siguientes aspectos:

   - La influencia de la óptica del ojo, cuya MTF había sido previamente medida en nuestro laboratorio.
   - El efecto de la apertura finita de los conos; se ha comprobado que en condiciones de visión normales su efecto es despreciable frente al de la MTF.
   - La distribución de conos en la fovea. Para simular el mosaico de fotoreceptores en la fóvea, se ha diseñado una red de muestreo pseudo-hexagonal de espaciado variante siguiendo un método de crecimiento en espiral. La distribución de fotoreceptores en dicha red reproduce fielmente la existente en la fóvea, tanto en cuanto al espaciado de los conos en función de la excentricidad, como al número total de conos dentro de un radio determinado.
   - Los campos receptivos de células simples del córtex se modelan con funciones de Gabor, implementadas en filtros hexagonales con 61 muestras.

8. Usando este modelo se han reproducido diversos efectos observados experimentalmente.

   - El uso de diferentes ganancias para los distintos canales de frecuencia ha permitido reproducir ilusiones visuales como las conocidas **bandas de Mach**.
   - Se han reproducido diversos fenómenos de **aliasing** ("zebra patterns", aliasing de orientación, etc) observados experimentalmente cuando se elimina el efecto de la óptica del ojo mediante métodos interferométricos.

- Se ha simulado la detección de estímulos sinusoidales por el sistema visual. De esta forma se ha caracterizado la función de sensibilidad al contraste (CSF) para nuestro modelo, comprobandose que puede reproducir su forma típica.
- Repitiendo el mismo procedimiento para estímulos situados a una excentricidad de 1° estudiamos la evolución de la respuesta en frecuencias de nuestro modelo con el campo visual. La nueva CSF obtenida muestra un comportamiento notablemente similar al de las curvas obtenidas en experimentos psicofísicos en condiciones equivalentes.

9. Se ha propuesto un modelo simple de sistema visual basado en una red neuronal con dos sistemas de aprendizaje: uno hebbiano y el otro supervisado. Los resultados obtenidos a partir de este modelo son:

   - La generación de forma espontánea (tras un aprendizaje) de campos receptivos selectivos en frecuencia y orientación, similares a los observados en células simples del córtex visual.
   - Los campos receptivos con las mismas características situados en distintas posiciones espaciales se agrupan en diferentes niveles.
   - El modelo es capaz de corregir los efectos de variaciones en la posición de los puntos de muestreo.

   Los puntos arriba considerados muestran que un modelo similar podría explicar el desarrollo y organización de campos receptivos en el cortex a traves de mecanismos de aprendizaje.

Bibliografía

[1] W.K. Pratt, Digital image Processsing, Wiley, New York, 1978

[2] R.C. Gonzalez and P. Wintz, Digital image Processsing, Addison Wesley, 1977

[3] J. Beck, B. Hope, and A. Rosenfeld, eds., Human and Machine Vision, Academic Press, 1981

[4] A.C. Slade, ed., Physical and Biological processing of Images, Springer-Verlag, Berlin 1983

[5] R. Navarro, J. Santamaría, and J. Bescós, "Accomodation-dependent model of the human eye with aspherics", *J. Opt. Soc Am., A*, **2**, pp 1273–1281, 1985

[6] R. Navarro, P. Artal, and D.R. Williams, "Optical quality of the human eye across the visual field", Ophthalmic & Visual Optics meeting, pp. 48–51, Santa Fe, Jan 1992

[7] P. Artal, R. Navarro, D.H. Bruinard, S.J. Galvin, and D.R. Williams, "Off-axis optical quality of the eye and retinal sampling", ARVO meeting, 3240-2, Sarasota, FL, 1992

[8] R.M. Mersereau, "The processing of hexagonally sampled two-dimensional signals", *Proc. of the IEEE*,, **67**, pp. 930–949, 1979

[9] G. Ostemberg, "Topography of the layers of cones and rods in the human retina", *Acta Ophthalmologica, Supplement*, **6**, pp. 1 –103, 1935

[10] V.H. Perry and A. Cowey, "The ganglion cell and cone distribution in the monkey's retina: implications for central magnification factors", *Vision Res.*, **25**, pp 1795–1810, 1985

[11] J. Hirsch and W.H. Miller, "Does cone positional disorder limit resolution?", *J. Opt. Soc. Am. A*, **4**, pp. 1481–1492, 1987

[12] J. Hirsch and C.A. Curcio, "The spatial resolution capacity of human foveal retina" , *Vision Res.*, **29**, pp 1095–1101, 1989

[13] C.A. Curcio, K.R. Sloan Jr., O. Packer, A.E. Hendrickson, and R.E. Kalina, "Distribution of cones in human and monkey retina: individual variability and radial asimmetry", Science , **236**, pp. 579–582, May 1987

[14] C.A. Curcio, K.R. Sloan, R.E. Kalina, and A.E. Hendrikson, "Human photoreceptor topography", *Journal of Comparative Neurology*, **292**, pp. 497–523, 1990

[15] D.R. Williams, "Aliasing in human foveal vision", *Vision Res.*, **25**, pp. 195–205, 1985

[16] D.R. Williams, "Seeing through the photoreceptor mosaic", *Trends in Neuroscience*, **9**, pp. 193-198, 1986

[17] N.J. Coletta, D.R. Williams, and C.L.M. Tiana, "Consequences of spatial sampling for human motion perception", *Vision Res.*, **30**, pp. 1631-48, 1990

[18] N.J. Coletta and D.R. Williams, "Psychophysical estimate of extrafoveal cone spacing" , *J. Opt. Soc. Am. A*, **4**, pp. 1503–1513, 1990.

[19] J.I. Yellot Jr., "Spectral analysis of spatial sampling by photoreceptors: topological disorder prevents aliasing" , *Vision Res.* , **22**, pp. 1205–1210, 1982

[20] W.H. Miller and G.D. Bernard, "Averaging over the foveal receptor aperture curtails aliasing", *Vision Res.* , **23**, pp. 1365–1369, 1983.

[21] J.I. Yellot Jr., "Image sampling properties of photoreceptors: a reply to Miller and Bernard", *Vision Res.* , **24**, pp. 281–282, 1984

[22] A.S. French, A.W. Snyder, and D.G. Stavenga, "Image degradation by an irregular mosaic", *Biol. Cybernetics*, **27**, pp. 229–233, 1977

[23] P. Artal and R. Navarro, "High-resolution imaging of the living human fovea: measurement of the intercenter cone distance by speckle interferometry", *Opp. Lett.* , **14**, pp. 1098–1100 , 1989

[24] R. W. Rodieck, The Vertebrate Retina, W.H. Freeman, San Francisco, 1973

[25] L. Spillmann and J.S Werner, eds. , Visual perception. The neurophysiological foundations, Academic Press, San Diego, CA, 1990

[26] D. Marr, Vision, Freeman, San Francisco, CA, 1982

[27] P. Gouras, "Identification of cone mechanisms in monkey ganglion cells", *J. of Physiology*, **199**, pp. 533–547, 1968

[28] S.W. Kuffler, "Discharge patterns and functional organization of mammalian retina", *Journal of Neurophysiology*, **16**, pp. 37–68, 1953

[29] H. Wassle, U. Grunert, J. Rohrenbeck, and B.B. Boycott, "Retinal ganglion cell density and cortical magnification factor in the primate", *Vision Res.* , **30**, pp. 1897–1911. 1990

[30] C.A. Curcio and K.A. Allen, "Topography of ganglion cells in human retina", *Journal of Comparative Neurology*, **300**, pp. 5–25, 1990

[31] L.N. Thibos, F.E. Cheney, and D.J. Walsh, "Retinal limits to the detection and resolution of gratings", *J. Opt. Soc. Am. A*, **4**, pp. 1524–1529, 1897

[32] S.J. Anderson and R.F. Hess, "Post-receptoral undersampling in normal human peripheral vision", *Vision Res.* , **30**, pp. 1507–1515, 1990

[33] D.H. Hubel and T.N. Wiesel, "Receptor fields, binocular interaction, and functional arquitecture in the cat's visual cortex", *Journal of Physiology*, **160**, pp. 106–154, 1962

[34] D.H. Hubel and T.N. Wiesel, "Anatomical demonstration of columns in the monkey striate cortex", Nature (London), **225**, pp. 747–750, 1969

[35] J.C. Horton and E.T. Hedley-White, "Mapping of cytochrome oxidase patches and ocular dominance columns in human visual cortex", Trans. of the Royal Soc. of London, Series B, **304**, pp. 255-272, 1984

[36] R.L. De Valois, D.G. Albrecht, and L.G. Thorell, "Spatial frequency selectivity of cells in macaque visual cortex", *Vision Res.* , **22**, pp. 545–559 , 1982

[37] F.W. Campbell and J.J. Kulikowski, "Orientation selectivity of the human visual system", *J. of Physiology*, **187**, pp. 437–445, 1966

[38] S. Marcelja , "Mathematical description of the response of simple cortical cells" , *J. Opt. Soc. Am. A* , **70** , pp. 1297–1300 , November 1980

[39] D.A. Pollen and S.F. Ronner, "Visual cortical neurons as localized spatial filters" , *IEEE Trans. on Systems, Man, and Cybernetics*, **SMC–13** , pp. 907–916 , 1983

[40] J.G. Daugman , "Spatial visual channels in the Fourier plane", *Vision Res.*, **24** , pp. 891–910 , 1984

[41] J. Jones and L. Palmer, "An evaluation of the two–dimensional Gabor filters model of simple receptive fields in cat striate cortex" , *J. Neurophisiology* , **58**, pp. 538–539 , 1987

[42] H.R. Wilson and J. R. Bergen, "A four mechanism model for threshold spatial vision", *Vision Res.*, **19**, pp. 19–32, 1979

[43] A.B. Watson , "Detection and recognition of simple spatial forms" , A.C. Slade, ed., Physical and Biological processing of Images Springer-Verlag , Berlin 1983

[44] D.V. Van Essen, W.T. Newsome, and J.H.R. Maunsell, "The visual field representation in striate cortex of the macaque monkey: asymmetries, anisotropies, and individual variability", *Vision Res.* , **24**, pp. 429–448, 1984

[45] L. Maffei and A. Fiorentini, "The visual cortex as a spatial frecuency analyser", **Vision Res.**, **13** pp. 1255–1268, 1973

[46] E. Wigner, "On the quantum correction for thermodinamic equilibrium", *Phys. Review*, **40**, pp. 749–759, 1932

[47] G. Cristobal, J. Bescós, and J. Santamaría , "Image analysis through the Wigner distribution function" , *Applied Optics*, **28** , pp. 262–271 , January 1989

[48] T.D. Reed and H. Wechsler , "Segmentation of textured images and geltalt organization using spatial/spatial–frequency representations" , *IEEE Trans. on Patt. Anal. and Mach. Intell.* , **PAMI–12**, pp. 1–12, January 1990

[49] L. Cohen, "Generalized phase-space distribution functions", *J. Mathematical Physics*, **7**, pp. 781–786, 1966

[50] D. Gabor , "Theory of communications", *J. Inst. Elect. Eng.*, **93**, pp. 429–457 , 1946

[51] J. Daugman , "Uncertainty relation for resolution in space, spatial–frequency and oriented optimized by two–dimensional visual cortical filters", *J. Opt. Soc. Am. A*, **2**, pp. 1160-1169 , 1985

[52] M.J. Bastiaans , " A sampling theorem for the complex spectrogram, and Gabor's expansion of a signal in gaussian elementary signals", *Opt. Engineering* , **20** , pp. 594–598 , 1981

[53] A.B. Watson, "The cortex transform: rapid computation of simulated neural images" , *Computer Vision, Graphics, and Image Processing.* , **39**, pp. 311–327 , 1987.

[54] A.B. Watson , "Efficiency of a model human image code" , *J. Opt. Soc. Am. A* , **4** , pp. 2401–2417 , December 1987

[55] A.B. Watson and A.J. Ahumada, Jr. , "A hexagonal orthogonal-oriented pyramid as a model of image representation in visual cortex", *IEEE Trans. on Biomedical Engineering* , **36** , pp. 97–106 , January 1989

[56] J. Daugman, "Complete discrete 2D Gabor transform by neural networks for image analysis and compression" , *IEEE Trans. on Acoust. Speech & Sign. Proc.* , **ASSP–36**, pp. 1169–1179 , 1988

[57] M. Porat and Y.Y. Zeevi , "The generalized Gabor scheme for image representation in biological and machine vision" , *IEEE Trans. on Patt. Anal. & Mach. Intell.* , **PAMI–10**, pp. 452–468, 1988

[58] R. Navarro and A. Tabernero, "Gaussian wavelet transform: two alternative fast implementations for images", *Multidimensional System and Signal Processing*, **2**, pp. 421–436, 1991

[59] P.L. Burt and E.H. Adelson , "The Laplacian pyramid as a compact image code" , *IEEE Trans. on Communications* , **COM–31** , pp. 532–540 , April 1983

[60] J. Morlet, I. Forgeau, and D. Giard, "Wave propagation and sampling theory", *Geophysics*, **47**, pp. 203–236, 1982

[61] S. G. Mallat , "A theory for multiresolution signal decomposition: the wavelet representation" , *IEEE Trans. on Pattern Anal. and Mach. Intell.* , **PAMI-11** , pp. 674–693 , 1989

[62] N.S. Jayant y P. Noll, Digital coding of waveforms, Englewood Cliffs, NJ, Prentice-Hall, 1984

[63] J.W. Woods and S.D. O'Neil, "Subband coding of images", *IEEE Trans. on Acoustics, Speech, and Signal Processing*, **ASSP-34**, 1986

[64] E.P. Simoncelli and E.H. Adelson, "Subband transforms", J.W Woods, ed., Subband Image Coding, Cap. 4, Kluwer Academics Publishers, Norwell, MA, 1990

[65] E.P. Simoncelli and E.H. Adelson, "Non-separable extensions of quadrature mirror filters to multiple dimensions", *Proc. of the IEEE*, **78**, 1990

[66] A. Rosenfeld, ed., Multiresolution image processing and analysis, Springer-Verlag in Information Sciences , 1984

[67] A.J. Ahumada, Jr. and K. Turano, "Calibration of a system with progresive receptor dropout", 13th European Conference on Visual Perception, Paris, France, September 1990

[68] C. Youdelis and A. Hendrikson, "A qualitative and quantitative analysis of the human fovea during development", *Vision Res.*, **26**, pp. 422–427, 1986

[69] R.P. Lippmann, "An introduction to computing with neural nets", IEEE ASSP Magazine, pp. 4–22 April 1897

[70] J. Hertz, A. Krogh, and R.G. Palmer, Introduction to the Theory of Neural Computation, Santa Fe Institute Addison-Wesley, Redwood City, CA, 1991

[71] L.T. Maloney and A.J. Ahumada, Jr., "Learning by assertion: two methods for calibrating a linear visual system", *Neural Computation*, **1**, pp. 392–401, 1989

[72] A.J. Ahumada, Jr. and J.B. Mulligan, "Learning receptor positions from imperfectely known motions", in B. Rogowitz and J. Allebach, eds., Human Vision, Visual Processing, and Digital Display, Proc. SPIE 1249, pp. 124–134, 1990

[73] A.J. Ahumada, Jr. and J.B. Mulligan, "Network compensation for missing sensors", Human Vision, Visual Processing, and Digital Display II, Proc. SPIE 1453, pp. 134–146, 1991

[74] A.J. Ahumada, Jr., "Learning receptor positions", in M. Landy and J.A. Movshon, eds., Computational Models of Visual Processing, pp. 23–34, MIT Press, Cambridge, MA, 1992

[75] S. Amari, "Dynamical stability of formation of cortical maps", M.A. Arbib and S. Amari, eds., Dynamic Interations un Neural Networks; Models and Data, Springer Verlag, New-york, pp. 15–34, 1988

[76] T. Kohonen, "Self-organization and associative memory", Springer, New-York, 1989

[77] E. Oja, "A simplified neuron model as a principal component analysis", *J. Math. Biology*, **15**, pp. 267–273, 1982

[78] T.D. Sanger, "Optimal unsupervised learning in a single layer linear feed-forward neural network", *Neural Networks*, **2**, pp. 459–473, 1989

[79] A.B. Watson, "Perceptual-components architecture for digital video", *J. Opt. Soc. Am. A*, **7**, pp. 1943–1954 1990

[80] B. Julesz, E.N. Gilbert, L.A. Sheep and H.L. Frisch, "Inability of humans to discriminate between textures that agree in second-order statistics", *Perception*, **2**, pp. 391–405, 1973

[81] R.M. Haralick , "Statistical and structural approach to texture", *Proc. IEEE*, pp. 786–804, May 1979

[82] W.K. Pratt, O.D. Faugeras, A. Gagalowicz , "Visual discrimination of stochastic texture fields" , *IEEE Trans. on System, Man, and Cybernetics*, **SMC–8**, pp. 796–804 , 1978

[83] R.M. Haralick, K. Shanmugan, and I. Dinstein , "Textural features for image classification" , *IEEE Trans. on System, Man, and Cybernetics*, **SMC–3**, pp. 610–621 , November 1973

[84] M. Turner , "Texture discrimination by Gabor functions", *Biol. Cybernetics*, **55**, pp. 71–82 , 1986

[85] I. Fogel and D. Sagi, "Gabor filters as texture discriminator", *Biol. Cybernetics*, **61** , pp. 103–113 , 1989

[86] A. Sutter, J. Beck, and N. Graham, "Contrast and spatial variables in texture segregation: testing a simple spatial-frequency channels model", *Perception & Psychophysics*, **46 (4)**, pp. 312–332, 1989

[87] A.C. Bovik, M. Clark and W.S. Geisler , "Multichannel texture analysis using localized spatial filters" , *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **PAMI–12**, pp. 55–73 , 1990

[88] A. Tabernero and R. Navarro, "Performance of Gabor functions for texture analysis", enviado a *IEEE Trans. on Patt. Anal. and Mach. Intell.*

[89] A.P. Pentland , "Fractal-based description of natural scenes" , *IEEE Trans. on Patt. Anal. and Mach. Intell.*, **PAMI-6**, pp. 661-674 , 1984

[90] J.M. Keller and S. Chen , "Texture description and segmentation through fractal geometry" , *Computer Vision, Graphics, and Image Processing*, **45**, pp. 150-166 , 1989

[91] N. Yokoza, K. Yamamoto, and N. Funakubo, "Fractal-based analysis and interpolation of 3D natural surfaces shapes and their applications to terrain modeling" , *Computer Vision, Graphics, and Image Processing*, **46**, pp. 284-302 , 1989

[92] A.V. Oppenheim, R.W. Schafer, "Digital signal processing" , Englewood Cliffs, NJ, Prentice-Hall , 1975

[93] L.O. Harvey,Jr., and V.V. Doan , "Visual masking at different polar angles in the two-dimensional fourier plane" , *J. Opt. Soc. Am. A*, **7** , pp. 116-127, January 1990

[94] J.L. Dannemiller, and J.N. Ver Hoeve , "Two-dimensional approach to psychophysical orientation tunning" , *J. Opt. Soc. Am. A*, **7**, pp. 141-151 , January 1990

[95] R. Navarro, J. Santamaria , and R. Gomez, "Automatic log spectrum restoration of atmospheric seeing" , *Astron. Astrophys.* , **174**, pp. 344-351 , 1987

[96] D.J. Field , "Relation between the statistics of natural images and the response properties of cortical cells" , *J. Opt. Soc. Am. A* , **4** , pp. 2379-2394 , December 1987

[97] T. Caelli and B. Julesz, "On perceptual analizers underlying visual texture discrimination. Part I", *Biol. Cybernetics*, **28**, pp. 167-175, 1978

[98] M.M. Galloway , "Texture analysis using gray level run lenghts", *Computer Vision, Graphics, and Image Processing*, **4**, pp. 172-179 , 1975

[99] J.S. Weszcka, C.R. Dyer, and A. Rosenfeld , "A comparative study of texture measure for terrain modeling" , *IEEE Trans. on System, Man, and Cybernetics*, **SMC-6** , pp. 269-285 , 1976

[100] G. Lendaris and G. Stanley, "Diffraction pattern sampling for automatic pattern recognition" , *Proc. IEEE* , **58** , pp. 198-216, 1970

[101] J. Malik and P. Perona, "Preattentive texture discrimination with early vision mechanisms", *J. Opt. Soc. Am. A*, **7**, pp. 923–932, 1990

[102] Brodatz , Textures: a photographic album for artists and designers, Dover, 1966

[103] H. Nyblack, Digital image processing, Englewood Cliffs, NJ, Prentice-Hall, 1986

[104] P.A. Devijter and J. Kittler , Pattern recognition , Englewood Cliffs, NJ, Prentice-Hall , 1986

[105] T.M. Cover and P.E. Hart , "Nearest neigbour pattern classification", *IEEE Trans. on Information Theory* , **IT–13**, pp. 21–27, January 1967

[106] T. Kailath , "The divergence and the B-distance in signal selection" , *IEEE Trans. on Communication Technology* , **COM–15**, pp. 52–60 , February 1967

[107] A.K. Jain , "On a estimation of the Bhattacharyya distance" , *IEEE Trans. on System, Man, and Cybernetics*, **SMC–6**, pp. 763–766, 1976

[108] M. Ichino, "Multiclass pattern recognition systems based on independent subrecognition systems" , *IEEE Trans. on System, Man, and Cybernetics*, **SMC–6**, pp. 256–269 , April 1976

[109] J.Y. Hsiao and A.A. Sawchuk , "Unsupervised textured image segmentation using feature smoothing and probabilistic relaxation techniques" , *Computer Vision, Graphics, and Image Processing*, **48**, pp. 1–21 , 1989

[110] M. Fang and G. Hausler , "Class of trasforms invariant under shift, rotation, and scaling" , *Applied Opt.*, **29**, pp. 704–708 , 1990

[111] R. Wu and H. Stark, "Rotation invariant pattern recognition using optimum feature extraction" , *Applied Opt.*, **24** , pp. 179–184 , 1985

[112] E. Freysz, B. Pouligny, F. Argoul, and A. Arneodo, "Optical wavelet transform of fractal aggregates" , *Physical Review Letters*, **64** , pp. 745–748 , February 1990

[113] B.B. Mandelbrot , "The fractal geometry of nature" , Freeman , 1982

[114] R.F. Voss , "Random fractal forgeries" in Fundamentals Algorithms in Computer Graphics , pp. 805–835 , Springer–Verlag, Berlin, 1985

[115] M.J. Bastiaans , "On the sliding-window representation in digital signal processing" , *IEEE Trans. on Acoustic, Speech, and Signal Processing*, **ASSP–33** , pp. 868–873 , August 1895

[116] P. Artal and R. Navarro, "Simultaneous measurements of two point-spread functions at different positions across the human fovea", *Applied Optics*, **31**, to appear in 1992.

[117] A.J. Ahumada, Jr. and A. Poirson, "Cone sampling array models", *J. Opt. Soc. Am. A*, **4**, 1987

[118] J. Hirsch and R. Hylton, "Quality of the primate photoreceptor lattice and limits of spatial vision", *Vision Res.* , **24**, pp. 347–355, 1984

[119] A.W. Snyder and W.H. Miller, "Photoreceptor diameter and spacing for highest resolving power", *J. Opt. Soc. Am. A* , **67**, pp. 696–698, 1977

[120] A. Johnson, "Spatial scaling of central and peripheral contrast-sensitivity functions", *J. Opt. Soc. Am. A*, **4**, pp. 1583–1593 1987

[121] E.T. Davis, D. Yager, and B.J. Jones, "Comparison of perceived spatial frequency between the fovea and the periphery", *J. Opt. Soc. Am. A*, **4**, pp. 1606–1611, 1987

[122] J. Rovamo, V. Virsu, and R. Nasanen, "Cortical magnification factor predicts the photopic contrast sensitivity of peripheral vision", Nature, **271**, pp. 54–56, 1978

[123] V. Virsu, R. Nasanen, and Smoviita, "Cortical magnification and peripheral vision", *J. Opt. Soc. Am. A*, **4**, 1987

[124] R.F. Quick, "A vector-magnitude model of contrsat detection", *Kybernetic* **16**, pp. 65–67, 1974

[125] A.B. Watson, "Summation of grating patches indicates many types of detectors at once retinal position", *Vision Res.*, **21**, pp. 1115–1122, 1982

[126] R. Sekuler, H.R. Wilson, and C. Owsley, "Structural modeling of spatial vision", *Vision Res.*, **24**, pp. 689–700, 1984

[127] S. LeVay, M.P. Stryker, and C.J. Shatz, "Ocular dominance columns and their development in layer iv of the cat's visual cortex", *Journal of Comparative Neurology*, **179**, pp. 223–244, 1978

[128] O.J. Braddick, J. Wattam-Bell, and J. Atkinson, "Orientation-specific cortical responses develop in early infance", Nature (London), **320**, pp. 617–619, 1986

[129] C. von der Malsburg, "Self-organization of orientation sensitive cells in the striate cortex", *Kybernetik*, **14**, pp. 85–100, 1973

[130] A.B. Widrow and M.E. Hoff, "Adaptative swiching circuits", INst. of Radio Engineers, WESCOM Record, Part 4, pp. 96–104, 1960

[131] A.B. Widrow and S.D. Stearns, "Adaptative signal processing", Englewood Cliffs, NJ, Prentice-Hall, 1985