

UNIVERSIDAD POLITÉCNICA DE MADRID
ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE
TELECOMUNICACIÓN



PROYECTO FIN DE CARRERA

**MODELOS VISUALES EN EL ANÁLISIS DE
LA CALIDAD DE IMAGEN**

Ana Rosa Gallego Aguilar

Febrero 2006

UNIVERSIDAD POLITÉCNICA DE MADRID
ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE
TELECOMUNICACIÓN

PROYECTO DE FIN DE CARRERA

TÍTULO: MODELOS VISUALES EN EL ANÁLISIS DE LA CALIDAD DE IMAGEN.

AUTOR: ANA ROSA GALLEGOS AGUILAR.

TUTOR: GABRIEL CRISTÓBAL PÉREZ.

TRIBUNAL:

PRESIDENTE: ANDRÉS DE SANTOS Y LLEÓN.

VOCAL: FERNANDO CALLE GÓMEZ.

SECRETARIA: MARÍA JESÚS LEDESMA CARBAYO.

SUPLENTE: ANTONIO GABRIEL LOMEÑA MORENO.

FECHA DE LECTURA:

CALIFICACIÓN:

Departamento de Ingeniería Electrónica.
Instituto de Óptica “Daza de Valdés” (C.S.I.C.)

Resumen:

En este proyecto se van a estudiar los modelos visuales y su aplicación concreta en sistemas de medida de calidad de imagen. Se va desarrollar una herramienta para la evaluación automática de la calidad perceptible de imágenes monocromáticas que hayan sido procesadas mediante algoritmos de compresión o de marcado de agua. Para ello se investigan cuáles son los atributos visuales perceptivos de mayor relevancia en la determinación de la calidad de imagen. Se analizan los distintos componentes del Sistema Visual Humano (adaptación de luminancia, CSF, enmascaramiento, etc.) y para cada uno de ellos se realizarán diferentes implementaciones dentro de un esquema de medida de calidad perceptible. En base a los resultados que proporcionan los distintos modelos de estos atributos se indican las implementaciones que resultan más fiables.

Palabras clave:

Calidad de imagen, modelos visuales, adaptación de luminancia, CSF, contraste, enmascaramiento, índices de calidad, visibilidad error.

Agradecimientos:

A todos los compañeros del CSIC. En especial a mi tutor Gabriel, por su paciencia, sus indicaciones y sus consejos.

A todos los que os pasasteis por el laboratorio del CSIC para la realización de los test, gracias.

A mis amigos de la facultad: Isa, Mariajo, Elena, Fede, Noelia, Andrés, Vanessa, Raquel, Augusto y Pablo. Porque habéis llenado de historias y recuerdos mis cinco añitos en la escuela y porque todavía me seguís aguantando.

A Leo, Carlos y Ari por los veranos de estudio, de proyecto y de fiesta.

A Elena, Ana e Inma por estar ahí siempre apoyando moralmente. A Pepa además por las consultas estadísticas y a Alex por los libros de medicina.

A mi familia, con todo mi cariño.

ÍNDICE

1. Introducción.	
1.1. Estado del Arte	4
1.2. Objetivos del Proyecto	6
1.3. Estructura del Proyecto.....	7
2. Sistema Visual Humano.	
2.1. Anatomía y fisiología visual.....	10
2.1.1. El ojo	10
2.1.2. Campos receptores de la retina.....	12
2.1.3. Vías visuales. Integración de la información en la corteza	16
2.2. Propiedades de la visión	20
2.2.1. Adaptación visual. Sensibilidad a la intensidad luminosa.....	20
2.2.2. Contraste	22
2.2.3. Sensibilidad en frecuencia.....	25
2.2.4. Enmascaramiento	30
2.3. Movimientos oculares y atención visual	32
3. Evaluación de la calidad en imágenes.	
3.1. Calidad de imagen. Clasificación de métodos de medida.	35
3.2. Modelos de calidad de imagen subjetivos.....	37
3.3. Modelos de calidad de imagen objetivos.....	40
3.3.1. Modelos de medida de error	40
3.3.2. Modelos de medida de calidad perceptible.....	43
3.3.3. Modelos híbridos de medida de calidad	51
3.4. Distorsión de imagen.....	56
3.4.1. Técnicas de compresión y artefactos.....	56
3.4.2. Marcado de agua.....	60

4. Implementación de modelos visuales para la evaluación de la calidad de imagen.	
4.1. Preprocesado. Características del monitor	63
4.2. Adaptación de luminancia	66
4.3. Función de Sensibilidad al contraste (CSF)	68
4.3.1. Modelo 1 de CSF (B.W.Rust)	70
4.3.2. Modelo 2 de CSF (Daly)	71
4.3.3. Modelo 3 de CSF (Barten)	74
4.4. Descomposición en canales	75
4.4.1. Filtros de Gabor	76
4.4.2. Transformada Cortex. Filtros anillo	82
4.5. Cálculo de contrastes	85
4.6. Enmascaramiento espacial.....	89
4.6.1. Enmascaramiento con CSF como filtro.....	90
4.6.2. Enmascaramiento con CSF por pesos	92
4.7. Mapas e índices de error.....	94
5. Comportamiento y comparativa de los modelos de medida de calidad.	
5.1. Validación de modelos de medida de calidad perceptible.....	100
5.2. Sistema de medida de calidad con CSF como filtro	109
5.3. Sistema de medida de calidad con CSF por pesos.	120
5.4. Otros modelos.....	127
5.4.1. Medida de calidad basada en información estructural(SSIM)	127
5.4.2. Medida de calidad basada en distribuciones Wigner.....	132
6. Conclusiones.	
6.1. Conclusiones.....	139
6.2. Posibles ampliaciones.	141
7. Anexos.	
Anexo A. Imágenes del test DSCQS y resultados MOS.....	144
Anexo B. Entorno gráfico.....	155
Anexo C. Manual de programas básicos.	164
Bibliografía.	171

GLOSARIO

CSF	Contrast Sensitivity Function.
CRT	Cathode Ray Tube.
DCT	Discrete Cosine Transform.
DSIS	Double Stimulus Impairment Scale.
DSCQS	Double Stimulus Continuous Quality Scale.
DWT	Discrete Wavelet Transformation.
FFT	Fast Fourier Transform.
ITU	International Telecommunication Union.
JND	Just Noticeable Difference.
JPEG	Joint Photographic Experts Group.
LUT	Look-Up Table.
MOS	Mean Opinion Score.
MRI	Magnetic Resonance Imaging.
MSE	Mean Square Error.
NMSE	Normalized Mean Square Error.
PEM	Perceptual Error Measure.
PQS	Picture Quality Scale.
PSF	Point Spread Function.
PSNR	Peak Signal to Noise Ratio.
PWD	Pseudo Wigner Distribution.
SNR	Signal to Noise Ratio.
SSCQE	Single Stimulus Continuous Quality Evaluation.
SSIM	Structural SIMilarity.
SVH	Sistema Visual Humano.
VDM	Visual Discrimination Model.
VDP	Visible Differences Predictor.
WVD	Wigner- Ville Distribution.

1

Introducción.

Las imágenes digitales se ven afectadas por una amplia gama de distorsiones durante su adquisición, procesado, almacenado, transmisión y reproducción. Todos estos procesos, lógicamente, pueden dar como resultado una degradación de la calidad visual. Para aplicaciones en las que finalmente la imagen va a ser observada por personas, el único método realmente correcto de evaluar la calidad de esa imagen sería mediante la valoración subjetiva que daría el observador. En la práctica la evaluación subjetiva mediante test resulta bastante complicada y en algunos casos, según la aplicación, puede llegar a ser inabordable. La meta en la investigación y el desarrollo de métricas de calidad objetivas consiste precisamente en conseguir medidas cuantitativas que puedan valorar de forma automática la calidad de imagen percibida.

Una métrica de calidad de imagen objetiva puede cumplir diferentes funciones según su aplicación. Por ejemplo, se puede utilizar para monitorizar y ajustar dinámicamente la calidad de la imagen en un servidor de video digital de forma que se adecue a los recursos disponibles. También puede servir para optimizar algoritmos y parámetros en sistemas de procesado de imágenes como en el caso de sistemas de comunicación en los que es posible utilizarla para optimizar el diseño de filtros y algoritmos de codificación y decodificación. En todas estas aplicaciones lo que interesa es que a pesar de la degradación introducida el observador final perciba ésta lo menos posible. Por ello, las métricas de calidad objetivas que resultan efectivas son las que evalúan esta calidad percibida. Para ello es necesario saber qué es lo que este

observador percibe, siendo imprescindible conocer el funcionamiento del SVH y modelarlo en la medida de lo posible.

Los modelos del sistema visual humano, por tanto, van a ser el fundamento para la definición de métricas objetivas de calidad perceptible, pero además, no sólo van a estar presentes como sistema de valoración sino que lo ideal sería que las principales características del SVH sean integradas directamente en el corazón de la herramienta diseñada. Este planteamiento adquiere una gran importancia en el desarrollo de métodos y sistemas de compresión ya que las capacidades limitadas de procesamiento del SVH permiten que una imagen pueda ser modificada sin que las distorsiones introducidas sean detectadas con lo que se podrán conseguir tasas de compresión mayores sin que se modifique la calidad percibida de forma significativa. Por tanto, el objetivo es hacer que las distorsiones introducidas por el compresor sean invisibles al ojo, es decir, un compresor perceptual sin pérdidas. De hecho, ya desde sus inicios en la primera reunión del grupo de expertos JPEG 2000 en Sydney en 1997, se comenzaron los primeros trabajos para incorporar las características del SVH como una herramienta de optimización dentro del estándar. El método más habitual para mejorar la compresión mediante técnicas de optimización visual está basado en transformar las amplitudes de la imagen a un dominio perceptivo uniforme. Puesto que la respuesta global del sistema visual a la escala de grises puede aproximarse por una no-linealidad de tipo de raíz cúbica, el proceso a realizar consistiría en convertir la imagen a un dominio inverso para cuantificar posteriormente. Esta técnica es la base de la denominada corrección gamma que está presente en la mayoría de los monitores de video. Los otros dos elementos claves a optimizar son el contenido de las frecuencias espaciales y el enmascaramiento. El primero de ellos involucra el cálculo de la función de sensibilidad al contraste (CSF) para encontrar los pasos de cuantificación, mientras que el segundo está relacionado con la importancia del contexto en donde se encuentren los objetos a codificar. El fenómeno del enmascaramiento (“masking”) es un efecto perceptivo bien conocido tanto para el caso de imágenes como para el caso de señales acústicas, etc. Se debe básicamente al hecho de que estímulos de gran intensidad van a enmascarar a estímulos más pequeños. Este efecto puede ser aprovechado en compresión para concentrar los errores en aquellas regiones en donde van a ser visualmente imperceptibles. Sin embargo, el efecto de enmascaramiento admite múltiples interpretaciones y su comportamiento no es bien

conocido en algunas situaciones, por lo que constituye un área activa en la investigación actual del sistema visual.

En definitiva, la necesidad de compresión cada vez mayor y con una calidad aceptable, motivada por la gran demanda actual en la transmisión y almacenamiento de la información visual ha impulsado, como se ha comentado, la incorporación de los modelos visuales a la mayoría de algoritmos de compresión [10, 12, 22, 25, 26, 35]. En general, la compresión es del todo necesaria ya que las imágenes digitales son de tamaño considerable. Esto se agrava en el caso de aplicaciones con imágenes médicas. Por ejemplo, en el contexto de la radiología, la medida final de la calidad de una imagen reside en lo útil que es la imagen para conseguir un diagnóstico certero y preciso. El verdadero test para un radiólogo es la detección precisa de incluso pequeñas anormalidades. Es por ello que el mantenimiento de la mayor fidelidad posible en la representación de las imágenes es crucial en la práctica radiológica y no se pueden tolerar artefactos ni pérdida de calidad. Se necesitan, por tanto, esquemas de compresión sin pérdidas lo que hace que el problema del tamaño de la imagen médica sea crítico en cuanto a su almacenamiento, consulta, etc. Por tanto, si en general los modelos de SVH son útiles para alcanzar tasas de compresión elevadas con las menores pérdidas de calidad posible en el caso de las imágenes médicas resultan tanto más útiles puesto que lo que se busca es un sistema de compresión sin pérdidas *perceptibles*.

Finalmente, no se puede olvidar que las aplicaciones de los modelos del SVH en entornos de compresión o en cualquier otro no tendría sentido si no se pudiese evaluar a posteriori el valor de calidad percibida de la imagen. En conclusión, se van a necesitar las *métricas de calidad objetiva perceptible* para evaluar si se han conseguido los objetivos deseados después del procesado de la imagen.

1.1. Estado del arte.

A lo largo de muchos años de investigación en el procesado visual, uno de los problemas fundamentales ha sido la ausencia de una métrica universalmente aceptada que proporcione una medida de la calidad de una imagen con unas prestaciones aceptables en un amplio rango de situaciones. Los test subjetivos son usados normalmente para obtener la estimación más precisa de la calidad de imagen. Sin embargo, llevar a cabo estos test en la práctica resulta muy pesado en cuanto a la carga temporal que conlleva, aparte de resultar caros y difíciles de repetir puesto que requieren unas condiciones de visualización especiales. Debido a estos inconvenientes se hace necesario el uso de métricas de calidad objetivas. De estas, las más conocidas, y que se siguen empleando en casi todos los casos, son medidas simples matemáticas basadas en la relación señal/ruido de pico (PSNR) o el error cuadrático medio (MSE). A pesar de ser las más utilizadas está comprobado que dichas medidas no guardan ninguna relación directa con la calidad de la imagen percibida. La búsqueda y desarrollo de métricas más robustas que proporcionen medidas de calidad acordes con la percibida es actualmente un área activa de investigación. De hecho, se han definido un gran número de posibles métricas que incorporan de una u otra manera factores perceptivos. El desarrollo de estas métricas objetivas basadas en modelos del SVH ha sido posible gracias a los avances en el conocimiento del SVH y su funcionamiento. En general lo que se suele modelar es el procesado a bajo nivel desde el ojo hasta la corteza visual primaria. Las propiedades del SVH o características visuales que incorporan normalmente estas métricas son la adaptación de luminancia, la función de sensibilidad al contraste (CSF) y el enmascaramiento. La forma en la que se lleva a cabo esta incorporación es mediante modelos de las características visuales mencionadas que se obtienen gracias a los datos experimentales conseguidos mediante la realización de test psicofísicos de detección de umbral con estímulos simples.

Los distintos modelos de medida de calidad perceptible existentes, puesto que se han definido teniendo en cuenta las características del SVH, van a compartir una estructura básica similar y luego, dependiendo de la aplicación para la que fueron propuestos tendrán diversas particularidades. En general en todos ellos se pueden encontrar una etapa de calibrado, un filtrado de la imagen a diferentes frecuencias espaciales y orientaciones, ajustes según la sensibilidad al contraste, mecanismos para

modelar el enmascaramiento y una etapa final que obtiene el error o bien como un número que describe la calidad de la imagen o bien como un mapa de detectabilidad de la distorsión o ambos.

La implementación de estas métricas es en muchos casos compleja, son bastante elaboradas y con un coste computacional elevado. El grado de complejidad depende de las características visuales que se contemplen dentro de la métrica y el modelo escogido para las mismas. Dos de los modelos de referencia para la mayor parte de métricas que se han desarrollado ya que resultan bastante generales y robustos son el “Visible Difference Predictor” (VDP) de Daly [7] y el “Visual Discrimination Model” (VDM) de Lubin [19]. Sin embargo, estos modelos presentan algunas limitaciones como son el no tener en cuenta características del SVH de alto nivel (extracción de características, procesos cognitivos, reconocimiento de patrones, atención visual...), que se diseñan para condiciones de visibilidad del estímulo cercanas al umbral (los experimentos psicofísicos en los que se basan son en condiciones de detección umbral) y otras que se verán con más detalle a lo largo del proyecto. Es por ello, que algunos modelos han intentado incorporar características del procesado del SVH de alto nivel [26] o directamente buscar nuevas métricas que de alguna forma simulen ciertas características de la percepción sin llegar a incorporar los modelos visuales [30, 31].

En definitiva este es un área con distintos frentes de investigación abiertos puesto que aún no hay una métrica aceptada estándar para la evaluación de la calidad percibida de imagen. Además, el estudio del comportamiento de las métricas ya existentes es complicado ya que la validación de las mismas ha de hacerse mediante test subjetivos sobre un conjunto de imágenes dadas con lo que, en cualquier caso, su evaluación es parcial.

1.2. Objetivos.

El objetivo fundamental es investigar cuales son los atributos visuales perceptivos más importantes que juegan un papel decisivo en la determinación de la calidad de una imagen, como interactúan y se correlacionan entre si. En definitiva, se trata de analizar los modelos visuales en la evaluación de la calidad de imagen.

En concreto, la realización de este análisis se va a llevar a cabo con dos propósitos determinados que en realidad van a surgir como resultado directo de la investigación realizada en el ámbito de los modelos y métricas de calidad de imagen. En primer lugar se trata de desarrollar una herramienta de trabajo en forma de entorno gráfico que sea global, amplia y abierta. Global en el sentido de que incluya todo tipo de métricas para la evaluación de la imagen y no sólo las que se basan modelos de SVH. De esta forma se van a poder comparar y contrastar los resultados obtenidos mediante los distintos sistemas de medida de calidad sacando las pertinentes conclusiones en cuanto a tiempos de ejecución, precisión de los resultados, adecuación de los métodos a los distintos tipos de imágenes, etc. Amplia en el sentido de abarcar el mayor número de modelos posibles para los distintos componentes del SVH de manera que una vez centrados en los modelos visuales sea posible también obtener distintas conclusiones en cuanto a qué modelos en concreto o combinación de ellos obtienen mejores resultados. Por último, se trata de que sea una herramienta abierta que facilite la incorporación futura de nuevos métodos o modelos de forma sencilla, que se pueda implementar en cualquier sistema sin más que realizar algunos ajustes y que sirva para una gama amplia de imágenes. En segundo lugar, se pretende desarrollar un modelo que permita explicar como un observador humano combina los atributos visuales para lograr una valoración de la calidad de imagen percibida. El desarrollo de una métrica perceptible permitirá tener una herramienta de tipo “observador ideal” que proporcione una medida de la calidad más próxima a la que se obtendría a través de un observador humano. Este modelo de medida de calidad de imagen se implementará de forma dinámica dentro del entorno gráfico antes mencionado.

En el presente proyecto se desarrollan por tanto herramientas para la evaluación automática de la calidad subjetiva de imágenes monocromáticas. En concreto, las imágenes que se pretenden analizar y que resultan de mayor interés serán aquellas que

hayan sido procesadas mediante algoritmos de compresión o de marcado de agua. En particular serán objeto de estudio por una parte las imágenes médicas de video-microscopía y radiológicas (MRI, PET, etc.) y por otra y para que el estudio resulte lo más completo posible se utilizarán imágenes conocidas y de uso común en entornos de compresión.

1.3. Estructura del proyecto.

La organización de este proyecto para conseguir los objetivos que se han planteado va a estar claramente dividida en tres bloques: investigación, implementación y evaluación.

En primer lugar, se necesitará una investigación y análisis del SVH y de sus componentes. Este estudio previo se desarrolla en el capítulo 2, donde se examinan las características de comportamiento del SVH. Para comprender su funcionamiento se realizará una descripción de la anatomía y fisiología visual y posteriormente se determinarán las distintas propiedades de la visión y su respuesta. Esta es la base sobre la que después se podrán realizar modelos de medida de calidad perceptible. Además, será necesario aportar un marco general en cuanto a modelos ya existentes. La descripción de algunos de los modelos de medida de calidad de imagen, su estructura básica y los distintos problemas y ventajas que llevan asociados se hace en el capítulo 3. También se definen y clasifican aquí los distintos tipos de métricas de calidad de imagen y lo que es más importante, se define el concepto de calidad que se tendrá en cuenta en este proyecto. Finalmente, se da una visión general en cuanto al tipo de distorsiones con las que van a tratar estos métodos.

En segundo lugar, para la construcción de un modelo de medida de calidad perceptible se van a llevar a cabo distintas implementaciones de los componentes del SVH (adaptación, CSF, enmascaramiento, etc.). El capítulo 4 reúne para cada uno de ellos las propuestas que se han desarrollado mediante la herramienta MATLAB, explicando cómo se relacionan y unen entre sí los distintos bloques que componen el modelo.

En tercer lugar se van mostrar y analizar los resultados obtenidos para las diferentes imágenes y distorsiones que son evaluadas con los sistemas de medida de calidad implementados. En el capítulo 5 se describe primero el método DSCQS que sirve para la validación de los modelos y después se examinan los resultados para los modelos de medida de calidad perceptible y los de otras métricas híbridas ya descritas dentro de la clasificación del capítulo 3. En el capítulo 6 se resumen las conclusiones obtenidas y se señalan posibles ampliaciones y mejoras.

Finalmente, en el anexo A se van a mostrar los valores subjetivos que se han obtenido mediante la realización del test DSCQS y las imágenes utilizadas. El anexo B describe el entorno gráfico que se ha construido y su utilización. En el anexo C se enumeran y explican brevemente los programas y funciones más importantes generadas mediante MATLAB para el desarrollo de este proyecto así como su relación.

2

Sistema Visual Humano.

La calidad de una imagen puede verse influenciada por diversos factores entre los que pueden destacarse las técnicas de adquisición, de procesamiento y de visualización a las que se somete a esa imagen. En la mayoría de los casos el resultado obtenido va a ser contemplado por un observador que formará parte activa en la evaluación de la imagen final. Por ello, para poder dar una medida de calidad acorde con la percibida por ese observador es necesario estudiar el SVH y los mecanismos de percepción, entendiéndose como tal los procesos de extracción de la información y la formación de representaciones sobre las que pueda operar el sistema cognitivo.

En el estudio de la visión se pueden diferenciar claramente dos etapas. Por un lado, se realiza un procesado a bajo nivel que no utiliza información de experiencias pasadas, ni razonamiento. Este es el procesado que incluyen la mayor parte de las métricas de calidad de imagen basadas en las características del SVH y es en el que se centra este capítulo. Por otra parte, el procesado de alto nivel utiliza recursos como la atención y la memoria, siendo estos de un mayor nivel de complejidad y por tanto mucho más difíciles de modelar.

En este capítulo, en resumen, se pretende analizar las características de comportamiento del SVH, comprendiendo su funcionamiento y determinando, en la medida de lo posible, su respuesta. Esta será la base sobre la que después se podrán modelar métricas de calidad perceptible.

2.1. Anatomía y fisiología visual.

2.1.1. El ojo.

Es el órgano del SVH en el que comienza el procesamiento visual y cuyo funcionamiento básico consiste en recoger la luz y enfocarla en su superficie posterior, transformando energía luminosa en eléctrica. Es sensible a las radiaciones electromagnéticas con longitudes de onda comprendidas entre los 400 y los 780 nm aproximadamente.

La figura 2.1 muestra una sección del globo ocular, con una forma prácticamente esférica y de unos 20 mm de diámetro. En la figura, se pueden apreciar los componentes principales de los que se compone el ojo, que van a ser enumerados a continuación, explicando su funcionamiento y características más relevantes.

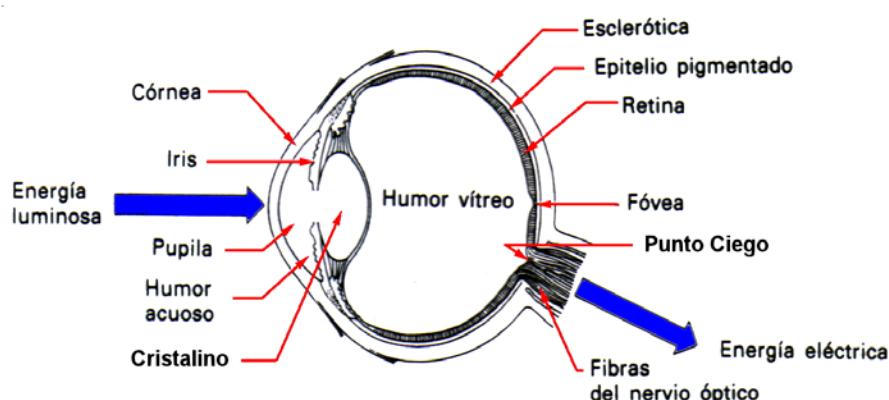


Figura 2.1: Sección vertical del ojo.

La **esclerótica** es el tejido duro, blanco y fibroso que conforma la parte exterior del globo ocular. Es opaca y se conoce comúnmente como el blanco del ojo. La parte delantera y central del ojo es la **córnea**. Se trata de una membrana transparente y dura cuya función principal consiste en refractar la luz. Además, la córnea tiene una cierta curvatura que contribuye a dirigir la luz y concentrarla en una pequeña apertura redonda (de 2 a 8 mm de diámetro) situada en el centro del ojo y que permite que la luz pase al interior del mismo, la **pupila**. Esta a su vez está rodeada por el **iris**, un músculo circular que permite regular la cantidad de luz que entra en el ojo, de forma que si la intensidad luminosa es elevada se contraerá haciendo menor el diámetro de la pupila y viceversa.

La parte anterior del iris es la que contiene el pigmento visible que caracteriza el color de ojos.

Tras el iris se encuentra el **cristalino** que está formado por una serie de capas concéntricas con distintos índices de refracción que actúan como una lente flexible enfocando la luz en el fondo del ojo. El cristalino absorbe cerca del 8% de la luz visible del espectro, incrementándose esta absorción en las zonas del infrarrojo y del ultravioleta, pudiendo llegar a dañarse el ojo por un exceso de radiación en estas frecuencias. Entre la córnea y el cristalino se encuentra una sustancia líquida, clara y transparente denominada **humor acuoso**.

La **retina** recubre la parte interior del ojo y es en ella donde se enfoca la luz incidente. La luz se convierte aquí en señales nerviosas mediante células sensibles a la luz. Entre el cristalino y la retina se encuentra el **humor vítreo** que es una sustancia gelatinosa transparente e incolora. El humor vítreo llena todo el espacio entre el cristalino y la retina y ocupa alrededor de 2/3 del volumen ocular. Para poder enfocar objetos cercanos y lejanos es necesario que el ojo humano cambie la forma del cristalino. Este proceso, que se denomina *acomodación*, es controlado mediante un grupo de músculos situados alrededor del iris y sucede prácticamente en tiempo real. Una vez realizada esta acomodación, la luz se proyecta en la retina y es transformada en impulsos eléctricos mediante dos tipos de fotorreceptores que reciben el nombre de conos y bastones.

Los conos se distribuyen fundamentalmente por la zona central de la retina denominada **fóvea**, son sensibles al color y responsables de la visión fotópica. Proporcionan una visión en detalle y pueden ser de tres tipos según la respuesta que presentan al contenido de la radiación. Es un sistema que funciona con tres receptores de color distintos, dando lugar a la teoría tricromática para la percepción del color. Cada ojo posee entre 6 y 7 millones de conos y su umbral de visibilidad se sitúa alrededor de 1 μ L (micro-lumen).

Los bastones proporcionan información general del campo visual. Son sensibles a valores muy bajos de iluminación con umbrales de 1nL, de tal forma que hacen posible la visión nocturna o con poca iluminación pero sin información de color. Son

responsables, por tanto, de la visión escotópica. El número de bastones es de 75 a 150 millones.

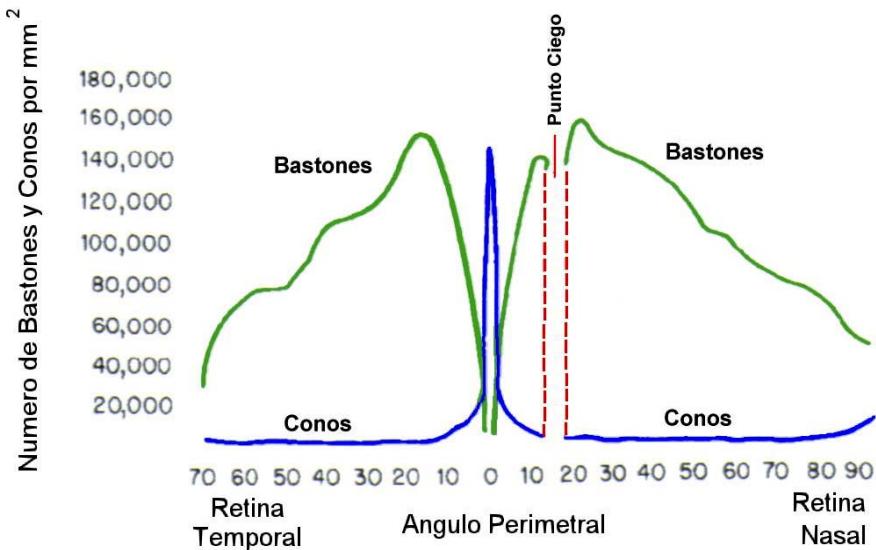


Figura 2.2: Distribución de conos y bastones en la retina.

Como puede apreciarse en la figura 2.2 la distribución de estos fotorreceptores en la retina es poco uniforme y más o menos simétrica respecto a la fóvea. Esta distribución hace que tengamos un amplio ángulo de visión y una resolución espacial elevada a nivel local que se resuelve con el continuo movimiento del ojo enfocando los objetos de interés en la fóvea.

El **disco óptico** se caracteriza por ser la única zona de la retina en la que no hay fotorreceptores, constituyendo, por tanto, un *punto ciego*. De aquí parten los vasos sanguíneos que riegan la retina y además en esta zona convergen los axones formando el nervio óptico que parte con la información que será enviada al cerebro.

2.1.2. Campos receptores de la retina.

Un campo receptor es un área de la retina formada por fotorreceptores (conos y bastones) cuyas respuestas influyen en células más avanzadas del procesamiento visual: las células ganglionares. La figura 2.3 muestra una sección de la retina con las células que la forman y su conexiónado. Como puede observarse, entre las células ganglionares y los fotorreceptores existen otros tres tipos de células: horizontales, amacrinas y bipolares.

Las células horizontales unen los fotorreceptores y las células bipolares mediante conexiones relativamente largas y paralelas a las capas de la retina. Las células amacrinas funcionan de forma similar pero como puente de unión entre células bipolares y ganglionares.

Las células bipolares pueden recibir la información directamente de los conos y bastones o a través de las células horizontales y posteriormente se manda a las células ganglionares o bien de forma directa o a través de las células amacrinas. La información, por tanto, puede seguir diversos caminos dentro del conexionado de las distintas células retinianas.

Este esquema general es válido para toda la retina con algunas puntualizaciones. En la fóvea cada cono se conecta con una célula bipolar y esta a su vez con una ganglionar siguiendo el camino más directo. A medida que nos alejamos de la fóvea más receptores convergerán en células bipolares y mayor número de estas en ganglionares.

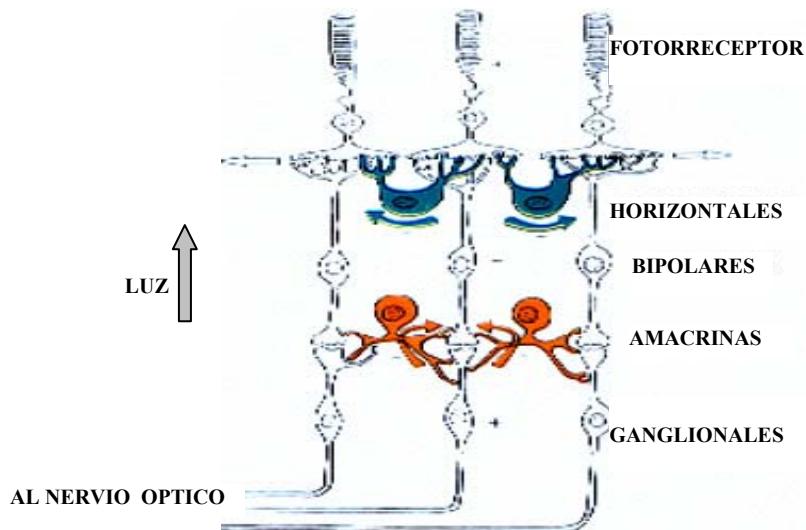


Figura 2.3: Sección simplificada de la retina (longitud aproximada de $\frac{1}{4}$ de mm).

Las células bipolares se dividen en dos clases: de centro “off” y “on”. Se debe aclarar que las únicas células que disparan impulsos en la retina son las ganglionares, y al hablar aquí de centro “on” u “off” nos referimos a si responden a la luz o a la oscuridad. Así, el proceso da comienzo en los receptores, que se encuentran activos en

la oscuridad (la luz los hiperpolariza apagándolos). Entre el receptor y la célula bipolar de centro “on” aparece una sinapsis inhibidora que mantiene la célula inactiva en la oscuridad, y la luz, apagando el receptor, elimina esta inhibición de la célula haciendo que se liberen neurotransmisores en los terminales de la misma. Si la célula bipolar es de centro “off”, la sinapsis es excitadora, la célula está activada en la oscuridad y disminuye la liberación de neurotransmisores cuando se somete a la luz. La periferia de este tipo de células se origina a través de la información que suministran las células horizontales que tienen sinapsis excitadoras con las bipolares “on” e inhibidoras con las “off” (efecto de la luz en la periferia excitador).

De las células amacrinas se conocen más de 20 tipos diferentes en cuanto a su anatomía y pueden desarrollar diversas funciones entre las que se encuentran respuestas específicas a objetos en movimiento.

Las **células ganglionares** tienen un comportamiento similar a las bipolares y se han identificado también dos clases: células ganglionares de centro encendido (excitadoras u “on”) y células de centro apagado (inhibidoras u “off”). Las células ganglionares de centro encendido se activan cuando el centro de su campo receptor se encuentra excitado y la periferia del mismo no. En las células de centro apagado sucederá a la inversa. De esta manera, si las dos partes del campo receptor reciben el tipo de estimulación que necesitan, sus efectos se suman y se alcanza un nivel de activación máximo en la ganglionar, pero si el estímulo produce efectos opuestos en el centro y en la periferia, las dos regiones compiten entre sí y la célula ganglionar correspondiente se mantiene casi inactiva. Esta interacción es conocida como *inhibición lateral*. La inhibición lateral es uno de los procesos más importantes en la explicación de muchos fenómenos perceptivos, y en concreto, este particular funcionamiento hace que ya en las primeras etapas de la visión se pierda prácticamente la información de intensidad luminosa en niveles absolutos, a favor de una percepción del contraste.

De la descripción que se ha hecho hasta ahora se podría pensar que los campos receptores forman un mosaico de pequeños círculos retinianos, pero no es así, los campos receptores de células ganglionares vecinas se solapan como puede apreciarse en la figura 2.4 b. Un fotorreceptor puede influir sobre varias células ganglionares estando situado en el centro del campo receptor de algunas y en la periferia de otras. Por tanto,

excitará a varias células a través de sus centros si son células de centro encendido y a través de sus periferias si son células de centro apagado.

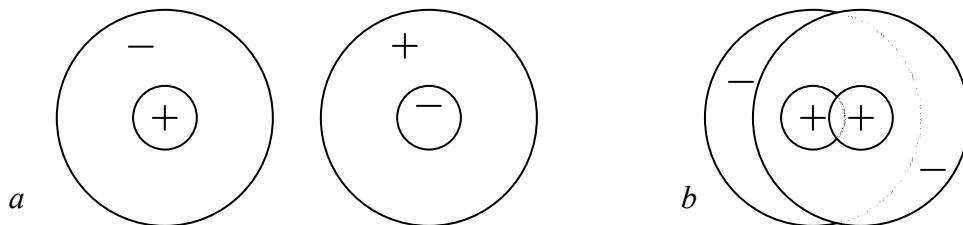


Figura 2.4: Campos receptores. **a.** Campos receptores de célula ganglionar de centro encendido y centro apagado respectivamente. **b.** Campos receptores de dos células ganglionares vecinas.

Además, los campos receptores difieren en tamaño de una célula ganglionar a otra. En concreto, los centros de los campos receptores son más pequeños en la fóvea y se hacen mayores a medida que nos alejamos de ella. Esto, nuevamente, explica el hecho de que nuestra agudeza visual, la capacidad de distinguir objetos pequeños, sea mayor en la fóvea. El tamaño del campo receptor está pues relacionado con la frecuencia espacial de forma que células con campos receptores grandes responderán a bajas frecuencias espaciales y viceversa. De acuerdo con esto se puede hacer una nueva clasificación de las células ganglionares en magno (10% células) que son las que poseen campos receptores grandes y parvo (90% células) con un campo receptor pequeño. Las células magno responden al movimiento y a objetos grandes, son rápidas y de alta sensibilidad al contraste. Las parvo son lentas y están implicadas en la detección del color y del detalle.

En conclusión, los mensajes que el ojo envía hacia el cerebro a través del nervio óptico tienen poco que ver con intensidad absoluta ya que las células ganglionares no responden bien a cambios de luz difusa. Lo que la célula señala es el resultado de una comparación entre la cantidad de luz que impacta en un cierto punto de la retina con la cantidad media de luz que ilumina su periferia inmediata. Esto permite ver el mismo objeto en condiciones de iluminación completamente distintas. Por ejemplo, si leemos un periódico a plena luz del sol o en una habitación poco iluminada lo que vemos son letras negras sobre un fondo blanco. Sin embargo, si tuviésemos en cuenta la luz que

reflejan, las letras negras a plena luz reflejarían más que el papel blanco en la habitación.

2.1.3. Vías visuales. Integración de la información en la corteza.

La vía visual de un cerebro humano es el camino que sigue la información visual desde el ojo hasta que llega a la corteza visual primaria.

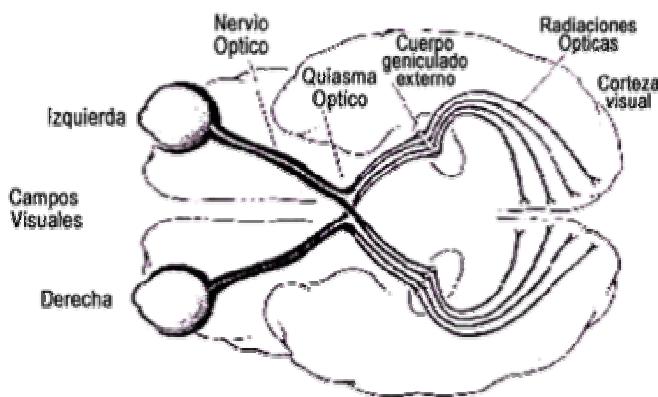


Figura 2.5: Vías ópticas hacia la corteza visual.

Las fibras que salen del ojo formando el nervio óptico llegan sin interrupción al quiasma óptico y a partir de ahí se dirigen a distintas zonas del cerebro. La mayoría de ellas (80% aproximadamente) envía la información a través del tracto óptico al núcleo geniculado lateral, el resto es información para el control de funciones de movimiento ocular y reflejo pupilar a la luz (cerebro medio, colículo superior) y también para procesos de sincronización de ritmos biológicos (hipotálamo).

En el **quiasma óptico** se cruzan los nervios ópticos de forma que a cada **núcleo geniculado lateral** le llega información procedente de ambos ojos pero que se corresponde con la mitad del campo visual opuesta. Es decir, al núcleo geniculado lateral derecho le llega toda la información visual correspondiente con el campo visual izquierdo. El campo visual es la zona del espacio que puede ser percibida para una posición fija del ojo. El campo visual izquierdo se refiere a la zona del espacio a la izquierda del punto al que estemos mirando. De esta forma, cada hemisferio cerebral se encarga de la mitad opuesta del entorno.

Cada cuerpo geniculado lateral se encuentra dividido en 6 capas y una célula de una capa cualquiera recibe información de un solo ojo. Así, por ejemplo, en el caso del cuerpo geniculado lateral izquierdo la secuencia al ir de arriba hacia abajo sería: proyección de la hemirretina izquierda del ojo derecho, izquierdo, derecho, izquierdo, izquierdo y derecho.

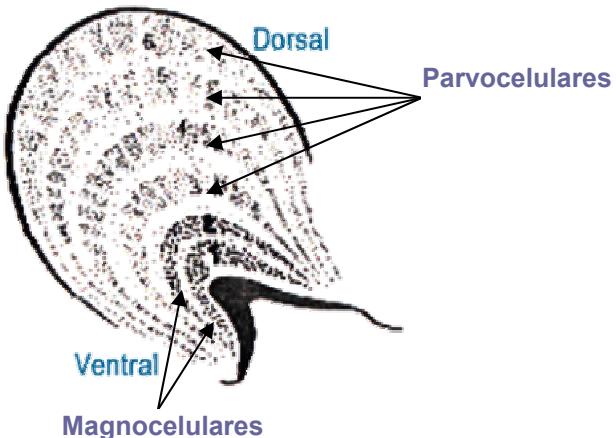


Figura 2.6: Sección de cuerpo geniculado lateral.

Además, como puede apreciarse en la figura 2.6, las capas se pueden dividir en parvocelulares (dorsales) y magnocelulares (ventrales) según el tamaño de las células que las forman y su respuesta al estímulo visual. Las capas se encuentran alineadas según la zona del campo visual o más concretamente de la zona de la retina de la que reciben la información, es decir, las conexiones están organizadas topográficamente. Esta representación topográfica también se mantendrá en la corteza visual primaria. Así si se produjera alguna lesión en una porción pequeña de la corteza visual se correspondería con una ceguera en algún punto del campo visual.

La información visual sale de los cuerpos geniculados laterales mediante unas bandas anchas denominadas **radiaciones ópticas** que la encaminan hasta la **corteza visual primaria (V1)**. La corteza visual primaria es una capa de células de unos 2mm de grosor en la parte occipital del cerebro. Contiene aproximadamente 200 millones de células frente a los 1,5 del cuerpo geniculado lateral. Estas células fueron clasificadas por Hubel [16] en células simples y complejas.

Las células simples tienen campos receptores de tipo “on” y “off” pero con patrones de excitación e inhibición como los mostrados en la figura 2.7 a. Estímulos que cubran una mayor área excitadora darán lugar a mayores respuestas y los que cubran regiones excitadoras e inhibidoras a la vez producirán respuestas menores puesto que hay una cancelación mutua. Estas células poseen, además, tres o cuatro geometrías diferentes con todas las orientaciones y posiciones dentro del campo visual posibles. Son, por tanto, las primeras células que presentan sensibilidad a la orientación.

Las células complejas son mucho más numerosas que las simples, constituyendo el 75% [16] de las células de la corteza visual primaria. Responden a estímulos lineales orientados estacionarios y en movimiento (donde presentan una mayor respuesta). Una célula compleja recibe información de varias simples de la misma orientación y cuyos campos receptores se solapan sobre el campo receptor de la célula compleja como muestra la figura 2.7 b. Si una línea excita una célula simple se producirá una respuesta en la compleja, si la línea está en movimiento, la activación sucesiva de células simples hace que la respuesta en la célula compleja sea sostenida. Además, entre un 10% y un 20% de las células complejas de la corteza visual primaria muestran una fuerte selectividad direccional. En algunas células de la corteza aparece un efecto de *inhibición final*. Este fenómeno consiste en que la célula responde máximamente a una línea con una determinada longitud y hacerla mayor no aumentará la respuesta. Una explicación para las células con inhibición final sería la aferencia de células complejas excitadoras e inhibidoras sobre otras células complejas [16].

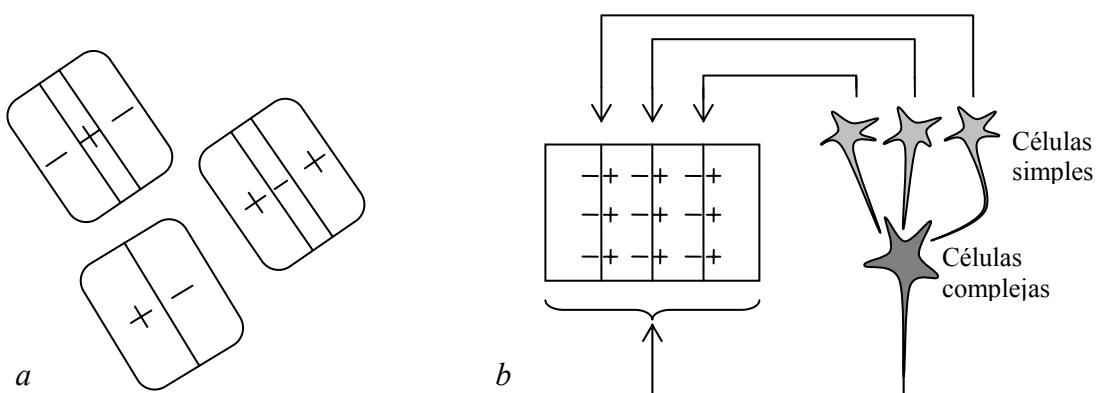


Figura 2.7: Campos receptores. **a.** Mapas típicos de campos receptores de células simples. **b.** Posible campo receptor y diagrama de conexión de célula compleja.

Finalmente, y por primera vez en todo el proceso de percepción, aparecen células en la corteza visual primaria que reciben aferencias de ambos ojos, es decir, hay una convergencia binocular que permitirá construir una única imagen de la escena. En general todas las gradaciones de dominancia ocular relativa están presentes (desde células monopolizadas por el ojo izquierdo pasando por células con igual influencia de ambos ojos hasta células con respuesta al ojo derecho).

La corteza visual primaria se estratifica en 6 capas. Las más densas son la 4C y la 6 mientras que la 1 apenas contiene células nerviosas y está formada en su mayor parte por axones. En la figura 2.8 se pueden observar las principales conexiones efectuadas entre el cuerpo geniculado lateral y las distintas capas de la corteza visual primaria, así como las conexiones entre capas y con otras regiones del cerebro.

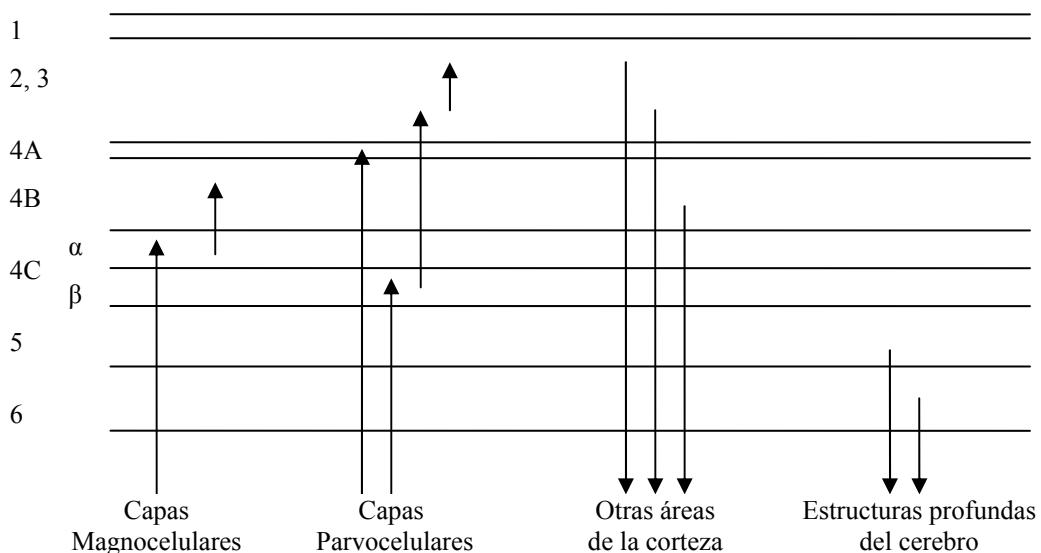


Figura 2.8: Estructura e interconexiones en la corteza visual primaria.

Las células que forman la corteza visual primaria se encuentran dispuestas en columnas verticales relacionadas con la orientación del estímulo. Además, se produce también una agrupación de células sensibles a una determinada frecuencia espacial. Se configuran así *detectores de frecuencia espacialmente orientada* con las células del V1. Se debe tener en cuenta que hay también otras capas de la corteza (desde V2 a V5) que participan en el proceso visual realizando diversas tareas. Así, por ejemplo, la zona V4 estaría relacionada con el procesamiento del color.

En conclusión, el sistema visual humano que se ha descrito es modular y paralelo. Se podrían diferenciar básicamente tres etapas. La primera de ellas sería la óptica (equivalente a un enfoque), la segunda la retiniana (transducción de la señal luminosa en determinados impulsos eléctricos) y finalmente el procesado cerebral.

2.2. Propiedades de la visión.

A continuación se van a analizar algunos de los mecanismos más relevantes de la visión. Se pretende realizar una caracterización del SVH en cuanto a su sensibilidad y respuesta al estímulo.

2.2.1. Adaptación visual. Sensibilidad a la intensidad luminosa.

La magnitud del estímulo percibido por el cerebro será mayor cuanto mayor sea la magnitud del estímulo luminoso que se proyecta en la retina. En este apartado se examinará cómo es esta relación y qué rango abarca.

El rango de niveles de intensidad luminosa que se presentan en la naturaleza y a los que el SVH se tiene que adaptar es muy amplio, alrededor de 10^{10} . Sin embargo, simultáneamente, el SVH es capaz de discriminar sólo unos pocos niveles. La adaptación se realiza para un nivel de intensidad luminosa I_o dentro de todos los posibles del rango. Para este I_o , se pueden discriminar unos 50 valores distintos de intensidad luminosa. Por tanto, habrá un nivel por debajo del cual no se distingan otros niveles, que se verían como negro y de igual forma sucederá con la percepción del blanco. A medida que el ojo recorre otros niveles I_o' del rango total, va definiendo niveles de negro y blanco de forma análoga.

La adaptación proporciona la capacidad de ver en condiciones de muy distinta iluminación. A pesar de ello, hay que tener en cuenta que la percepción visual va a variar mucho dependiendo de la iluminación con la que nos encontramos. Así, como ya se ha comentado, en condiciones de baja iluminación se detectan muy bien las diferencias de luminosidad, pero la distinción del color y del detalle es pobre y sucederá

justo al contrario en condiciones de elevada luminosidad. Además, el proceso de adaptación que realiza el SVH no es instantáneo y es más costosa temporalmente la adaptación de luz a oscuridad.

En cuanto a la relación entre el estímulo percibido y el valor real del mismo se obtiene una respuesta no lineal. Para calcularla se considera una zona iluminada con una intensidad I y otra con $I + \Delta I$, como muestra la figura 2.9 a. Se va a medir el valor necesario de ΔI para que el observador vea dos zonas diferenciadas partiendo de un ΔI inicial igual a cero. El valor así obtenido de ΔI es la mínima diferencia notable (JND). Este experimento se realiza sobre un amplio margen de intensidades (en visión fotópica, con valores sobre 10 cd/m^2) dando como resultado la curva de la figura 2.9 b. Se observa una relación $\frac{\Delta I}{I}$ constante y de valor 0.02 que no depende del nivel de intensidad en el que se trabaja.

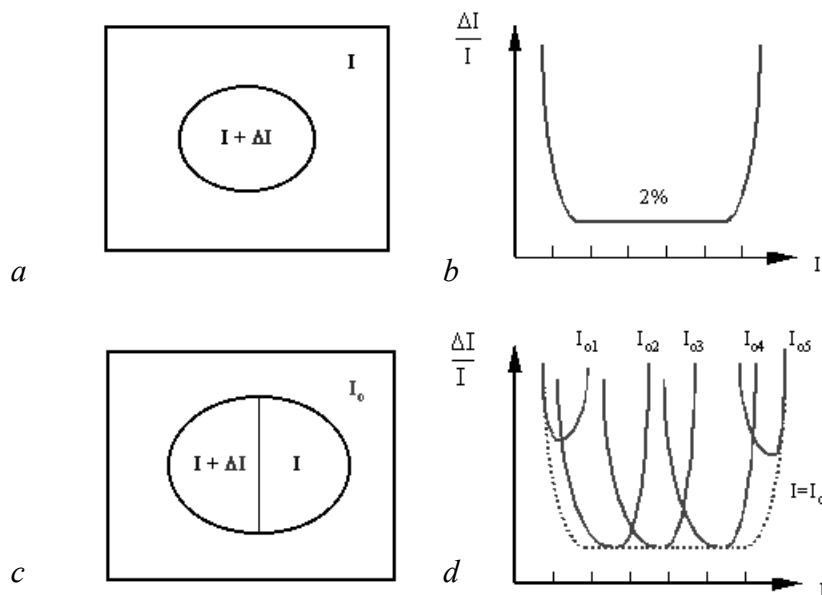


Figura 2.9: Medida de la respuesta del SVH a cambios de intensidad.

Si nos planteamos la misma medida pero para las condiciones indicadas por la figura 2.9 c. obtenemos curvas de respuesta que son constantes en un margen mucho menor pero que para un valor fijado I_{oi} es comparable al margen dinámico de muchos sistemas de procesado electrónico de imágenes.

A la relación $\frac{\Delta I}{I} = cte$ se le conoce como *ley de Weber* y se verifica también para otros sentidos como el oído, el olfato, el gusto y el tacto. Este resultado es de especial interés en el estudio de la percepción de ruido en imágenes. Basándonos en él podemos explicar porqué un nivel uniforme de ruido aleatorio es menos visible en las regiones más luminosas de la imagen. Así, para que el ruido sea perceptible con un fondo de imagen de nivel de intensidad luminosa I_1 será preciso un valor ΔI_1 y para un $I_2 < I_1$ entonces el valor $\Delta I_2 < \Delta I_1$. Con un nivel uniforme de ruido, este será más visible en la zona de luminosidad I_2 que en la I_1 . En conclusión, para una imagen dada, *se percibe más el ruido uniforme en las regiones oscuras*.

Para valores de luminancia menores de 10 cd/m^2 la ley de Weber deja de cumplirse y tendremos un ΔI proporcional a la raíz cuadrada de I que se conoce como relación de *De Vries-Rose*. Para valores todavía menores de luminancia (1cd/m^2) existe una relación lineal.

2.2.2. Contraste.

Por la descripción fisiológica previa del SVH se sabe que los mecanismos básicos de la visión trabajan con el contraste. Experimentos psicofísicos que evidencian los fenómenos visuales relativos a la percepción de contraste pueden verse en la figura 2.10. En la figura de las *bandas de Mach* se muestra el fenómeno de inhibición lateral que se produce en los campos receptores y a pesar de que cada barra representa un nivel constante se aprecian diferencias subjetivas en los bordes de separación de las bandas. Se puede observar también un ejemplo de contraste simultáneo, en el que se muestra un rectángulo interior del mismo nivel de intensidad pero que se percibe más claro u oscuro según el nivel de luminosidad del rectángulo exterior.

Por tanto, como el SVH es sensible al contraste y no a la luminancia en términos absolutos es de gran importancia la definición que se haga del contraste. Lo que podría parecer en un primer momento sencillo es, en realidad, bastante complicado y su dificultad aumenta cuando se quiere definir el contraste para imágenes complejas. El objetivo sería encontrar una medida de contraste que se corresponda con la percepción visual que se tiene del mismo.

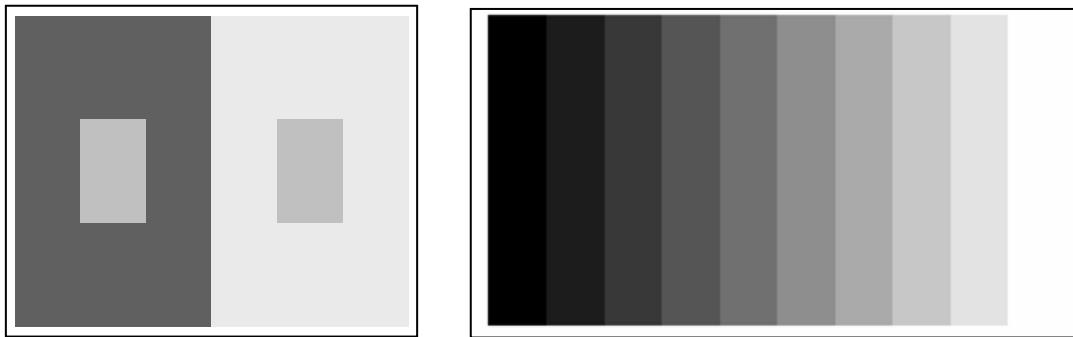


Figura 2.10: Fenómenos relativos al procesamiento de contraste. **a.** Bandas de Mach. **b.** Contraste simultáneo.

El problema se ha resuelto en primer lugar con imágenes simples (patrones) como por ejemplo funciones sinusoidales o discos de luminancia. Para este tipo de imágenes se utilizan dos definiciones de contraste diferentes. Por una parte, Michelson definió el contraste de este tipo de imágenes como:

$$C_{\text{Michelson}} = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}}$$

donde L_{\max} y L_{\min} son los valores máximos y mínimos de luminancia. Weber propone otra definición para medir el contraste local de un objeto simple (blanco) sobre un fondo uniforme:

$$C_{\text{Weber}} = \frac{\Delta L}{L}$$

donde ΔL es el incremento o decremento de la luminancia del blanco comparada con la luminancia del fondo L .

La diferencia entre ambas definiciones es clara y se hace más obvia si expresamos el contraste de Michelson en la siguiente forma:

$$C_{\text{Michelson}} = \frac{\Delta L}{L + \Delta L}$$

donde $\Delta L = (L_{\max} - L_{\min})/2$ y $L = L_{\min}$. Estas dos medidas de contraste no coinciden a pesar de que pueden proporcionar valores parecidos para estímulos simples de bajo contraste y ambas definen el contraste como un radio adimensional de la variación de luminancia respecto a la luminancia media de fondo. Sin embargo para contrastes más elevados obtenemos dos medidas de contraste diferentes y lo que es más, estas dos

definiciones ni siquiera comparten el mismo rango. Mientras que el $C_{Michelson}$ tiene un margen de valores entre 0 y 1 el C_{Weber} varía entre -1 e ∞ .

Además, se plantean otros problemas derivados del hecho de que tanto el $C_{Michelson}$ como el C_{Weber} se definen para imágenes simples y de forma global. En imágenes más complejas sería más apropiada una definición local del contraste puesto que la luminancia (y con ella también el contraste) varía en toda la imagen. Por tanto, las fórmulas anteriores no serán válidas para el cálculo de contraste en imágenes complejas como las que se tratan en cualquier modelo de visión. Estas medidas tampoco tienen en cuenta que la sensibilidad al contraste del SVH es dependiente de la frecuencia, sobre todo si nos encontramos en condiciones umbrales de visibilidad.

Para solucionar estos problemas Peli [23] propone un cálculo del contraste por bandas de frecuencia. Es un contraste local y limitado en banda que tiene en cuenta el nivel medio de luminancia local para la obtención de un valor en cada punto. Puesto que el cálculo del contraste se efectúa por bandas es necesario disponer primero de una versión filtrada de la imagen. Más adelante se tendrán en cuenta las distintas implementaciones que se pueden hacer al realizar este banco de filtros y la importancia que pueda tener esta elección. El contraste en una banda frecuencial k determinada se representa por:

$$C_{Peli_k}(i,j) = \frac{b_k(i,j)}{l(i,j)}$$

donde $b_k(i,j)$ es la imagen filtrada para una determinada banda k y $l(i,j)$ es una versión paso bajo de la imagen que contiene toda la energía por debajo de la banda k en la que se está realizando el cálculo del contraste. Obtenemos de esta forma un valor de contraste que no sólo depende de la energía en esa banda sino también de la luminancia de fondo *local*. La fórmula así planteada no tendría en cuenta la distinta sensibilidad al contraste del SVH dependiendo de la orientación, por lo que se podría ampliar sin más que hacer un filtrado de la imagen para distintas frecuencias y orientaciones.

Otras posibles métricas de contraste para imágenes complejas realizan una modificación del contraste de Michelson adaptándolo al filtrado previo de la imagen como se muestra en la siguiente ecuación:

$$C_k(i,j) = \frac{b_k(i,j) - \bar{b}_k}{\bar{b}_k}$$

donde $b_k(i,j)$ es la imagen filtrada para una banda k y \bar{b}_k es la media de esa banda. El problema que plantea esta redefinición del contraste es que esta media es cero para todas las bandas de frecuencia excepto para la banda base. Por ello, se deben hacer algunas modificaciones a esta fórmula. Daly [7] propone un cálculo global y otro local del contraste:

$$C_{Global_k}(i,j) = \frac{b_k(i,j)}{\bar{b}_K} \quad C_{Local_k}(i,j) = \frac{b_k(i,j)}{b_K(i,j)}$$

donde $b_K(i,j)$ es la banda base y \bar{b}_K es la media de la misma.

El contraste en un píxel de una imagen determinada también puede calcularse en función de la luminancia de los píxeles que hay en su vecindario. Esta técnica [29] considera los valores máximos y mínimos de luminancia en los píxeles vecinos a la izquierda y derecha del píxel para el que se calcula el contraste. Tanto en este caso, como en todos los anteriores, se debe tener en cuenta que la definición que se hace del contraste debe ser consecuente con la implementación del modelo visual que se esté llevando a cabo así como con el objetivo que va a tener este modelado.

Finalmente, una vez expuesta la variedad de medidas de contraste de las que se dispone, quedaría por preguntarse cuál de ellas tiene una mayor correspondencia con la percepción del contraste del SVH. Para ello se hace necesario el uso de test psicofísicos que evalúen esta relación. Este tipo de comparativa se estudió en [24] entre las métricas propuestas por Michelson, King-Smith y Kulikowski ($C_{KK} = (l_{max} - l_{fondo}) / l_{fondo}$) y una modificación del contraste de Peli sugerida por Lubin. Los patrones que se utilizaron son Gabores y el resultado de los tests realizados mantiene que la fórmula de Peli para el cálculo del contraste es la que mejor se adapta a los resultados subjetivos.

2.2.3. Sensibilidad en frecuencia.

El estudio de la anatomía y fisiología del SVH ha permitido que se conozca que la percepción del estímulo depende de la frecuencia del mismo. Tradicionalmente una de las caracterizaciones más importantes que se hacen del SVH consiste en la máxima

frecuencia que es capaz de detectar o diferenciar y que se denomina agudeza visual. Este dato proporciona únicamente un límite y es insuficiente si se quiere conocer la variación en la detectabilidad de un estímulo en función de su frecuencia espacial. Con este objetivo se empezó a medir la función de sensibilidad al contraste (CSF) que refleja la sensibilidad o capacidad de detección del SVH a estímulos de distinta frecuencia. La CSF fue determinada por primera vez en 1956 por Schade pero su uso no se generalizó hasta que las técnicas de Fourier empiezan a utilizarse en visión en los años 70.

Puesto que lo que se pretende medir es la respuesta del SVH a distintas frecuencias, el procedimiento de medida consistiría en mantener un estímulo de contraste constante y variarlo en frecuencia para ver como el SVH atenúa cada frecuencia. Sin embargo, esto resulta imposible puesto que no podemos medir el contraste de la imagen percibida por lo que es necesario realizar el procedimiento justo al contrario. La medida se realiza para cada frecuencia variando el contraste del estímulo de entrada y lo que mantenemos constante es la salida, es decir, la imagen percibida.

Un procedimiento experimental para calcular la CSF consistiría, por ejemplo, en ir reduciendo el contraste de un estímulo sinusoidal en el que la luminancia media se mantiene constante hasta que se alcanza el umbral (el estímulo deja de ser visible). El inverso de este valor umbral es el valor de la sensibilidad para esa frecuencia. Al realizar este proceso para las diferentes frecuencias obtenemos la CSF. Se pueden hacer varias consideraciones sobre este método de medida. En primer lugar, la CSF así calculada sería válida para estímulos simples, no para imágenes complejas. Aunque cualquier estímulo complejo pueda ser analizado como una serie de estímulos sinusoidales, el problema reside en que el SVH es no lineal y por ello, la respuesta obtenida varía y no se puede calcular como una combinación lineal. En segundo lugar, este método es de cálculo de umbrales de detección y es válido sólo en condiciones umbral, la forma de la CSF para condiciones por encima del umbral no se corresponde con esta medida. Para condiciones por encima del umbral habría que llevar a cabo otro tipo de tests, con el inconveniente de que son experimentos menos estables y de mayor dificultad de medida, por lo que no suelen utilizarse.

Existen muchas medidas experimentales de la CSF, la mayoría de ellas para estímulos monocromáticos. En [22] se dispone de un resumen de algunas de las CSF's

existentes y los parámetros que se han tenido en cuenta en la medida. La forma típica de la CSF puede observarse en la figura 2.11 y es la de un filtro paso banda con un pico que se sitúa entre los 4 y 8 ciclos/grado para niveles de iluminación fotópicos. Se produce una fuerte atenuación en altas frecuencias debida a factores ópticos como imperfecciones en la córnea y cristalina, difracciones en la pupila... Además, el muestreo espacial que realizan los fotorreceptores de la retina impone un límite a partir del cual no se detectan frecuencias espaciales mayores (agudeza visual). La atenuación que aparece en bajas frecuencias se debe a las interacciones inhibitorias de centro-periferia que se producen en los campos receptores. La medida del valor de la CSF en frecuencia cero es imposible para el método de medida que se ha explicado con anterioridad y según los parámetros que se hayan escogido al realizar los experimentos para la obtención de la CSF se dará una frecuencia mínima a partir de la cual se empieza a medir.

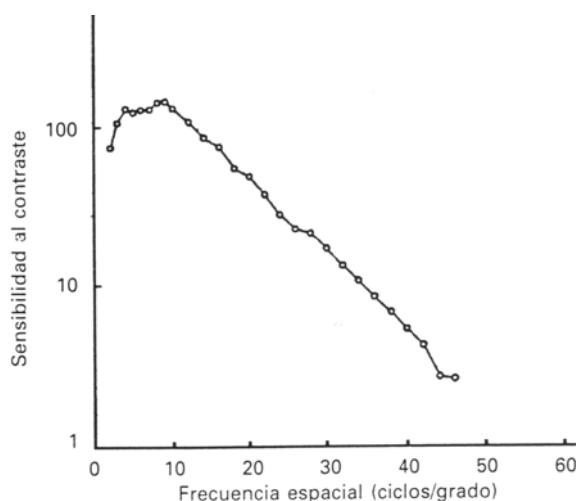


Figura 2.11: CSF del sistema visual (Campbell & Green, 1965).

La forma que la CSF adopta depende de muy diversos parámetros. Por una parte habría que considerar los parámetros del estímulo que se presenta para la medida y las condiciones en las que esta se realiza y por otra los que son inherentes al sistema visual. La descripción e importancia de los mismos ha sido analizada de diversa manera en muchos estudios [1, 22, 26] y a continuación se enumeran brevemente comentando cómo afectan a la CSF.

- *Número de ciclos presentes en el test:* este parámetro afecta a la zona de bajas frecuencias. Se ha comprobado que existe un número de ciclos críticos por debajo del cual la visibilidad del estímulo aumenta si aumentamos el número de ciclos que se le

presentan al observador y este efecto no dependerá de la frecuencia espacial. Por tanto, puesto que el tamaño del test suele ser fijo para todas las frecuencias que se miden en el experimento el problema se planteará precisamente en bajas frecuencias donde no se representa el suficiente número de ciclos para que la medida sea independiente de este efecto. Para corregirlo se debe variar el tamaño del test en bajas frecuencias para que el número de periodos de la señal representada esté por encima de este valor crítico, al hacerlo se obtienen curvas de CSF con característica paso bajo en lugar de paso banda.

- *Luminancia*: influye notablemente en el aspecto de la curva. La sensibilidad aumenta al aumentar la luminancia media, el pico de la CSF se desplaza hacia frecuencias mayores y la agudeza visual es mayor. Así, con niveles de iluminación bajos (menores de 10 cd/m^2), la curva se acerca más a un paso bajo con un pico de sensibilidad en torno a 1 ó 2 ciclos/grado y para niveles mayores de luminosidad presenta una característica paso banda con el pico alrededor de 6 ciclos/grado.
- *Tamaño del test*: si se incrementa el tamaño la sensibilidad aumentará. Esto sucede hasta que se alcanza un valor crítico a partir del cual aunque aumentemos más el tamaño del test no se obtiene una mayor sensibilidad (considerando que hay un número de ciclos suficientes para que la medida sea correcta). El tamaño crítico dependerá de la frecuencia [22] y se ha demostrado que no tiene efecto sobre la medida para valores de contraste por encima de $0'1$. Además, la extensión espacial del estímulo dentro del área de medida (ocupación total o parcial de la misma) modifica también la forma de la curva [26].
- *Orientación*: la CSF es una función anisótropa. La sensibilidad es mayor para estímulos orientados horizontalmente o verticalmente que para orientaciones oblicuas (45°). Este efecto oblicuo no repercute en bajas frecuencias y se hace patente para valores de frecuencia espacial mayores de 10 ciclos/grado típicamente.
- *Excentricidad*: como norma general la CSF suele determinarse para visión en la fóvea, donde la sensibilidad es mayor. A medida que aumenta la excentricidad la sensibilidad al contraste decrece y este decrecimiento es más rápido cuanto mayor es la frecuencia espacial del estímulo.
- *Color*: si calculamos la CSF para el rojo, verde y azul obtenemos tres curvas diferenciadas y que son distintas de la CSF acromática.

- *Tiempo de exposición*: cuanto mayor sea el tiempo del que dispone el observador para determinar el umbral, la sensibilidad obtenida será mayor. Esto sucede hasta que se alcanza un valor (unos cientos de msg) a partir del cual ya no surge este efecto.
- *Presentación temporal del estímulo*: para presentaciones graduales del estímulo la CSF presenta una característica más paso banda y con un valor de pico algo mayor que en presentaciones del estímulo abruptas
- *Frecuencia temporal*: lleva a un concepto de CSF espacio-temporal. La CSF espacial es mayor para frecuencias temporales medias (3 a 10 ciclos/sg) y decrece a ambos lados.
- *Parámetros inherentes al observador*: son la posición de mejor enfoque del ojo, el tamaño de la pupila y la edad. Al variarlos estamos analizando sistemas diferentes y por tanto obtendremos curvas de sensibilidad distintas.

Otro parámetro de importancia en el cálculo de la CSF es la distancia de observación para la que se realizan las medidas, ya que esto va a fijar la frecuencia espacial del estímulo, expresada en unidades de ciclos por grado. Si el observador se aleja de la imagen, un periodo del estímulo va a corresponder a un ángulo visual menor y por tanto la frecuencia espacial representada será mayor sin que se haya modificado en nada el estímulo presentado. Por tanto, las medidas que se toman de la CSF se hacen para una distancia de observación determinada.

Es necesario destacar que aunque se ha expuesto únicamente y de forma muy sencilla un método de medida de la CSF umbral, éste, por supuesto, no es el único método de medida. Existe una gran variedad de métodos entre los que se encuentran los tests impresos [1], que aunque no son muy precisos tienen como ventaja su gran sencillez. Además, dentro de la medida de CSF umbral se pueden seguir distintos métodos psicofísicos. En definitiva, la elección de uno u otro método dependerá de la aplicación en la que se vaya a utilizar la CSF y de la precisión o las características especiales que se requieran para la misma. Así, un posible uso de la CSF sería, por ejemplo, la detección y seguimiento de algunas afecciones visuales que tienen un efecto determinado en la forma de la curva, y que presenta como ventaja que es un método no invasivo y de detección temprana. En este caso bastaría con medir la CSF para el paciente bajo observación. Otra posible utilización de la CSF, sería dentro de un

modelado del SVH genérico, lo que reviste una mayor complejidad puesto que no existe una CSF “universal”. Más adelante, en la aplicación de la CSF dentro de modelos visuales para el análisis de la calidad se verán algunos de los problemas que se plantean derivados de este hecho.

2.2.4. Enmascaramiento.

El enmascaramiento es el fenómeno por el cual disminuye la visibilidad de una señal en presencia de otra que esconde o enmascara a la primera. No se debe olvidar que el método utilizado para la obtención de los umbrales de visibilidad trabaja con estímulos muy simples de una sola frecuencia y sobre campos de intensidad constante. Desde este punto de partida, el enmascaramiento es un paso más de complejidad en el modelado del SVH que ha sido observado y estudiado ampliamente por fisiólogos y psicofísicos.

El objetivo principal será conocer cómo varían los umbrales de visibilidad del estímulo cuando se encuentran en el entorno cambios espaciales y también temporales de luminancia. Se dan entonces dos fenómenos de enmascaramiento claramente diferenciados:

- *Enmascaramiento temporal*: asociado a la reducción visual que se produce sobre el estímulo cuando está rodeado de cambios temporales de luminancia. El estudio de este tipo de enmascaramiento es complicado debido a que habría que tener en cuenta las características de movimiento del ojo, es decir, hay una dependencia con la velocidad de movimiento retiniana. Además, habría que considerar si el estímulo en movimiento es seguido o no por el ojo, lo que depende del interés particular del observador, siendo esto muy difícil de analizar.

Por otra parte, está demostrado que cuando hay un cambio de escena, la resolución espacial puede ser reducida drásticamente sin que se llegue a percibir siempre que se restablezca la resolución original en un periodo breve de tiempo (en torno a 100 msg).

- *Enmascaramiento espacial*: está asociado a los cambios espaciales de luminancia que se producen alrededor del estímulo. Recibe también el nombre de enmascaramiento de contraste ya que el efecto del enmascaramiento será menor en regiones uniformes que en zonas de gran contraste. Así, si se considera cualquier imagen natural a la que se le añade un ruido aleatorio uniforme, la visibilidad del ruido en los bordes y texturas que presenta la imagen será menor que en zonas uniformes.

El fenómeno del enmascaramiento espacial se estudia mediante test psicofísicos en los que se evalúa la influencia del contraste, de la frecuencia y de la orientación tanto del estímulo (enmascarado) como de la señal de fondo (máscara). Las señales utilizadas suelen ser sinusoides y Gabores. Estos experimentos han demostrado que el umbral de detección que proporcionaba la CSF va incrementando a medida que el efecto del enmascaramiento se hace mayor, lo que ocurre cuando el contraste de la señal que enmascara se aumenta.

El comportamiento típico se muestra en la figura 2.12. C_T representa el valor de contraste para que el estímulo sea visible y C_M el contraste de la señal que enmascara. Para un valor pequeño de C_M , el umbral de visibilidad del estímulo C_T viene dado por la CSF y se representa como C_0 . Este valor se mantiene hasta que se alcanza un punto en el que $C_M = C_0$. A partir de ahí un incremento en C_M hace que se eleve el umbral de detección, es decir, se necesitaría aumentar el contraste del estímulo para que fuera visible. La línea discontinua de la figura ha sido observada en algunos de los experimentos psicofísicos y proporciona una medida por debajo del valor dado por la CSF, sin embargo, este efecto sólo se observa cuando la señal y el ruido están en fase.

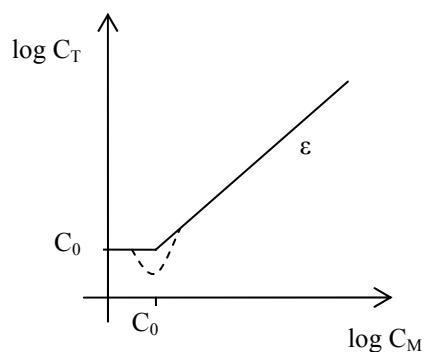


Figura 2.12: Curva de umbrales de elevación causados por enmascaramiento espacial.

El valor de la pendiente ϵ depende tanto de la complejidad y conocimiento del estímulo que se tengan, como del tipo de señal utilizada para el enmascaramiento. Si se conocen datos de la orientación, tamaño, frecuencia y localización del estímulo será más fácil reconocerlo y por tanto la pendiente de la curva será menor. El valor de ϵ varía entre $0'6$ y 1 dependiendo de que el tipo de señal utilizada para el enmascaramiento sea una sinusoide o un ruido limitado en banda. Además, aparece un efecto de aprendizaje. Si a un observador se le muestra el mismo patrón de ruido repetidas veces, el valor de ϵ decrece puesto que será capaz de reconocerlo con un valor menor de contraste.

La frecuencia espacial y la orientación también influyen en el grado de enmascaramiento, que será máximo cuando se produzca una coincidencia entre la frecuencia de la señal que enmascara y el estímulo enmascarado

2.3. Movimientos oculares y atención visual.

La cantidad de información con la que debe tratar el SVH y que se encuentra en el entorno visual hace que sean necesarios mecanismos selectores de información. Así, la atención puede ser considerada como un conjunto de redes de áreas neuronales que llevan a cabo operaciones específicas de procesamiento de información. Dentro de estas operaciones específicas se encontraría la detección o selección de objetivos dentro de la escena, que se corresponderá con determinados movimientos oculares.

Si se analiza la forma que tienen los ojos de explorar el entorno visual se van a encontrar tres tipos de movimientos oculares, atendiendo a la funcionalidad de los mismos:

- 1) Movimientos para el mantenimiento de la mirada: son aquellos que compensan el movimiento de la cabeza o de los objetos para que permanezca la mirada fija sobre el objeto.
- 2) Movimientos para el desplazamiento de la mirada: permiten pasar la atención de un objeto a otro. Fundamentalmente, se dan tres tipos: sacádicos, persecuciones o de seguimiento y vergencias.
- 3) Movimientos de fijación o micromovimientos: trémores, microsacádicos y fluctuaciones.

Si se considera el proceso seguido al visualizar una imagen estática se observa un primer paso en el que se fija la posición de un objeto o detalle por un breve periodo de tiempo (100-500 msg) para, posteriormente, realizar un salto a una nueva posición, este salto es el que se ha denominado movimiento sacádico. Por tanto, la exploración visual que se hace de una escena estática no es un movimiento continuo como cabría esperar, sino una sucesión de saltos de un punto de interés a otro. Además, cuando se tiene fijada una posición los ojos no permanecen quietos, se realizan movimientos muy pequeños y constantes, microsacádicos. Estos micromovimientos son continuados y de dirección aleatoria y su importancia radica en que permiten que se vean los objetos estacionarios. De hecho si se mantuviese una imagen estable en la retina, en vez de obtener una mayor agudeza visual (se evitaría el emborronamiento asociado al movimiento de la imagen que producen los movimientos oculares), lo que sucede es que los receptores de la retina se saturan y la imagen desaparece.

La particular forma de ver imágenes estáticas dando saltos de un punto de la imagen a otra hace que se ponga de manifiesto la importancia que tiene el estudio de la atención visual. De hecho, aunque es posible prestar atención a una parte de la escena sin que se produzca un movimiento ocular, el caso contrario es imposible: cualquier movimiento sacádico viene precedido por un cambio de atención a esa posición concreta. Así, parece poco probable que en la visualización de una imagen, se haga un salto al vacío, es decir, en una sacada se fije un punto que corresponde con fondo o con un punto que carece de cambios abruptos de luminancia.

El estudio de la atención visual controlando los puntos de parada de la visión resulta de gran interés y proporciona gran cantidad de información acerca de los objetos y detalles que son relevantes para el observador. Los factores que influyen en la atención visual [26] son muchos y variados, entre ellos se pueden destacar el tamaño, color, contraste, forma, bordes, texturas y por supuesto el movimiento.

Se han descrito brevemente los movimientos oculares y su relación con el proceso de atención y los factores que influyen en el mismo ya que su incorporación a los modelos del SVH puede resultar de gran interés en su aplicación a métricas de calidad perceptible.

3

Evaluación de la calidad en imágenes.

En este capítulo se hará un breve repaso de algunos de los modelos existentes de medida de calidad de imagen, comentando su estructura básica y los distintos problemas y ventajas que llevan asociados. Es necesario, para ello, definir previamente algunos conceptos elementales que sentarán las bases sobre las que se han construido estos modelos de medida. Se llevará a cabo, por tanto, una primera serie de definiciones y clasificaciones de los modelos, que después será desarrollada en posteriores apartados con más detalle. En cualquier caso, no es un análisis en profundidad de los diferentes modelos, sino una explicación breve con los detalles de mayor relevancia dentro de cada método particular. Se proporciona así un marco general en el que se circunscribe este proyecto y para una compresión más completa de los diferentes modelos que existen en la literatura se debe consultar la bibliografía.

Por otra parte, se pretende dar una visión general en cuanto al tipo de distorsiones con las que van a tratar estos métodos, haciendo especial hincapié en la compresión de imágenes y las técnicas de marcado de agua. Se va a explicar en qué consisten estas técnicas y cómo afectan a la imagen en cuanto a los artefactos que se generan. Esto es necesario no sólo porque en este proyecto se van a tratar precisamente este tipo de imágenes, sino también por la relevancia que tienen estándares como JPEG o JPEG2000 y la importancia que están adquiriendo en ámbitos diversos las marcas de agua.

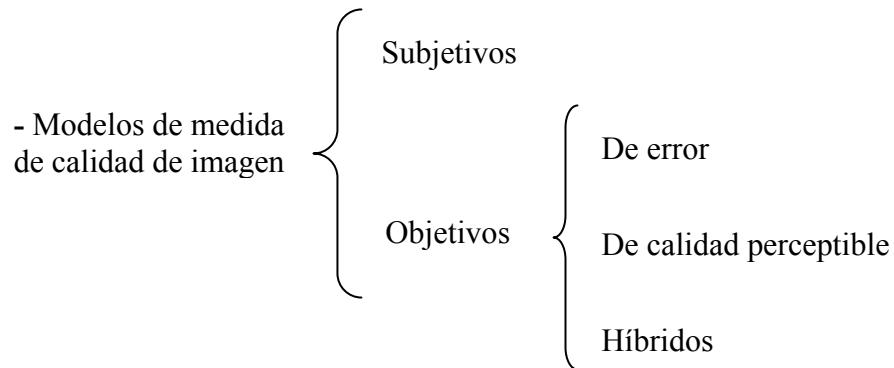
3.1. Calidad de imagen. Clasificación de métodos de medida.

En primer lugar, se debe establecer una definición de *calidad de imagen*. Es imprescindible comprender lo que significa este término para el observador ya que la forma de diseñar los modelos de medida se debe corresponder previamente con lo que se entiende por calidad. Se define calidad como propiedad o conjunto de propiedades inherentes a una cosa. Este concepto, llevado a imágenes, se entendería como la capacidad que una imagen tiene de representar el objeto original, es decir, la exactitud o parecido entre ambos. Dentro de los modelos de medida de calidad, en lugar de tener una imagen y un objeto, la definición se extiende a dos imágenes de las cuales una es considerada la de referencia u original (hace las veces de objeto) y otra sobre la que se ha efectuado alguna operación de compresión, marcado de agua, etc. En este caso, la *medida de calidad de imagen* es una medida de semejanza entre la imagen original y la distorsionada. Por tanto, para poder evaluar la calidad de una imagen en todos los modelos que aparecen en este capítulo será necesario disponer de la imagen original. El desarrollo de modelos de medida de calidad en los que no exista imagen de referencia, o sólo exista parcialmente, es realmente difícil y está fuera del contexto de este proyecto, aunque el sistema visual sea capaz de ello.

La definición de calidad de imagen que se ha dado corresponde con lo que se conoce comúnmente como *fidelidad de imagen* y presenta también algunos inconvenientes. En realidad, no está claro que la visibilidad del error esté relacionada con pérdida de calidad y de hecho, algunas distorsiones que pueden ser claramente visibles no son, a pesar de ello, molestas al observador. Por ejemplo, si se considera la medida de calidad entre dos imágenes, una de las cuales es la multiplicación de los valores de luminancia de la otra imagen por un factor global, la diferencia visual entre ambas será obvia aunque el observador no aprecie tal diferencia como pérdida de calidad.

En segundo lugar, en cuanto a la clasificación de los modelos de medida, se van a diferenciar dos grupos principales: métricas de calidad de imagen subjetivas (utilizando observadores) y objetivas (medidas matemáticas). Dentro de las objetivas se hará una nueva clasificación atendiendo únicamente a si se incorpora o no el SVH

descrito previamente en el capítulo 2. Así, la clasificación resultante quedará como muestra el siguiente esquema:



Los *modelos de medida de error* obtienen una medida de calidad sin tener en cuenta las características especiales del SVH. Esta sería una medida objetiva, relativa a la diferencia entre las imágenes en sí y no a nuestro modo de percibir. En un segundo grupo se incluyen aquellos que incorporan en mayor o menor medida las características propias del SVH y denominamos a este segundo grupo *modelos de medida de calidad perceptible*. Finalmente, se hace un tercer grupo de *modelos híbridos de medida de calidad*, en el que se describen algunos métodos de medida que a pesar de no incluir un modelo de SVH tratan de obtener valores que se adecuen a la calidad que el observador percibe explotando otras cualidades relativas a la percepción.

Por último, se van a enumerar algunas de las características o propiedades de los modelos de calidad sobre las que se puede evaluar la bondad del método seleccionado. Entre ellas se pueden destacar las siguientes:

- Velocidad: es deseable que el valor de calidad que resulta de la utilización del método se obtenga de forma rápida. Esto adquiere un especial interés cuando se va a hacer uso de la medida de calidad para mejorar los procesos o algoritmos de compresión, cuantificación...
- Coste: el computacional depende de la velocidad y de la complejidad de los algoritmos para la obtención de un resultado. Además, se deben considerar otros costes adicionales ya que para la validación o evaluación del método es necesario llevar a cabo test psicofísicos que tienen una serie de requisitos temporales y también en cuanto a número y características de los observadores.

- Complejidad: está estrechamente ligada con la velocidad y el coste. Lo ideal sería encontrar un método lo más sencillo posible que diese una medida de calidad igual a la percibida. En general, resultados más próximos a los obtenidos por un observador se consiguen al incorporar características propias del SVH, lo que forzosamente implica que el método sea complejo.
- Portabilidad: los resultados que proporciona el método no deben alterarse si se repiten las medidas en diferentes entornos o tiempos.
- Precisión: referida a como representa el resultado del método de medida la percepción de calidad que tendría el observador.
- Robustez: se pretende obtener resultados válidos sobre un amplio margen de variación de los parámetros asociados a la medida (tipo de imagen, tipo de distorsión, condiciones de visibilidad, etc.), es decir, se buscan métodos robustos.
- Forma del resultado: pueden ser valores numéricos (índices de calidad) o mapas de visibilidad del error. Según la aplicación a la que esté destinado un método de medida será conveniente una forma u otra y en general, lo ideal es que el método pueda proporcionar ambas salidas.

3.2. Modelos de calidad de imagen subjetivos.

Los modelos de medida subjetivos son aquellos en los que el valor de calidad es evaluado directamente por un observador al que se le muestran las imágenes. En la recomendación ITU-R BT.500-10 (“Methodology for the subjective assessment of the quality of television pictures”) [17] se regulan algunos de los test para la evaluación subjetiva de imágenes estáticas y de video. La norma incluye criterios para la selección del material de test y de los observadores, las condiciones de visibilidad, los procedimientos de evaluación y el análisis de los datos. Los tres procedimientos de evaluación más comunes son:

- Escala de artefactos por doble estímulo (**DSIS**). Se muestra a los observadores la imagen o secuencia original seguida de la distorsionada una sola vez. Se pide a los observadores que califiquen la imagen o secuencia que presenta los artefactos con respecto a la original dentro de una escala discreta compuesta por cinco valores:

imperceptible, perceptible pero no molesto, ligeramente molesto, molesto y muy molesto.

- Doble estímulo y escala continua de calidad (**DSCQS**). Se muestra a los observadores dos imágenes o secuencias, una de ellas de referencia y la otra test. El orden en el que aparecen es aleatorio, no sabiendo el observador cual de ellas es la referencia. Además, el par de imágenes o secuencias se presenta varias veces antes de que el observador emita su valoración. Ésta consiste en puntuar ambas imágenes dentro de una escala continua dividida en secciones a las que se les adjudica los calificativos: excelente, buena, justa, pobre y mala. La escala numérica equivalente es de 0 a 100 y se computan las diferencias dentro de cada par de imágenes evaluadas. Esta diferencia hace que se elimine parte de la subjetividad referida a la experiencia y al contenido de la imagen.
- Estímulo simple y evaluación de calidad continua (**SSCQE**). Se muestra a los observadores un programa de entre 20 y 30 minutos de duración que ha sido procesado por el sistema bajo test. La referencia no se muestra y se evalúa la calidad instantánea percibida de forma continua en una escala DSCQS. Esta técnica se introduce para evaluar imágenes en las que la calidad varía con el tiempo y depende de la escena (compresión digital de video).

La utilización de uno u otro método depende del problema que se pretende evaluar. La norma muestra una tabla con algunos de los problemas más usuales y los test que llevarían asociados. Así, el método DSIS es usado para medir la robustez de un sistema, es decir, para evaluar distorsiones visibles como pueden ser los artefactos causados por errores de transmisión. El método de DSCQS evalúa un sistema con respecto a otro de referencia y funciona bien para calidades similares ya que es bastante sensitivo a las pequeñas diferencias. El SSCQE sirve, por ejemplo, para medir la fidelidad entre dos secuencias de video distorsionadas.

Una de las principales características de estos métodos es que son los que dan un valor de calidad más preciso en la actualidad puesto que éste se saca directamente de la opinión del observador. Sin embargo, la realización de este tipo de test lleva asociado diversos inconvenientes y se debe tener en cuenta que en los datos obtenidos siempre habrá un cierto grado de error. A continuación se enumeran varios de los problemas que plantea su utilización:

- Suponen un coste elevado en tiempo y personas. Conseguir un conjunto de resultados aceptable puede llevar varias semanas. Primero se debe acondicionar el lugar donde se vayan a efectuar las medidas de acuerdo a la norma y después se deben realizar los test el mayor número de veces posible para conseguir buenos resultados. Es decir, el número de sujetos y las cualidades o capacidades de estos (edad, profesión, experiencia...) influyen en las puntuaciones de calidad dadas y eliminar en lo posible esta dependencia pasa por elegir un conjunto grande y heterogéneo de observadores.
- La recomendación ITU-R BT.500-10 expone otro de los problemas que aparecen y que denomina efecto de contexto. Se debe al orden e intensidad de las distorsiones que se presentan. Así, si después de varias secuencias o imágenes con distorsiones pequeñas se presenta una de gran distorsión el observador va a puntuarla por debajo de lo que normalmente lo haría.
- La escala de puntuación también influye. Escalas discretas introducen una aproximación que debe ser compensada con un mayor número de sujetos para reducir la varianza. Escalas fijas van a propiciar que los observadores no utilicen demasiado los valores extremos.
- Otro gran inconveniente de estos métodos es que se efectúan para unas condiciones de visibilidad determinadas y bajo otras diferentes no son válidos, siendo necesario repetir los test en las nuevas condiciones.
- No proporcionan información espacial o temporal de calidad. El observador da un valor global de calidad pero no evalúa dónde o cuándo se produce este error, datos que pueden resultar de gran utilidad para diseñar codificadores, métodos de marcado de agua, etc.

Finalmente, este tipo de métodos subjetivos a pesar de que sus resultados no sean perfectos y que su realización plantea bastantes problemas son necesarios para comprobar la precisión de los modelos de calidad objetivos. Es decir, debe existir correlación entre los resultados obtenidos mediante el uso de modelos subjetivos y los que se obtienen de los objetivos.

3.3. Modelos de calidad de imagen objetivos.

Son aquellos que proporcionan un valor objetivo de la calidad de una imagen mediante una serie de operaciones matemáticas en las que pueden incluirse o no funcionalidades propias del SVH. El desarrollo de este tipo de métodos es necesario para un gran número de aplicaciones y en general pueden utilizarse para la optimización de algoritmos en procesamiento de imágenes. Una de sus principales ventajas frente a los métodos subjetivos es que dan una medida de forma más o menos inmediata y por ello sirven para monitorizar de forma dinámica la calidad de la imagen.

La clasificación de modelos dentro de los métodos objetivos para la evaluación de la calidad se basa en la incorporación que se hace de modelos visuales en los mismos y ya ha sido explicada en apartados anteriores. A continuación se van a exponer las principales características y limitaciones de algunos de los modelos que tienen mayor relevancia dentro de cada uno de los grupos en los que se han dividido los modelos objetivos.

3.3.1. Modelos de medida de error.

Los modelos de medida de error se han definido como aquellos que no van a incorporar características propias del SVH. Proporcionan medidas que se basan en funciones matemáticas simples, normalmente con un procesamiento en el dominio espacial y punto a punto de la imagen. Estos métodos obtienen un valor de calidad en términos de desviaciones entre la imagen procesada y la original, es decir, un valor de error.

Los más utilizados son el error cuadrático medio *MSE* (“Mean Square Error”) y la relación señal a ruido de pico *PSNR* (“Peak Signal to Noise Ratio”). Se definen como:

$$\mathbf{MSE} = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (I(i, j) - \hat{I}(i, j))^2$$

y

$$\mathbf{PSNR} = 10 \log \frac{(2^b - 1)^2}{MSE} dB$$

donde N y M son el número de columnas y filas de la imagen, $I(i,j)$ es el valor de intensidad de la imagen original en el píxel de posición (i,j), $\hat{I}(i,j)$ es el valor en ese mismo punto para la imagen procesada y b es el número de bits por píxel.

Otra posible medida de calidad surge de la normalización del MSE para reducir la sensibilidad que este presenta a cambios globales en la intensidad de las imágenes. Así, el *NMSE* (“Normalized Mean Square Error”) normaliza las intensidades haciendo que la imágenes resultado tengan media cero y desviación estándar uno. Después, sobre estas imágenes se aplica la fórmula de MSE.

El uso de este tipo de medidas presenta numerosas ventajas entre las que se pueden destacar su simplicidad, su velocidad y sobre todo, el que son portables y están globalmente aceptadas. El valor obtenido no depende ni de las características específicas del proceso de medida (características del monitor, distancia de visualización...) ni del observador, lo que en principio es una ventaja. Sin embargo, este será también su principal problema ya que al no tener en cuenta las sensibilidades del sistema visual tratan todas las distorsiones dándoles la misma importancia independientemente de su tipo, localización en la escena, etc. y por ello no predicen de forma correcta la calidad que realmente es apreciada por el observador. En definitiva, son poco precisas.

Otro inconveniente añadido es que presentan poca robustez. Para estímulos particulares o imágenes con contenido específico (por ejemplo en áreas más o menos uniformes) pueden predecir bien el error. Son útiles, por ejemplo, para comparar un conjunto de imágenes que tienen una misma imagen de referencia y un mismo tipo de distorsión introducida pero de diferente valor (diferentes radios de compresión, de varianza de ruido gaussiano, etc.). Sin embargo, cuando se utilizan sobre diversas escenas o sobre una misma pero con distintos tipos de distorsión pone de manifiesto el hecho de que para valores de PSNR iguales se observan calidades diferentes. Un ejemplo de esto se muestra en la figura 3.1 en la que aparecen cuatro bloques distintos de la imagen Bike de la ISO.

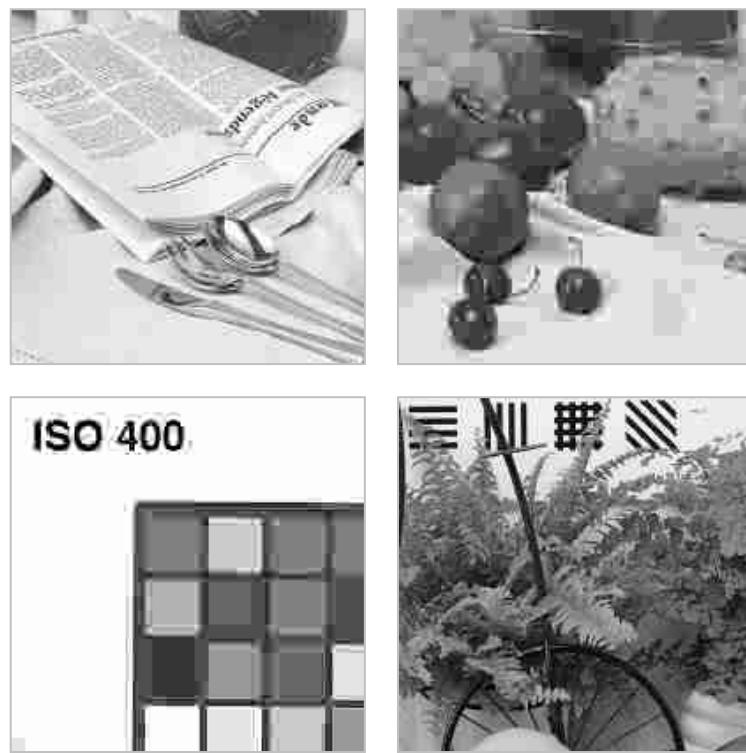


Figura 3.1: Imágenes con PNSR = 26 dB.

Las cuatro imágenes que se muestran están comprimidas con JPEG y todas ellas tienen un valor aproximado de PSNR de 26 dB, sin que ello implique que las calidades percibidas sean iguales, de hecho, dista mucho de ser así. Se aprecian calidades distintas, siendo la mejor la imagen inferior derecha en la que las distorsiones no son tan visibles y la peor la imagen superior derecha.

Una última consideración puede hacerse acerca de la forma del resultado obtenido. Es interesante, como ya se ha comentado, poder disponer tanto de un mapa que muestre la calidad como de un valor numérico que se corresponda con la calidad global observada. Tanto la PSNR como el MSE o sus variantes dan como resultado un valor numérico que no puede reflejar las variaciones espaciales. Esto es un inconveniente más de estos modelos que puede ser superado mediante una modificación del método basada en la utilización de ventanas deslizantes que operan de forma local, obteniéndose así un mapa de la variación espacial de la calidad. Un ejemplo de esto es el MSE local propuesto por Girod [14] para una señal de video. Calcula un valor que denomina visibilidad potencial tomando como entrada el ruido y haciendo un procesado espacial y temporal mostrando dónde y cuándo van a ser visibles los errores.

3.3.2. Modelos de medida de calidad perceptible.

Las métricas de calidad perceptible proporcionan o bien valores numéricos que cuantifican la “satisfacción” del observador con la imagen procesada o bien mapas de visibilidad del error. Es decir, dan medidas que tienen en cuenta no sólo la existencia de posibles distorsiones o errores sino lo visibles que son los mismos dentro de una imagen determinada.

La principal ventaja de este tipo de métricas consiste en que al incorporar de alguna forma las características del sistema visual humano van a dar resultados reales de la calidad que se percibe. Precisamente la incorporación de estas características hace que sean computacionalmente complejas. Todos estos modelos de medida de calidad comparten una estructura similar como la que muestra en la figura 3.2 y la diferencia que habrá entre unos y otros consiste en el modelado que se proponga para el SVH.

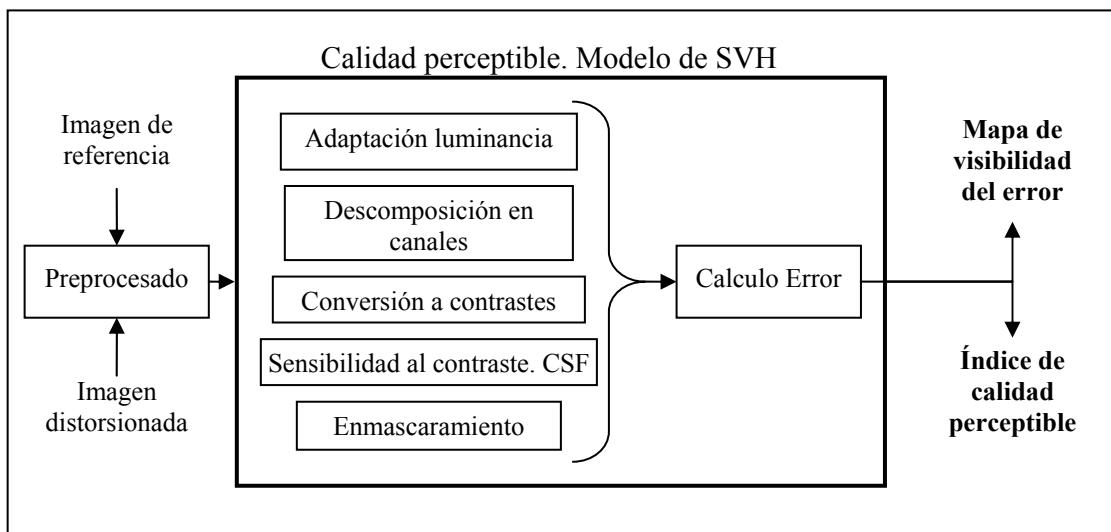


Figura 3.2: Esquema general para modelo de calidad de imagen perceptible.

- **Preprocesado.**

La etapa de preprocesado puede llevar a cabo diversas operaciones de las que se van a destacar las más importantes. En primer lugar, el método de medida normalmente puede requerir un escalado y alineación de las imágenes de entrada. Es necesario puesto que estos métodos se basan en el cálculo de la calidad por comparación y para que ésta sea consistente es preciso que las imágenes se representen dentro de un mismo margen y

con una correspondencia entre ambas punto a punto. Así, si por ejemplo se diese el caso de que moviésemos la imagen distorsionada un píxel a la izquierda o derecha y dejásemos de estar alineadas, cualquier método daría una medida de calidad errónea.

Otra funcionalidad posible correspondiente a esta etapa sería obtener a partir de las imágenes de referencia y distorsionada las entradas al modelo del SVH en las unidades adecuadas. Es decir, se conviertan los valores digitales almacenados en memoria con los que se representa la imagen en los valores de luminancia con los que se visualiza en pantalla. Esta conversión en monitores CRT se corresponde con una transformación no lineal que se conoce con el nombre de función gamma. Por tanto, es necesario realizar un calibrado del monitor y obtener la relación existente entre la entrada digital de niveles de gris y la salida de valores de luminancia en cd/m².

En general, se requiere una caracterización precisa del monitor dentro de cualquier modelo de calidad ya que el SVH será capaz de ver sólo aquello que sea reproducible en la pantalla. Diferentes parámetros del monitor y condiciones de visualización pueden ser requeridas también a la hora de convertir, por ejemplo, frecuencias espaciales a ciclos por grado.

■ **Modelo de SVH para medida de calidad perceptible.**

Los modelos visuales tratan de reproducir lo más fielmente posible el procesado realizado por el SVH y que se corresponde con las primeras etapas del procesamiento visual incluyendo la corteza visual primaria. Esto engloba típicamente los componentes que se muestran en la figura 3.2:

- Adaptación de luminancia y conversión a contrastes.
- Descomposición en canales.
- Sensibilidad al contraste (CSF).
- Enmascaramiento.
- Cálculo de error. Suma de canales.

A continuación se van a exponer las clasificaciones, limitaciones y modelos más representativos que existen en medida de calidad perceptible.

Posibles clasificaciones.

No todos los modelos visuales para calidad de imagen incluyen los componentes arriba enumerados. Existe gran variedad de métodos de medida de calidad basados en el SVH y se han hecho diversas clasificaciones de los mismos. En [10], por ejemplo, se dividen en tres grandes grupos. El primero de ellos para los modelos con adaptación de luminancia y CSF únicamente, el segundo con los que incorporan características para modelar condiciones por encima del umbral y el tercero para condiciones umbral. En [26,35] se hace una clasificación de los modelos visuales atendiendo a si el análisis de frecuencia es de un solo canal o varios. Los planteamientos hechos para un único canal son las primeras aproximaciones que se efectuaron para modelar el SVH y no se corresponden con la descomposición que realizan las neuronas corticales que son selectivas en orientación y frecuencia (ver capítulo 2). Por ello, los modelos visuales con varios canales representan mejor al SVH y aunque más complejos, con ellos se obtendrán resultados mejores.

Algunas limitaciones.

Antes de entrar en la descripción de los modelos de mayor relevancia es necesario comentar algunos de los puntos débiles que presentan estos métodos. La calidad perceptible, en todos ellos, es estimada mediante una cuantificación de la visibilidad del error que se consigue introduciendo en el modelo de calidad las funcionalidades de las primeras etapas del SVH (ver capítulo 2). Sin embargo, uno de los principales problemas planteados es la gran complejidad del SVH y su no linealidad. Por tanto, no puede obviarse que el modelado que se hace del SVH usando operadores lineales o casi lineales, aunque aceptado, es bastante simplista en comparación con esta gran complejidad. Esta simplificación hace que se realicen ciertas hipótesis en casi todos los modelos. En primer lugar, no se tienen en cuenta procesamientos de alto nivel como extracción de características, procesos cognitivos, reconocimiento de patrones y atención visual. Se asume que la calidad percibida es determinada en las primeras etapas del sistema visual. En segundo lugar, se va a considerar que el efecto de la CSF y el enmascaramiento dentro de un mismo canal son los factores dominantes de la percepción. Se supone que las interacciones entre canales son suficientemente pequeñas para poder ignorarlas. Esto se ha demostrado que es falso, siendo considerable el efecto que tiene la interacción entre canales. Aun así, la mayoría de los modelos no incorporan

enmascaramiento entre diferentes canales. Por último, se ignora la posible interacción entre los coeficientes dentro de un mismo canal después del enmascaramiento y la CSF, considerándolos independientes.

Otras limitaciones presentes en este tipo de modelos están relacionadas con los experimentos psicofísicos de los que se extraen las curvas de la CSF y de los fenómenos de enmascaramiento. Estas curvas se obtienen con estímulos simples como puntos, barras o señales sinusoidales que no se corresponden con la variedad del contenido presente en las imágenes reales evaluadas. Se generalizan los modelos estudiados para patrones simples a imágenes que se formarían mediante la superposición de muchos de esos patrones y además, como ya se ha comentado, con el inconveniente añadido de que el SVH contiene no linealidades. Por otra parte, el diseño de los experimentos psicofísicos se hace para la detección de condiciones umbral, es decir, el valor para el cual el estímulo pasa a ser visible y por ello, el modelo será eficiente y preciso justamente en condiciones de visibilidad cercanas al umbral.

La estructura básica que comparten estos modelos incluye, además, la suposición de que el valor global de la distorsión percibida aumenta de forma monótona con la suma del error para todos los canales y en todas las posiciones. Por tanto, a la hora de conseguir un mapa de visibilidad del error o un índice global se utiliza la métrica de Minkowski que tiene la forma típica:

$$E = \left(\sum_k (e_k)^\beta \right)^{1/\beta},$$

donde e_k se corresponde con el valor del error en el canal k . La mayoría de los modelos hacen primero esta suma para obtener un mapa y luego se volvería a utilizar esta fórmula para obtener un índice. En el segundo caso k se correspondería con la posición del píxel (fila, columna) y sería una suma espacial. El uso de este tipo de métrica para el cómputo final de los errores plantea también algunos problemas ya que se está asumiendo que errores en diferentes localizaciones son independientes. Esto será cierto si el procesado previo al cálculo del error global hace que los errores no estén correlados. En el caso de los modelos de medida de calidad esto no sucede.

Aun con todas estas limitaciones, los modelos de medida de calidad que se exponen a continuación obtienen buenos resultados y suponen un gran avance frente a las medidas de error tradicionales como SNR y MSE.

Modelos más representativos.

Dos de los modelos propuestos a principios de los años 90 y que han recibido mayor atención son el “Visible Difference Predictor” (VDP) de Daly [7] y el “Visual Discrimination Model” (VDM) de Lubin [19]. Ambos representan un progreso importante en la incorporación de modelos visuales para conseguir evaluaciones de calidad precisas y con un rango amplio de imágenes de entrada. Tanto el VDP como el VDM hacen uso de resultados psicofísicos para conseguir un modelo del SVH completo, siguiendo la filosofía de que cuanto mayor parecido exista entre la estructura y funcionalidades del SVH y el modelo propuesto tanto mejores serán los resultados obtenidos. Para ello integran los diferentes modelos que existen de mecanismos visuales en uno general del SVH. Los experimentos psicofísicos en los que se basan estos modelos se caracterizan por tener estímulos físicos como entrada y respuestas cognitivas como salida. Por tanto, se plantea un primer inconveniente ya que la disposición en serie de varios de ellos hace que se tengan entradas diferentes para las que se han obtenido. A pesar de ello, ambos métodos son bastante generales y como ya se ha comentado son también bastante robustos.

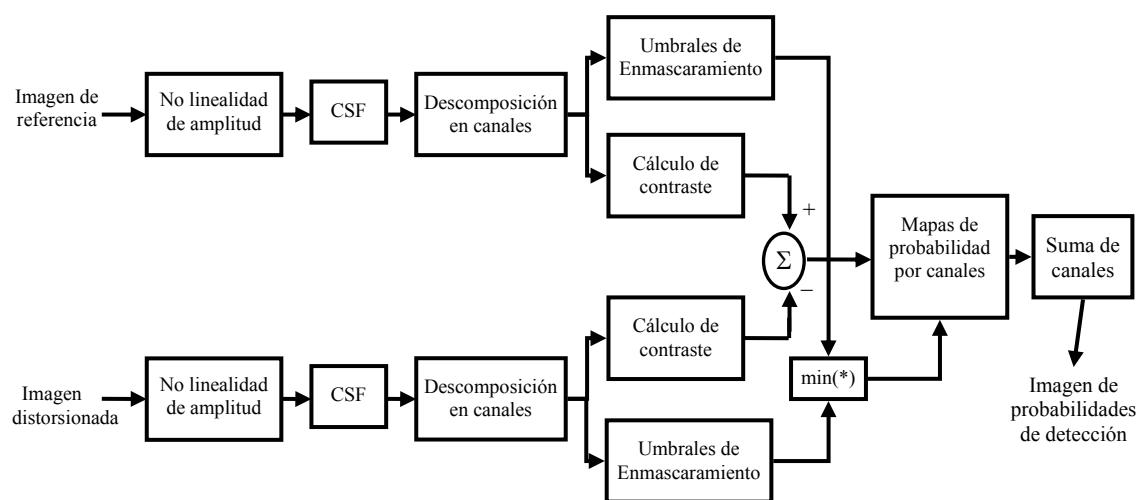


Figura 3.3: Esquema del VDP de Daly.

El VDP incluye todos los componentes típicos del modelado del SVH. Primero modela la adaptación de luminancia como una función no lineal de amplitud, luego efectúa un filtrado con la CSF y descompone en canales. La descomposición se hace mediante un banco de filtros de un octavo de ancho de banda y con diferentes orientaciones. Sobre los canales de la imagen se efectúa una conversión a contraste y se calculan los umbrales de enmascaramiento. Se modela el enmascaramiento de contraste como una elevación del umbral de visibilidad y se calcula a partir de la energía del canal, es decir, para una banda y una orientación. Finalmente, se obtienen los mapas de probabilidad de visibilidad del error y mediante una suma de los errores en los diferentes canales se consigue el mapa de predicción de visibilidad del error. Es decir, cada píxel de este mapa da la probabilidad de que el error sea percibido en esa posición. Un esquema del VDP de Daly puede observarse en la figura 3.3. En el modelo propuesto por Daly se predicen los niveles de visibilidad tanto de estímulos complejos como simples pero para condiciones umbral. No se incorpora ninguna característica para modelar la visión en condiciones por encima del umbral. Además, en este modelo el enmascaramiento entre diferentes orientaciones tampoco está presente.

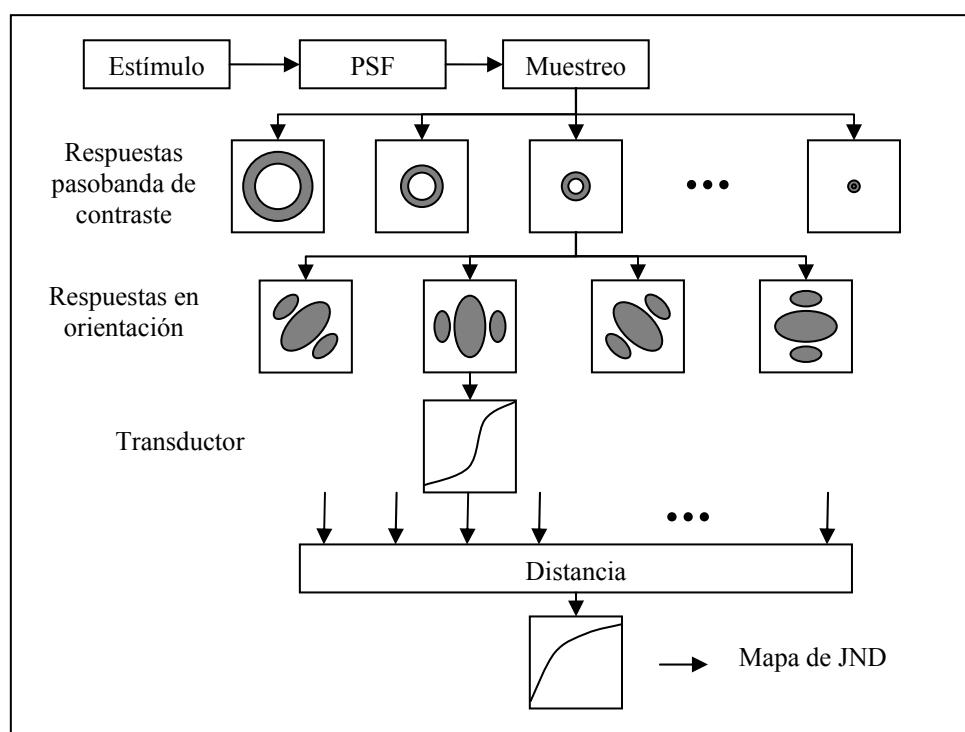


Figura 3.4: Estructura básica del VDM.

El VDM también mide la calidad de imagen estática y presenta la estructura que se muestra en la figura 3.4. Primero se realiza una convolución de la imagen con una aproximación de la PSF (“Point Spread Function”) que representa el efecto de visión en la fóvea y emborrana la imagen de entrada. Se simula también el muestreo realizado en la retina. La imagen se convierte a unidades de contraste local mediante el uso previo de una pirámide Laplaciana y de forma similar al contraste propuesto por Peli [23]. Se obtienen así contrastes localizados en espacio y frecuencia que pasan por una serie de filtros orientados espacialmente y se computa la energía. A continuación se realiza la etapa de enmascaramiento que incluye una normalización y una función no lineal (transductor). Finalmente, se calculan las diferencias y éstas se convierten en probabilidades, obteniéndose así el mapa de distorsiones visibles (JND “just noticeable difference”).

Ambos modelos comparten una estructura similar con un esquema de descomposición y suma. Además, ambos proporcionan resultados para condiciones umbral. Una de las diferencias básicas es que mientras que el VDP trabaja en el dominio de frecuencia con FFT y una jerarquía de filtros, el VDM opera en el dominio espacial. Una comparación de estos métodos se hace en [18] donde se comentan las ventajas de uno y otro en detalle.

De interés resulta también el modelo que se propone en [25] para la evaluación de calidad de imágenes digitales comprimidas. Se basa principalmente en el de Daly pero con algunas modificaciones interesantes. La CSF va a dar un umbral de contraste base para cada canal. Así, la imagen se va a clasificar en tres tipos de regiones: bordes, texturas y zonas uniformes. Los bordes se calculan mediante un detector de tipo Sobel y las texturas se distinguen de las zonas uniformes puesto que su valor de contraste va a estar por encima del valor umbral dado por la CSF. Es decir, en lugar de utilizar la CSF como un filtro, se usa como una serie de pesos que descomponen la imagen en regiones. El enmascaramiento se modela como una elevación de umbral aplicado a cada región. Las zonas uniformes por debajo del valor de CSF no enmascaran, los bordes provocan una elevación del umbral de visibilidad menor que las texturas por lo que la curva de umbral de elevación (ver fig. 2.12) presenta una pendiente de 1 para las texturas y de 0,7 para los bordes.

Otro modelo de medida de calidad perceptible es el descrito por Westen et al. [34]. Es, de nuevo, una combinación de factores que ya existen en otros métodos. La descomposición en múltiples canales y el enmascaramiento de contraste es similar al de Daly. Calcula un contraste local y limitado en banda como el propuesto por Peli [23] pero incorpora un par de parámetros dentro de este cálculo que van a ser ajustados según el modelo de CSF propuesto por Barten. Es decir, incorpora la CSF como una serie de pesos dentro del cálculo de contrastes. Resulta también interesante la obtención del índice de calidad perceptible, que denominan PEM (“Perceptual Error Measure”) y consiste en una suma de Minkowski del error de los canales, primero en frecuencia y luego en espacio, con la particularidad de que los exponentes de la suma se buscan para que el modelo se adecue lo mejor posible a los resultados obtenidos mediante la realización de experimentos subjetivos.

Otro ejemplo interesante de este tipo de modelos de medida de calidad es el DCTune propuesto por Watson [33]. Está pensado para una aplicación concreta: estándares de compresión basados en descomposiciones DCT. El modelo de Watson se desarrolló para conseguir la matriz de cuantificación que hace que la imagen tenga una calidad perceptible óptima. La descomposición utilizando DCT presenta como ventajas que es separable y eficiente desde el punto de vista computacional. Las imágenes se dividen en bloques de 8×8 a los que se aplica la DCT. Para cada uno de los coeficientes de estos bloques se calcula un umbral de visibilidad de la distorsión que incluye los efectos de la sensibilidad al contraste y el enmascaramiento de luminancia y contraste. Se obtiene el error como la diferencia entre los coeficientes DCT de la imagen original y la distorsionada divididos entre los valores de umbral calculados. Es decir, la distorsión está calculada en JND. Finalmente, sólo queda obtener un valor global para evaluar la calidad. Aquí, se observa una particularidad de este método ya que la suma de errores se hace primero espacialmente y no en frecuencia como suele ser lo habitual. Teniendo en cuenta la aplicación para la que está diseñado el método se busca, no un mapa de distorsiones visibles, sino matrices de 8×8 cuyos valores especifiquen la visibilidad del error en esa frecuencia y permitan la modificación independiente de cada elemento dentro de una matriz de cuantificación. Sin embargo, se debe destacar que la DCT no tiene consistencia desde el punto de vista del modelado de las características en frecuencia del SVH y por ello no va a predecir correctamente la calidad percibida para otro tipo de codificadores que no se basen en DCT.

Para calcular el error percibido cuando la distorsión introducida está por encima del umbral es necesario añadir características que simulen la visión en estas condiciones. Miyahara et al. [21] propone un modelo que incorpora algunas preferencias de los observadores frente a artefactos por encima del umbral de visibilidad. Se transforma primero la imagen de forma que sea proporcional a la percepción visual. Para ello se realiza una adaptación de luminancia usando la ley de Weber-Fechner y la sensibilidad al contraste. Se aplican después a los errores una serie de pesos. La novedad de este método consiste en describir una serie de artefactos o distorsiones perceptibles con sus correspondientes factores objetivos que los cuantifican, incluyéndose aquí el enmascaramiento. Finalmente se obtiene un valor PQS (“Picture Quality Scale”) basado en la combinación de estos factores de distorsión.

Aunque existen bastantes modelos más a parte de los descritos, la estructura que comparten todos ellos es bastante similar. No es el objetivo de este apartado el proporcionar una lista exhaustiva de todos los modelos existentes en la literatura, sino exponer las ideas generales que subyacen en estos métodos y su estructura básica, como se ha hecho.

3.3.3. Modelos híbridos de medida de calidad.

En este grupo se engloban modelos de calidad de imagen que tratan de dar una medida que se adecue a la percibida aunque no incorporan el modelo del SVH tal y como lo hacen los modelos anteriores. Siguiendo esta nueva línea se van a considerar 2 modelos que resultan especialmente interesantes y que calculan la calidad de forma completamente diferente. Esto da una idea de la gran variedad de métodos de medida de calidad existentes así como de las distintas formas que se adoptan para abordar un mismo problema.

En primer lugar, se expondrá brevemente un modelo de calidad que se basa fundamentalmente en la información estructural de la imagen para obtener una evaluación de calidad. En segundo lugar, se verán algunas de las ventajas que presentan las distribuciones de Wigner y como se puede obtener una medida de calidad a partir de las mismas.

- **Medida de calidad basada en información estructura.**

Es un modelo propuesto por Z. Wang y A. C. Bovik [30, 31]. Proporciona un valor de calidad acotado entre 0 y 1 al que denominan índice de similitud estructural: SSIM (“Structural SIMilarity”). Se basan en la asunción de que el SVH extrae información de la estructura de los objetos del campo visual e incorporan este nuevo aspecto al método de medida.

El valor del índice de similitud estructural va a depender de tres factores clave dos de los cuales, luminancia y contraste, no presentan ningún aspecto nuevo en métodos de calidad de imagen. El esquema del método se muestra en la figura 3.5. Como puede observarse primero se hace un cálculo de luminancia que consiste en una media matemática, después se realiza el cálculo de contraste usando la desviación estándar y finalmente los atributos que representan la estructura de los objetos en la escena se modelan como una correlación. Así, el valor del SSIM final será:

$$SSIM(I, \hat{I}) = \frac{\sigma_{I\hat{I}}}{\sigma_I \sigma_{\hat{I}}} \cdot \frac{2\mu_I \mu_{\hat{I}}}{(\mu_I)^2 + (\mu_{\hat{I}})^2} \cdot \frac{2\sigma_I \sigma_{\hat{I}}}{\sigma_I^2 + \sigma_{\hat{I}}^2}$$

donde μ_I es la media de la imagen original y $\mu_{\hat{I}}$ la de la distorsionada, σ^2 es la varianza y $\sigma_{I\hat{I}}$ la covarianza. En la ecuación anterior, el primer término se corresponde con la comparación estructural (correlación entre la imagen original I y la distorsionada \hat{I}), el segundo término es la distorsión en la luminancia y el tercero evalúa la diferencia de contraste.

Este resultado puede ser inestable si las sumas de los denominadores son próximas a cero, y por ello, el método introduce algunas constantes para evitar esta posible inestabilidad. Además, las estadísticas descritas anteriormente se calculan localmente, obteniendo un SSIM local para cada punto de la imagen desplazando una ventana gaussiana. Se tendría así un mapa de cómo varía espacialmente la calidad y de este mapa se obtiene un SSIM global haciendo simplemente la media.

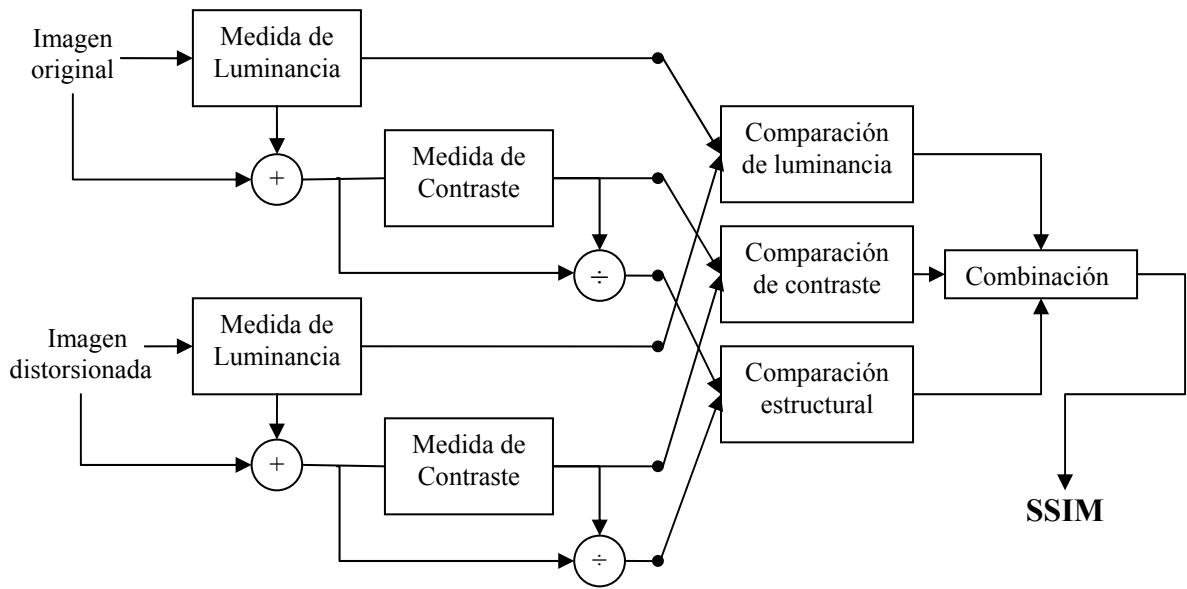


Figura 3.5: Esquema del sistema de medida de SSIM.

Entre las ventajas que se pueden encontrar en este método cabe destacar su sencillez, portabilidad, rapidez y escaso coste desde el punto de vista computacional. Además, proporciona tanto un índice como un mapa de la variación espacial de calidad. Otra ventaja añadida es que la definición hecha valdría para dos señales de entrada cualesquiera y no se restringe al ámbito de imágenes. Sin embargo, muestra una dependencia de la entrada al medir distorsiones de la señal, habría por tanto que estudiar en detalle su robustez y precisión.

■ Medida de calidad con distribuciones Wigner.

Las distribuciones de Wigner-Ville (WVD) se han utilizado con éxito en aplicaciones con señales no estacionarias para analizar las características tiempo-frecuencia. Su uso generalizado se ha dado con señales de una dimensión ya que su extensión a dos o tres dimensiones resulta algo más complejo. La WVD de una imagen 2-D es una función 4-D que incluye una transformada de Fourier por cada punto de la imagen, por lo que son costosas desde el punto de vista computacional. Un ejemplo de aplicación de estas distribuciones en imágenes estáticas se encuentra en [15, 27] donde se emplean para la segmentación de texturas.

Una nueva aplicación para estas distribuciones se propone en [4], donde se calcula una medida de calidad basada en la WVD. La idea general consiste en aprovechar la representación conjunta espacio-frecuencia espacial que permite la utilización de estas distribuciones teniendo presente que en las primeras etapas del procesamiento visual se crea precisamente este tipo de representación. Sin embargo, en comparación con los modelos visuales que implementan los métodos de medida de calidad perceptible, esta aproximación es bastante pobre y no incluye fenómenos de gran importancia como el enmascaramiento.

Una de las propiedades más importantes de la distribución de Wigner-Ville para su aplicación sobre imágenes es que es una función real y al mismo tiempo contiene la información de fase. La distribución continua 2-D de una imagen $f(x,y)$ se define como:

$$W_f(x, y, f_x, f_y) = \int_{\mathbb{R}^2} f\left(x + \frac{\alpha}{2}, y + \frac{\beta}{2}\right) f^*\left(x - \frac{\alpha}{2}, y - \frac{\beta}{2}\right) e^{-j2\pi(\alpha f_x + \beta f_y)} d\alpha d\beta$$

donde x e y son las coordenadas espaciales y f_x y f_y las frecuencias espaciales. Esta función presenta una característica de gran interés ya que localiza las estructuras en espacio y frecuencia. Es decir, si la energía de la imagen está concentrada espacialmente en (x_0, y_0) y en frecuencia en (f_{o_x}, f_{o_y}) , entonces, la energía de W_f estará centrada en $(x_0, y_0, f_{o_x}, f_{o_y})$.

Para incluir esta distribución dentro de un modelo de medida de calidad de imagen estática es necesario calcular una WVD discreta. Esto plantea dos problemas principales: el aliasing y la dispersión espectral por el uso de ventanas espaciales, que será tanto mayor cuanto menor sea el tamaño de las ventanas. La implementación discreta de la WVD se denomina pseudo distribución de Wigner (PWD). Es necesario, además, elegir una señal analítica asociada a la imagen real al hacer la PWD para evitar los principales problemas que surgen en la discretización: el aliasing y la aparición de términos cruzados. No se entrará aquí en detalle en el análisis de PWD y del cálculo de imágenes analíticas, se mostrará simplemente su uso para el cálculo de una SNR de Wigner:

$$SNR_W = 10 \log_{10} \frac{\sum_i \sum_j \max_{u,v} (|PWD_I(i, j, u, v)|)}{\sum_i \sum_j \max_{u,v} (|PWD_I(i, j, u, v) - PWD_{\hat{I}}(i, j, u, v)|)}$$

Se calcula la PWD para la imagen original (I) y para la distorsionada (\hat{I}), donde (i,j) son las coordenadas espaciales y (u,v) son las frecuencias espaciales. En cada posición (i,j) se va a guardar el componente de máxima energía de la PWD. Se trata de simular una especie de enmascaramiento al suponer que este máximo oculta las contribuciones de los demás componentes. Es, por supuesto, una aproximación muy básica, ya que los fenómenos de enmascaramiento son de gran complejidad.

La medida de SNR_W presentada da resultados más precisos (mejores en cuanto a la relación entre este valor y lo que percibe el observador) que los obtenidos mediante medidas como la SNR. Esto se debe básicamente a que las distorsiones que presentan una estructura clara son mucho más molestas para el SVH que las que no la poseen. Las primeras van a estar concentradas en el dominio espacial y de frecuencia mientras que las segundas no. Así, para el caso de dos distorsiones que presenten la misma energía y por tanto, la misma SNR si se aplican sobre la misma imagen original, va a ser mucho más molesta aquella que tenga una distribución con un pico de energía localizado en el plano espacio-frecuencia, es decir, la distorsión estructurada. Estos picos son capturados por la función máximo que existe en el denominador dando como resultado diferencias mayores y SNR_W menores que las que produciría una distorsión no estructurada.

En conclusión, el uso de WVD presenta como mayores ventajas la representación conjunta espacio y frecuencia espacial así como la localización de estructuras. Sin embargo, el precio que se paga por ello es elevado al pasar de 2-D a 4-D. El cálculo es costoso en tiempo y las aplicaciones que pueden darse a esta medida serían limitadas. Además, el valor SNR_W es un índice global y sería interesante poder disponer de mapas de visibilidad del error. Por todo ello, si se utilizan este tipo de distribuciones convendría poder explotar mejor sus propiedades e incluirlas en un modelo más complejo de sistema visual humano incluyendo el enmascaramiento, etc. para obtener valores precisos de calidad percibida. Es, a pesar de todo, una línea de investigación abierta para nuevos modelos de medida de calidad que puede resultar de gran interés.

3.4. Distorsión de imagen.

Tanto en la definición dada de calidad de imagen como en la descripción hecha de los métodos de medida de calidad en apartados anteriores es imprescindible disponer de una imagen considerada original o de referencia y otra que es una modificación de la misma y que ha sido sometida a algún tipo de procesado. Se pretende concretar aquí los distintos tipos de procesado que dan lugar a estas imágenes distorsionadas.

En la implementación práctica de modelos de medida realizada en este proyecto se va a disponer de un modulo de ruido que a la hora de cargar la imagen distorsionada permitirá abrir una imagen ya distorsionada o procesar la original e introducir distintos tipos de ruido o compresión. En el Anexo B correspondiente al entorno gráfico se muestra este modulo que incluye ruido gaussiano, ruido de sal y pimienta, filtrado con un filtro de media, JPEG, JPEG2000 y técnicas de marcado de agua. Este apartado se va a centrar en las técnicas de compresión y marcado de agua, dada su gran utilización e importancia para todo tipo de aplicaciones. Evaluar la calidad de este tipo de imágenes es uno de los objetivos de este proyecto.

3.4.1. Técnicas de compresión y artefactos.

Los estándares de compresión dan al usuario los mecanismos necesarios para que la interoperabilidad entre diferentes aplicaciones que utilizan compresión de imágenes sea posible. Tanto el estándar JPEG como JPEG2000 presentan un conjunto de técnicas de compresión para *imágenes estáticas*, dando soporte a aplicaciones que requieren un almacenamiento o transmisión de imágenes con un límite o bien en capacidad de almacenamiento o en ancho de banda. Ambos estándares están ampliamente difundidos.

▪ JPEG.

El estándar JPEG fue aprobado en Junio de 1991 como resultado del trabajo de un comité con este mismo nombre: “*Joint Photographic Experts Group*”. JPEG tuvo una gran aceptación debido sobre todo a su sencillez, rapidez y eficacia. Esta norma se aplica a imágenes digitales con escalas de grises o color, pero no será aplicable sobre imágenes binivel. Se contemplan cuatro modalidades dentro de la norma: sin pérdidas,

secuencial, progresiva y jerárquica. La compresión sin pérdidas es para aquellas aplicaciones que requieren una recuperación exacta de la imagen comprimida y las tres modalidades restantes son de compresión con pérdidas. Los valores utilizados normalmente para tasas de compresión en JPEG oscilan entre 12:1 y 96:1 (2 a 0,25 bits/píxel), con una degradación apreciable de la imagen a partir de grados de compresión superiores a 20:1.

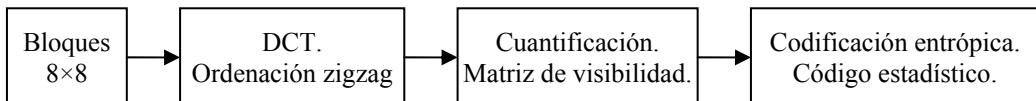


Figura 3.6: Esquema básico de codificador para modalidad secuencial.

El algoritmo más utilizado es el secuencial e incluye los bloques mostrados en la figura 3.6 para el codificador. En el decodificador se seguirá un procesamiento inverso. La imagen original se va a descomponer en bloques de 8×8 píxeles a los que se aplica una DCT directa y los coeficientes resultantes (1 coeficiente DC y 63 coeficientes AC) van seguir una ordenación en zigzag. La cuantificación consiste en un escalado y truncamiento de los coeficientes en función de una tabla o matriz de visibilidad, que da el valor por el que se divide cada coeficiente. Según los valores dados a esta tabla se van a tener tasas de compresión mayores (valores de la tabla más elevados que posteriormente, al truncar, van a dar ceros) o menores. Se denomina matriz de visibilidad ya que pretende dar un escalado a cada coeficiente en función de la visibilidad que tenga esa frecuencia. Esta es una de las primeras incorporaciones que tratan de tener en cuenta, aunque de manera muy elemental, las características del sistema visual. Finalmente, se realiza la codificación, que para el coeficiente DC es diferencial respecto al coeficiente DC del bloque anterior y para los 63 restantes es una codificación estadística mediante un código Huffman o aritmético. La matriz de visibilidad y el código estadístico deben ser enviados también al decodificador.

■ JPEG2000.

JPEG2000 surge fundamentalmente porque las cualidades del estándar JPEG se muestran insuficientes. Se pretende obtener una mejor calidad con tasas de compresión mayores. En este empeño juega un papel de gran importancia el concepto de calidad percibida. Así, se plantea un esquema de compresión que contenga aspectos del SVH y que abandona la DCT en busca de transformaciones que permitan incluir esos factores (típicamente la CSF y el enmascaramiento). Un esquema básico de JPEG2000 se muestra en la figura 3.7, donde puede observarse que la descomposición se realiza mediante una transformada wavelet, que es la base del nuevo estándar.

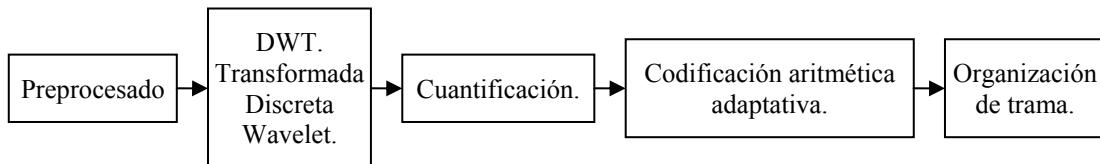


Figura 3.7: Esquema básico de sistema de compresión en JPEG2000.

JPEG2000 proporciona también modalidad con y sin pérdidas. Algunas de las características principales que se pretenden obtener son: compresiones eficientes, representaciones con múltiples resoluciones, protección frente a errores, decodificación SNR progresiva y en resolución, formatos flexibles, acceso y procesamiento aleatorio de la trama, regiones de interés en la codificación, fragmentación de la imagen y posibilidad de interacciones en la codificación. Todas estas características hacen de JPEG2000 una herramienta de gran utilidad en un mayor rango de aplicaciones y que además, proporciona una mayor calidad para tasas de compresión iguales a las de JPEG. Con JPEG2000 se van a obtener calidades de imagen buenas para tasas altas de compresión, incluso por encima de 80:1 y con calidades comparables reduce el tamaño de los ficheros en torno a un 25-35%.

El estándar está dividido actualmente en 11 partes. La primera de ella es el núcleo básico, las partes 2, 4, 5 y 6 definen extensiones. La parte 3 corresponde a JPEG movimiento y de la 7-11 tenemos: JPSEC (seguridad), JPIP (protocolo interactivo), JP3D (imágenes volumétricas) y JPW (transmisión inalámbrica).

■ Artefactos.

La compresión de imágenes con pérdidas hace que parte de la información presente en la imagen se pierda. La imagen original y la comprimida no serán iguales, apareciendo distorsiones. Se pretende que la información de la cual se prescinde sea aquella que presenta una menor visibilidad aprovechando las cualidades propias del SVH para que estas distorsiones no sean percibidas. Los artefactos son, por tanto, estas distorsiones introducidas por la pérdida de información inherente a la compresión y que además, son visibles. A continuación se van a enumerar algunos de los artefactos más comunes que aparecen en esquemas de compresión de imagen estática:

- De bloque (“blocking”): es uno de los más conocidos y consiste en la aparición de pequeños bloques cuadrados sobre toda la imagen. Se debe a la utilización de bloques en el algoritmo de compresión, como es el caso de JPEG que muestra la imagen, donde se observan claramente los bloques de 8×8 .



- De color: se debe al proceso que siguen algunos algoritmos al comprimir la información de color. Puesto que el ojo es más sensible a la información contenida en la luminancia, ésta se mantiene con la misma resolución, mientras que en la crominancia se utiliza una resolución menor. Este problema aparece en la mayoría de los estándares de compresión, siendo más significativo para imágenes comprimidas en JPEG.

- De anillo (“ringing”): se debe a que los algoritmos de compresión utilizan transformaciones para operar en el dominio espectral, que no es especialmente apropiado para la extracción de bordes. Por tanto, aparecen en compresión con DCT y también con DWT. Se presentan como una serie de ondas que aparecen como contorno alrededor de los bordes y son también bastante comunes.

- Emborronamiento (“blurring”): está presente también en la mayoría de estándares de compresión y consiste en el suavizado o emborronamiento de toda la imagen. Los detalles de la imagen se aprecian peor y los bordes son menos abruptos.

3.4.2. Marcado de agua.

El marcado de agua (“watermarking”) digital se propone como solución a los problemas que plantea la facilidad de distribución y reproducción actual de las imágenes digitales, para proteger los derechos de autor y también como una forma para seguir la distribución de los datos marcados. Es un sistema de protección que consiste básicamente en insertar cierta información sobre la imagen. En realidad, el marcado de agua no sólo se efectúa en imágenes, sino que es un campo que se aplica también a audio y video. Un sistema genérico de marcado de agua para imágenes es el que se aprecia en la figura 3.8. Este sistema está compuesto por dos módulos principales que realizan los procesos de codificación o inserción de la marca y de decodificación o extracción. Primero el sistema coge la marca de agua y la inserta en la imagen, pudiendo ser necesaria la utilización de una clave para reforzar la seguridad. El sistema de detección busca las marcas de agua en las imágenes que recibe y trata de extraerlas e identificarlas. Según el método utilizado puede ser necesario el uso de claves e incluso la imagen original y/o la marca.

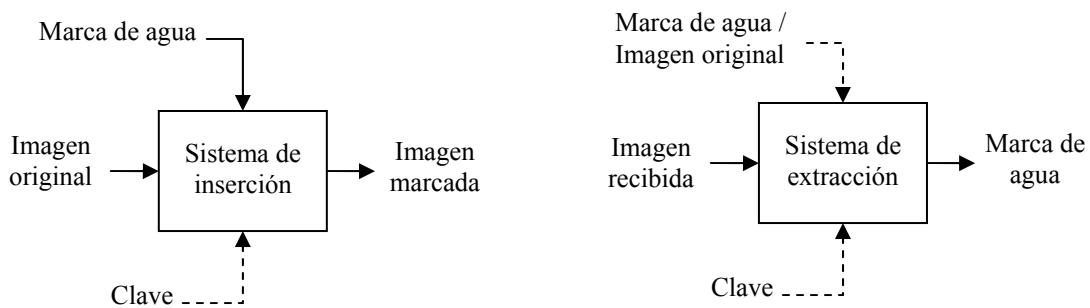


Figura 3.8: Estructura genérica de un sistema de marcado de agua.

Las marcas de agua pueden contener diversos tipos de información como por ejemplo derechos de autor, número de copias autorizadas, el creador, etc. y pueden ser perceptibles o no. El marcado de agua que no es o no debe ser percibido por el usuario es el que resulta de interés a la hora de evaluar la calidad de una imagen marcada. En este sentido, de nuevo, la incorporación de las características del SVH en los procesos de marcado hacen posible que se inserte la misma marca con mayor amplitud en las zonas que pasan desapercibidas para el observador como bordes de imágenes y también texturas, haciendo que mejore la calidad medida.

Desde el punto de vista de las características deseables que debe cumplir la marca de agua se destacan tres: robustez (debe ser difícil de distorsionar o manipular, resistente a ataques), no ambigüedad (no dé falsos positivos en la detección) e imperceptibilidad (el observador no distingue la imagen original de la marcada).

Según la técnica de inserción utilizada para marcar la imagen se pueden hacer dos grupos: las técnicas que trabajan en el dominio espacial modificando directamente el valor de luminancia y/o crominancia de los píxeles de la imagen y las técnicas en el dominio de frecuencia, donde se modifica el valor de los coeficientes de la transformación. En el laboratorio del Depto. de Imágenes y Visión del Instituto de Optica (CSIC) se dispone del código Matlab necesario para realizar marcado de agua sobre imágenes, para la posterior evaluación de calidad, siguiendo las dos filosofías básicas. Este material ha sido proporcionado por el Prof. Fernando Pérez de la ETS de Ing. De Telecomunicación de Vigo e implementa dos técnicas de marcado de agua: “spread spectrum modulation” y “distortion-compensated dither modulation”. La figura 3.9 muestra un ejemplo de marcado de agua para la primera de estas técnicas.

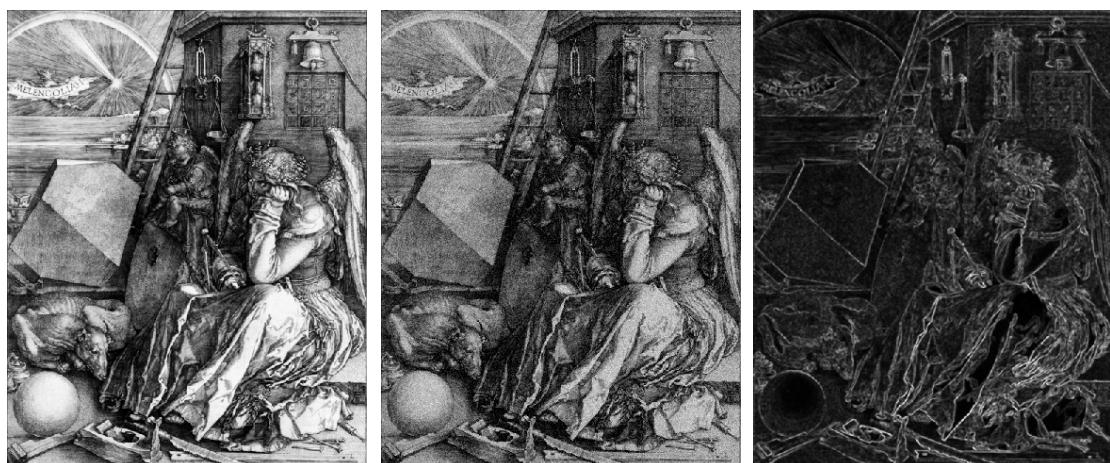


Figura 3.9: De izquierda a derecha imagen original, imagen marcada y marca.

Implementación de modelos visuales para la evaluación de la calidad de imagen.

4

En este capítulo se hace una descripción detallada de todas las implementaciones realizadas en la práctica tanto para el modelado del SVH, como para su integración en sistemas de medida de calidad de imagen. Las funcionalidades del SVH y algunas de las aproximaciones para su simulación ya han sido ampliamente comentadas (Capítulo 2). Aquí, para cada una de ellas se analizarán las distintas propuestas que se han implementado. El objetivo de todo ello es poder comparar, no sólo la influencia que tienen dentro del sistema de medida la adaptación de luminancia, CSF, enmascaramiento, etc., sino también, para una misma funcionalidad, cuál de los modelos obtiene mejores resultados.

En conclusión, no se va a tener un sistema único de medida de calidad de imagen, sino que la combinación de las diferentes opciones compondrá en sí un método de medida. Aun así, no se dispone de un grado de libertad total. Los modelos que se forman están basados en el estudio previo de modelos existentes en la literatura (Capítulo 3) y los distintos componentes deben adecuarse en cada caso a la utilización que se haga de los mismos. Los sistemas de calidad que se forman por la combinación de los diversos factores visuales se manejan de forma sencilla gracias al interfaz de usuario descrito en el Anexo B.

4.1. Preprocesado. Características del monitor.

La adquisición de imágenes y su posterior visualización son sistemas independientes que no están emparejados. Es decir, las imágenes sobre las que se quiere medir la calidad pueden proceder de muy diversos sistemas, haber sido adquiridas mediante distintas técnicas, con distinto formatos, etc. Es necesario realizar un procesado previo que actúe como una especie de filtro regulando las entradas al sistema para que una posterior comparación sea posible, ya que, en esencia, es en esto en lo que se basa una medida de calidad.

En el modelo que se va a presentar éste preprocesado consiste básicamente en una conversión de las imágenes a valores de luminancia. Esta conversión resulta lógica teniendo en cuenta que lo que se pretende es simular el comportamiento del SVH, cuyos estímulos de entrada son precisamente de carácter luminoso. Además, no se debe olvidar que la mayoría de los resultados de los experimentos psicofísicos en los que se basan los modelos visuales están definidos en términos de luminancia de imagen (cd/m^2) y por tanto, los módulos que se utilizan para su implementación han sido caracterizados para entradas de luminancia en unidades de cd/m^2 .

Para poder transformar los valores digitales con los que se almacena la intensidad de las imágenes en cada punto (niveles de gris) en unidades físicas de luminancia es necesario conocer el dispositivo a través del cual se va a efectuar la visualización. La caracterización del monitor es indispensable en los modelos de medida de calidad, constituyendo un campo en sí mismo [5]. Si la métrica de calidad implementada se quisiese utilizar en diversas aplicaciones y con dispositivos de visualización diferentes habría que conocer la conversión que se realiza en cada uno de ellos de niveles de gris a luminancia. Para el sistema de medida desarrollado se ha considerado el caso de monitores CRT, por lo que para una aplicación sobre otro tipo de dispositivo habría que implementar un nuevo bloque de preprocesado en consonancia con las características propias del mismo. Por tanto, se va a analizar únicamente el funcionamiento básico que presentan los monitores CRT con el fin de construir un bloque de preprocesado válido para cualquiera de ellos, siempre que se proporcionen una serie de parámetros de entrada comunes.

En la figura 4.1 se muestra un esquema del funcionamiento básico del monitor partiendo del valor digital almacenado en el buffer hasta la emisión de luz en pantalla. Esta cadena sería la misma para los tres colores rojo (R), verde (G) y azul (B); en el esquema se muestra sólo la del rojo.

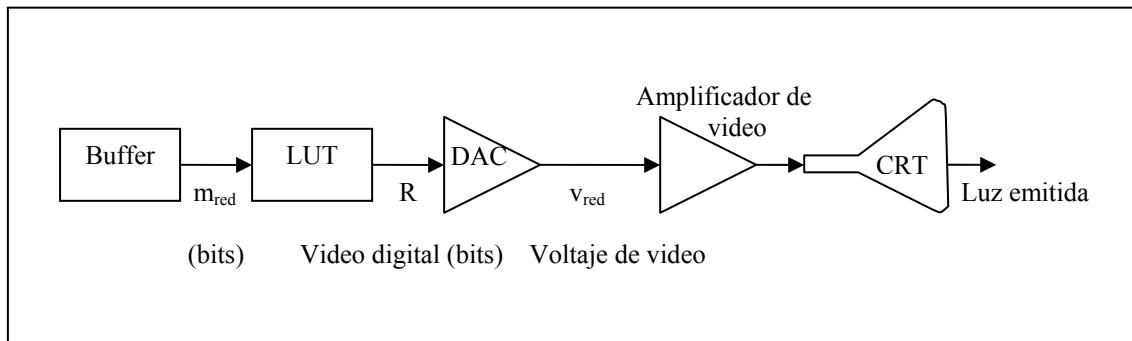


Figura 4.1: Esquema de la tarjeta gráfica y monitor CRT para el rojo.

La luz emitida por cada píxel del monitor se produce cuando un rayo de electrones excita una capa de fósforo en la pantalla. Los voltajes de video (v_{red} , v_{green} , v_{blue}) a la salida de la tarjeta gráfica pasan por un amplificador y se transmiten al CRT. La luz emitida por los fósforos del CRT es proporcional a la intensidad del rayo de electrones, pero esta no tiene una relación lineal con los valores de video. Esta relación no lineal recibe el nombre de función gamma. Es necesario conocer esta función para poder realizar la conversión a luminancia de las imágenes. Su forma típica viene dada por la siguiente expresión:

$$L = L_{\max} \left(\frac{g}{2^b - 1} \right)^\gamma \quad (4.1)$$

donde L es el valor de luminancia normalizado, g el valor de nivel de gris y $(2^b - 1)$ el número de niveles. El valor de γ para monitores CRT suele estar en torno a 2,2.

Para manejar esta no linealidad impuesta por el monitor y corregirla, se suelen seguir los siguientes pasos:

1. Caracterizar la función de transferencia del sistema.
2. Implementar la función inversa.
3. Usar esta función para modificar el estímulo antes de presentarlo en pantalla.

El segundo y tercer paso de esta lista no serían necesarios para un sistema en el que lo único que se pretende es obtener una conversión. Sin embargo, una vez calculada

la función de transferencia, implementar los otros dos pasos es sencillo y de gran utilidad ya que así se consigue obtener una relación lineal entre niveles de gris y valores de luminancia, que hará que la posterior visualización de las imágenes no se vea modificada por la función gamma del monitor. Si no se realiza este calibrado, las imágenes se verán afectadas por la no linealidad del monitor, lo que se traducirá en la práctica como un oscurecimiento de la imagen al presentarla en pantalla.

Para caracterizar la función de transferencia es necesario medir la luz emitida para los distintos valores de niveles de gris. El equipo de trabajo del laboratorio (del Depto. de Imágenes y Visión del Instituto de Óptica (CSIC)) consta de dos monitores SONY F520 de 21". Únicamente se va a calibrar el que se utiliza para la visualización de las imágenes. Se dispone también de un fotómetro ColorCal y del software "LightScan" para su utilización. Se tomaron medidas de luminancia para 11 valores distintos de niveles de gris, obteniéndose de esta forma la curva que se muestra en la figura 4.2 en azul. Esta curva se corresponde con un valor de gamma aproximado de 2. Una vez conocido este valor, el cálculo de la función inversa es inmediato y el resultado final obtenido al aplicar esta función es la relación lineal que se observa en la figura 4.2 en verde

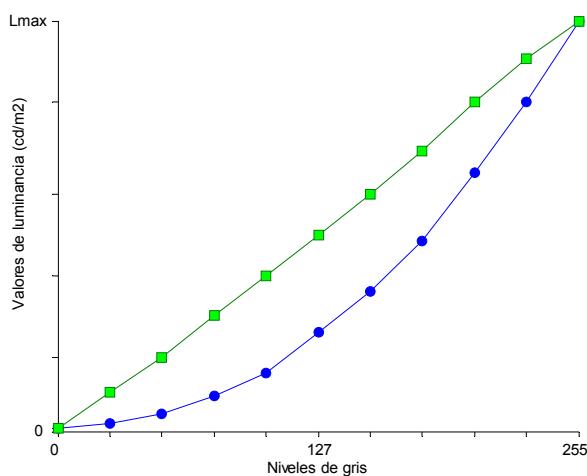


Figura 4.2: Característica gamma del monitor CRT del laboratorio (curva —●—) y resultado final al aplicar una función inversa con un valor de $\gamma=2$ (curva —■—).

Con esto se obtendrá finalmente una relación lineal entre niveles de gris y valores de luminancia que cumple los dos objetivos que se han propuesto. Por una parte sirve para realizar la conversión a cd/m^2 y por otra, para visualizar las imágenes sin que resulten modificadas por la característica gamma del monitor. Además, este módulo de

preprocesado permite que el sistema de medida de calidad sea portátil. Es decir, se puede montar sobre cualquier sistema con monitores CRT. El esquema general de este bloque puede observarse en la figura 4.3. Los parámetros de entrada necesarios, además de la imagen, para poder realizar la conversión son el valor de gamma (γ) y luminancia máxima (L_{max}), lo que puede deducirse también fácilmente de la expresión (4.1). La visualización de las imágenes es de gran importancia para cualquier sistema de medida de calidad perceptible ya que el valor de calidad percibido por el observador depende de lo que éste esté viendo, siendo por ello imprescindible que la relación de entre los valores digitales de la imagen almacenada y la luz emitida por cada píxel sea lineal y conocida.

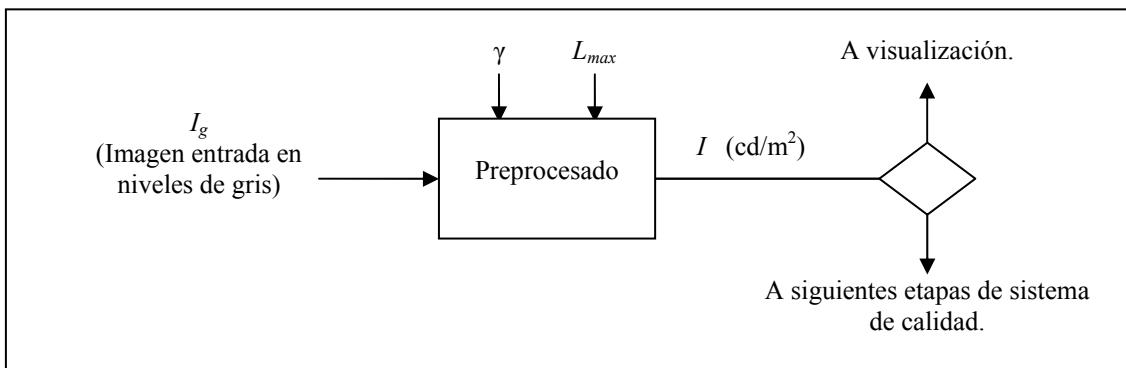


Figura 4.3: Esquema general de etapa de preprocesado.

4.2. Adaptación de luminancia.

Se basa principalmente en el hecho de que la sensibilidad a las diferencias de intensidad es dependiente de la luminancia local en las regiones de la imagen (ver 2.3.1). La ley de Weber-Fechner establece que las diferencias de luminancia en un estímulo son proporcionales a la luminancia media del mismo (el umbral de contraste permanece constante para valores de luminancia crecientes). La ley de Weber es válida para valores de luminancia por encima de unas 10 cd/m^2 , por debajo de este valor, el umbral de contraste aumenta cuando la luminancia disminuye. Así, nuestro modelo debe reflejar que si la luminancia local aumenta, un mayor nivel de distorsión puede ser tolerado.

La incorporación de la adaptación de luminancia a los modelos de medida de calidad puede hacerse en el dominio espacial o en el de frecuencia. En este proyecto se han implementado 3 posibles curvas de adaptación de luminancia en el dominio espacial y en las que el nivel de adaptación para un píxel de la imagen depende únicamente de ese píxel. Son transformaciones previas a la descomposición de la imagen en canales y tienen como entrada los valores de luminancia $I(x,y)$ en cd/m^2 para cada píxel, que se conocen gracias a la etapa de preprocesado. En la figura 4.4 se muestran las distintas opciones que se han propuesto en cuanto al modelado de la adaptación de luminancia y son:

- No se realiza ninguna adaptación de luminancia, es decir, no se modela. Se corresponde con la función lineal.
- Dos de las adaptaciones más comunes y utilizadas: la raíz cúbica y la función logarítmica.
- Una adaptación de luminancia local propuesta por Daly [7] y que se obtiene aplicando la siguiente fórmula: $I_l(x,y) = \frac{I(x,y)}{I(x,y) + 12.6 \cdot I(x,y)^{0.63}}$.

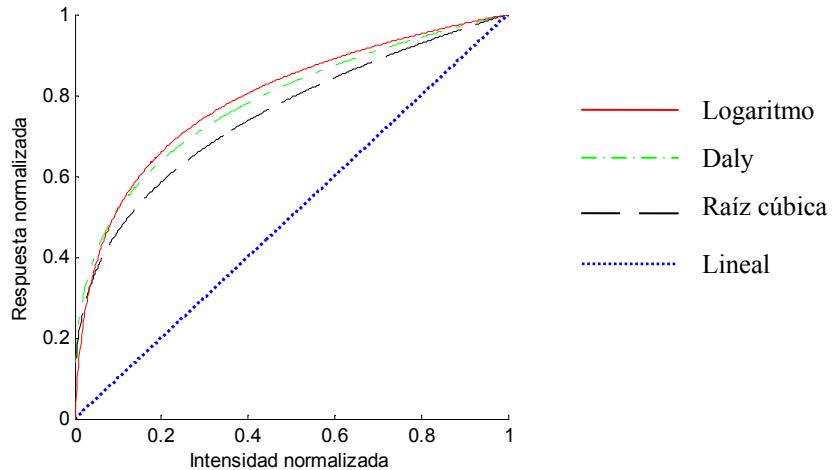


Figura 4.4: Curvas de adaptación de luminancia.

Del análisis de la esta figura se pueden sacar varias conclusiones. En primer lugar se observa que la función propuesta por Daly y la raíz cúbica muestran un comportamiento muy similar para el margen de niveles de luminancia de entrada con el que se trabaja (rango en torno a $100 \text{ cd}/\text{m}^2$), estando la curva de Daly ligeramente por encima de la raíz cúbica. Sin embargo, existe una gran diferencia entre las dos, ya que el modelo propuesto por Daly va a cambiar de forma en función del nivel de iluminación

que exista, mientras que la raíz cúbica no. En nuestro sistema de medida de calidad esta diferencia no será apreciable ya que se dispone de un margen de luminancias determinado sobre el que se trabaja y que además es en el que las dos funciones resultan similares. En cuanto a la función logarítmica, algunos autores [7] afirman que sobreestima la sensibilidad visual en regiones de baja intensidad luminosa y por tanto, la raíz cúbica sería un mejor modelo.

En general, la implementación que se haga para la adaptación de luminancia no tiene gran impacto en los modelos de medida de calidad a no ser que las imágenes presenten un gran contraste o la localización del enmascaramiento de luminancia sea significativa. Es un componente que da la forma de la no linealidad, pero en realidad, la sensibilidad va a ser calculada por la CSF.

4.3. Función de Sensibilidad al Contraste (CSF).

La CSF describe las variaciones de la sensibilidad visual en función de la frecuencia espacial siendo una de las piezas clave para el buen funcionamiento del sistema de medida de calidad. La CSF depende de una gran cantidad de parámetros y su medida resulta bastante compleja (ver 2.3.3). Por ello, se ha optado por implementar modelos ya propuestos en lugar de intentar realizar una medida de la CSF en el laboratorio. En este apartado se van a presentar los modelos de CSF que se incluyen en el sistema de medida de calidad y que son tres:

- CSF de B. W. Rust [28].
- CSF propuesta por Daly [7].
- CSF de Barten [11].

Para estos tres modelos se consideran dos posibles alternativas en el uso de la CSF, que determinarán, en parte, la estructura del modelo de medida. Estas dos posibilidades son:

1. La CSF actúa como una normalización en frecuencia previa a los mecanismos de detección posteriores. Como puede observarse en la figura 4.5, esto se implementa mediante un filtrado bidimensional de la imagen $I_l(x, y)$ que se pasa al dominio de

frecuencia. Lo que se obtiene así es una imagen $I_N(x, y)$ en la que cada componente de frecuencia ha sido dividido por el valor umbral de visibilidad de contraste para esa frecuencia. Es decir, da una transformada de Fourier $I_N(u, v) = \mathcal{F}\{I_l(x, y)\} \cdot CSF(u, v)$ en la que cada coeficiente tiene la misma visibilidad si presenta un valor de amplitud igual.

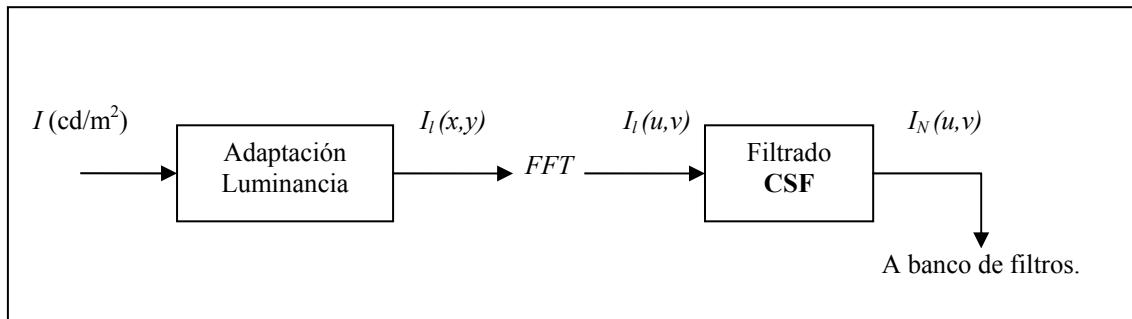


Figura 4.5: Implementación de CSF en sistema de calidad como filtro.

2. La imagen $I_l(x, y)$ se descompone en canales y se transforma a contrastes. Para cada una de las imágenes de contraste formadas se calcula un valor o peso de la CSF como la media en ese canal. Este valor de sensibilidad es la inversa del valor del umbral base $S_O^n = 1/TH_O^n$, siendo n el canal para el cual se calcula. En función del valor de estos pesos se divide la imagen en tres tipos de regiones: bordes, texturas y zonas uniformes. La extracción de bordes se puede hacer mediante un detector de Sobel. Las regiones cuyo valor de contraste esté por debajo de TH_O^n , se corresponden con zonas uniformes y todas las regiones restantes se corresponderán con texturas. Este proceso se verá con más detalle a lo largo de este capítulo y en el modelado del enmascaramiento.

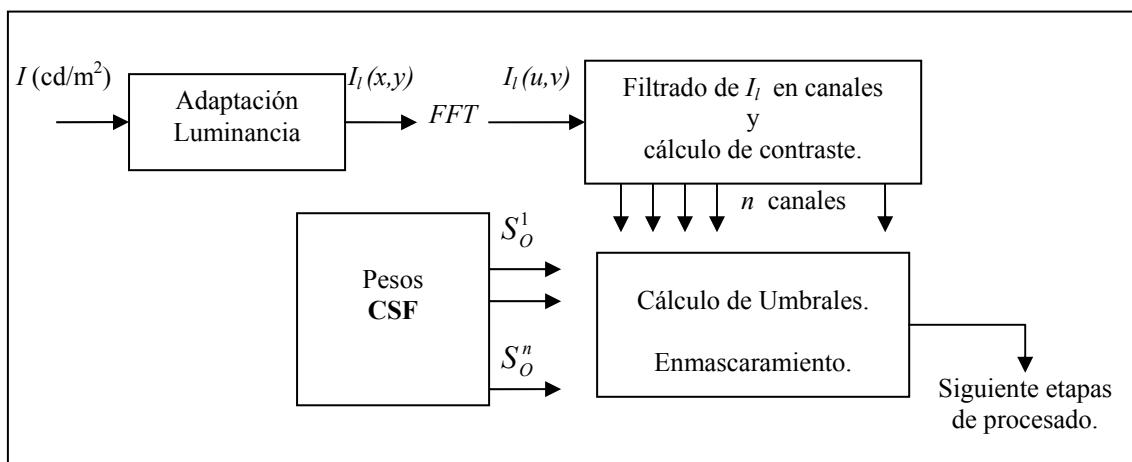


Figura 4.6: Implementación de CSF en sistema de calidad mediante pesos.

El uso de la CSF como filtro permite que se utilice un valor para cada frecuencia y orientación, no un cálculo global para todo un canal, que es lo que supone la segunda implementación. Sin embargo, este filtrado previo también presenta algunos inconvenientes. Daly [7] observó que causa una ligera distorsión en la forma de los canales. Además, se debe tener en cuenta que la elección que se haga para el uso de la CSF determinará la forma en que se calculan posteriormente los umbrales de enmascaramiento.

4.3.1. Modelo 1 de CSF (B.W. Rust).

Se implementa la CSF que obtuvo Rust [28], obteniendo las curvas que se muestran en la figura 4.7. La figura 4.7.a. representa la función en 1D y se aprecia la menor sensibilidad del SVH ante frecuencias oblicuas (curva $\Theta = 45^\circ$ por debajo de $\Theta = 90^\circ$). No es necesario representar el eje de frecuencias negativo ya que esta función es la transformada de Fourier de una función real y por tanto, es simétrica. La función en 2D se muestra también en la figura 4.7.b., donde se aprecia también claramente la característica anisótropa de la CSF.

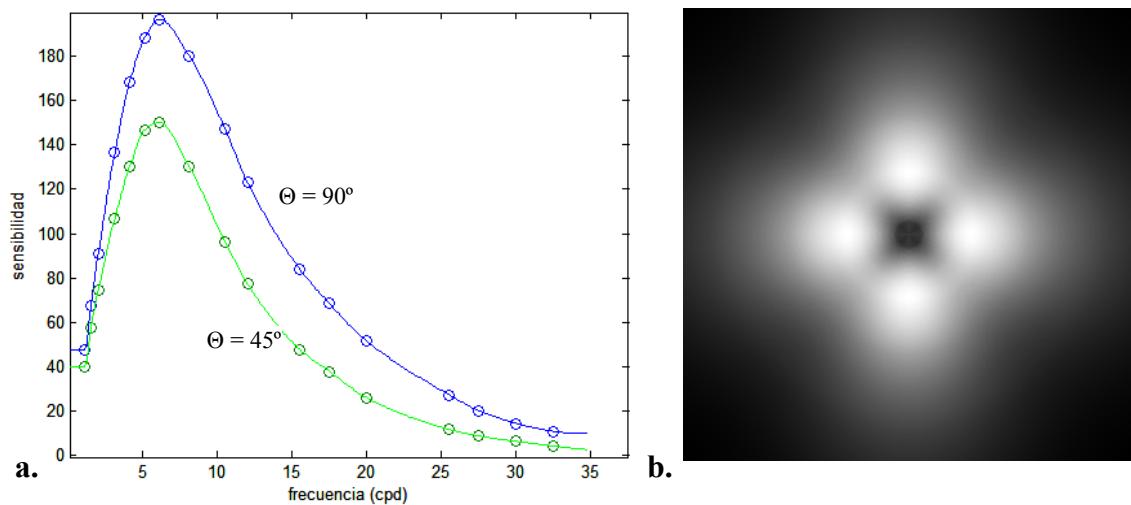


Figura 4.7: CSF de Rust. **a.** CSF para $\Theta = 45^\circ$ y 90° en 1D. **b.** CSF 2D.

Rust toma medidas de la sensibilidad al contraste para 17 valores de frecuencia espacial desde 1.1 a 32.5 ciclos por grado (cpd), con dos orientaciones (90° y 45°) y para tres sujetos. Las medidas tomadas en distintas orientaciones van a permitir que se modele el carácter anisótropo de la CSF que se ha observado en las figuras. El

procedimiento seguido por Rust no consiste en buscar una fórmula matemática que cuadre con el conjunto de datos obtenidos, sino que utilizará una spline cúbica que se ajuste a los valores medidos. Para ello va a servirse de las dos funciones descritas en (4.3), siendo S_i los valores medidos para cada frecuencia ρ_i . La CSF quedará expresada finalmente como se muestra en (4.4).

$$\begin{aligned} b(\rho_i) &= \frac{1}{2} [S_i(90^\circ) + S_i(45^\circ)] \\ A(\rho_i) &= \frac{1}{2} [S_i(90^\circ) - S_i(45^\circ)] \end{aligned} \quad (4.3)$$

$$CSF(\rho, \Theta) = b(\rho) + A(\rho) \cdot \cos(4\Theta) \quad (4.4)$$

Uno de los principales problemas que se plantean a la hora de conseguir una curva a partir del conjunto de medidas obtenido es la ambigüedad en el origen. No se dispone de un valor para frecuencia cero, por lo que Rust optó por dejar que los demás valores medidos fijasen la $CSF(\rho=0)$, ya que, aunque no hay garantías de que la spline cúbica que se utiliza sea la aproximación correcta para la región de bajas frecuencias, sí que presenta un buen comportamiento.

Por último, habrá que tener en cuenta el margen de frecuencias en el que se trabaja. Será necesario disponer de los parámetros distancia de visualización (d en cm) y resolución del monitor (r_0 en píxel/cm) para conseguir expresar la frecuencia espacial en ciclos por grado (cpd). Así, la frecuencia máxima que puede presentarse en pantalla será $f_{\max} = \sqrt{2} \cdot r_0 \cdot d \cdot \pi / 180$. Por tanto, cuanto mayor sea la distancia de visualización mayor será el rango de frecuencias espaciales. La interpolación hecha para conseguir una curva CSF 2D a partir del conjunto de medidas tomadas se debe efectuar en este rango.

4.3.2. Modelo 2 de CSF (Daly).

Daly [7] proporciona un modelo analítico para la CSF en el que va a incluir la influencia de diversos parámetros. La ventaja fundamental del modelo analítico es que permite una mayor flexibilidad y es posible estudiar la forma de las curvas, por ejemplo, para distintos niveles de adaptación de luminancia. La CSF va a depender de la frecuencia (ρ cpd), la orientación (Θ grados), la distancia de visualización (d metros), el

tamaño de la imagen (i^2 en grados visuales), el nivel de adaptación de luminancia (l cd/m²) y la excentricidad (e grados).

$$CSF(\rho, \Theta, l, i^2, d, e) = P * \min \left[S_1 \left(\frac{\rho}{r_a \cdot r_e \cdot r_\Theta}, l, i^2 \right), S_1(\rho, l, i^2) \right] \quad (4.5)$$

$$S_1(\rho, l, i^2) = \left(\left(3.23 \cdot (\rho^2 \cdot i^2)^{-0.3} \right)^5 + 1 \right)^{-1/5} \cdot A_l \cdot 0.9 \rho \cdot e^{-(B_l \cdot 0.9 \rho)} \cdot \sqrt{1 + 0.06 e^{(B_l \cdot 0.9 \rho)}} \quad (4.6)$$

donde r_a , r_e , r_Θ caracterizan la CSF en función de la distancia, excentricidad y orientación respectivamente. La variación con l viene determinada por A_l y B_l . Esta función tiene la forma de la figura 4.8. En 4.8.a. se muestra la curva para $\Theta = 0^\circ$ y dos distancias de visualización distintas. La curva se desplaza a la izquierda para d mayores. Además, la forma también va a variar ligeramente debido a la r_a de la expresión de la CSF.

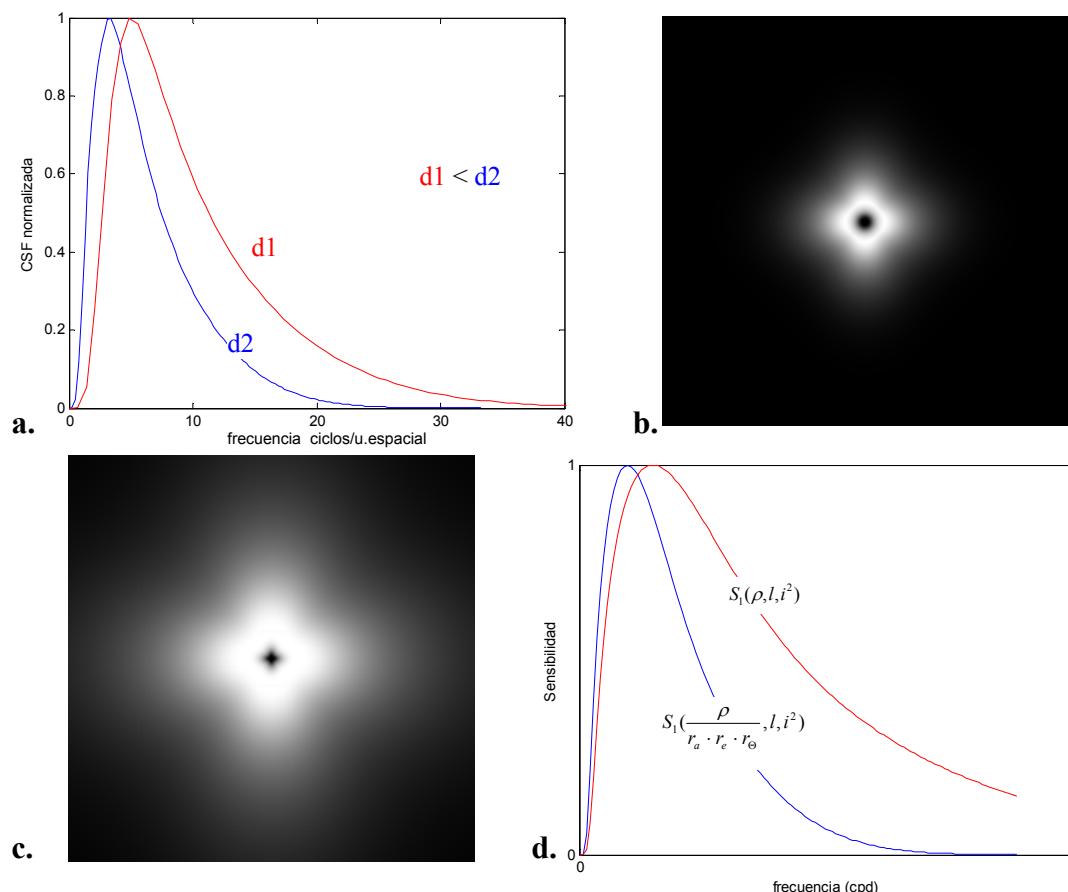


Figura 4.8: CSF de Daly. **a.** Comparación CSF en 1D con $\Theta = 0^\circ$ y 2 valores de d . **b.** CSF 2D. **c.** CSF de Daly modificada mediante la elección del máximo. **d.** Curvas S_1 .

En comparación con el modelo anterior se presentan varias diferencias. En primer lugar, aunque en la figura 4.8 se ha representado la CSF normalizada, el pico de sensibilidad P de la expresión analítica se sitúa en 250. Otra diferencia de gran importancia reside en el valor que se le da a la CSF para $\rho = 0$. Si se sustituye en la expresión (3.6) se obtiene que $CSF(0, \Theta, l, i^2, d, e) = 0$, por lo que al utilizar la CSF como un filtro se estarán eliminando las bajas frecuencias. La función modela también la anisotropía, como se observa en la figura 4.8.b., pero la diferencia de amplitud entre las direcciones más privilegiadas (verticales y horizontales) y las menos (oblicuas) no es tan grande. En general, la característica de esta función es menos suave que la anterior y para valores iguales de distancia de visualización la CSF propuesta por Daly presenta una banda más estrecha, es decir, elimina mucho más las bajas y altas frecuencias. Por tanto, en su incorporación al sistema de medida de calidad, en principio, debería proporcionar peores resultados que la de Rust, ya que se va a alejar más de los resultados experimentales de los test psicofísicos, que dan bandas más anchas. Para utilizar la CSF de Daly en el modelo visual para el análisis de calidad se han hecho algunas modificaciones. Como se ha comentado, la CSF de Daly tiene una banda muy estrecha que va a hacer que al utilizarla como filtro se pierdan bastante las altas y bajas frecuencias de la imagen. Para suavizarla se puede elegir el máximo, en lugar del mínimo, entre las dos curvas S_1 que se muestran en la figura 4.8.d. para obtener una banda de paso mayor. Así, la figura 4.8.c. muestra esta modificación sobre la CSF 2D con los mismos parámetros utilizados para la figura 4.8.b. y se aprecia que conserva la misma forma pero con un ancho de banda mayor. Además, si se pretende utilizar esta función como filtro ha de conservarse el canal paso bajo, ya que es imprescindible para el posterior cálculo de las imágenes de contraste. De hecho, el valor de la CSF para frecuencia cero no puede ser cero ya que esto significaría que en la expresión de contraste $(L_{\max} - L_{\min}) / (L_{\max} + L_{\min})$ utilizada para obtener los valores de sensibilidad al medir la CSF (ver 2.2.3) o bien $L_{\max} + L_{\min}$ es cero (oscuridad total) o $L_{\max} - L_{\min}$ es infinito. El problema es que la medida en frecuencia cero no es posible ya que supondría un periodo infinito de la señal test. La mayoría de las representaciones de la CSF existentes eluden el problema en bajas frecuencias truncando las gráficas para los valores cercanos a cero o simplemente tomando escalas logarítmicas y en general, los modelos matemáticos dan valor cero en frecuencia cero, como en este caso y en el que se va a mostrar a continuación.

4.3.3. Modelo 3 de CSF (Barten).

Se ha incluido un tercer modelo de CSF que aparece en [7, 34] y que ha sido utilizado también en diversos modelos visuales. Barten obtuvo la siguiente fórmula analítica basándose en datos experimentales:

$$CSF(\rho, l, i^2, d) = a \cdot \rho \cdot e^{-b \cdot \rho} \sqrt{1 + c \cdot e^{b \cdot \rho}} \quad (4.7)$$

donde $a = \frac{540 \cdot (1 + 0.7/l)^{-0.2}}{12}$, $b = 0.3 \cdot (1 + 100/l)^{0.15}$ y $c = 0.06$.

Esta función es muy similar a la S_1 de Daly. Aquí, al igual que en el apartado anterior la frecuencia es ρ en cpd, el tamaño de la imagen es i en grados visuales, el nivel de adaptación de luminancia se representa con l en cd/m². La gran diferencia con el modelo anterior es que no aparece la orientación (Θ), es decir, esta es una función isótropa. Aunque es claro el carácter anisótropo de la CSF, resulta de gran interés incluir un modelo que no posea esta característica para estudiar la importancia de la misma en comparación con los modelos 1 y 2 anteriores, que si la incluyen. Otra consideración que debe hacerse es el valor de pico de esta función. Para los mismos parámetros de entrada que el modelo 2 va a tener un pico de sensibilidad mayor. En la implementación práctica se puede elegir el pico de sensibilidad, con lo que es posible, de nuevo, ver la influencia de este parámetro y elegir el más adecuado para nuestro sistema de medida de calidad.

La figura 4.9 muestra la CSF en 2D que se obtiene con este modelo. La forma de anillo que presenta esta función se debe al hecho de que no existen direcciones privilegiadas, poniéndose de nuevo de manifiesto el carácter isótropo de este modelo. La curva en 1D, no aportará información de utilidad ya que será muy similar a las mostradas en el apartado anterior.

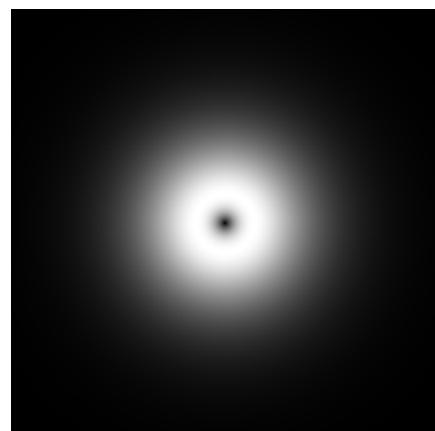


Figura 4.9: CSF de Barten.

4.4. Descomposición en canales.

El estudio del SVH ha mostrado que la representación de la señal visual se lleva a cabo mediante una serie de células que están sintonizadas en canales de frecuencia espaciales y con un rango de orientaciones determinado (ver capítulo 2). De este hecho se infiere que la forma más adecuada para el tratamiento de imágenes dentro de un modelo de SVH va a ser el análisis con múltiples escalas y orientaciones.

En este contexto se van a considerar dos opciones en cuanto a modelos para la descomposición de la imagen de entrada:

- Bancos de filtros con canales diseñados para un determinado rango de frecuencias espaciales y orientaciones. Se han realizado dos implementaciones distintas: filtros *Gabor 2D* [2, 12] y la transformada *Cortex* [7, 34].
- Banco de filtros con canales selectivos en frecuencia únicamente: filtros *anillo* [7, 34].

En principio, no parece que la elección que se haga para el banco de filtros tenga un papel crítico dentro del modelo de medida de calidad siempre que el diseño del mismo esté orientado a simular el procesado de las células simples en el sistema visual. Por ello, aunque resulte más compleja, es de esperar que la primera opción logre resultados mejores. Precisamente, lo que se pretende evaluar al introducir un modelo con descomposición de la imagen sólo por bandas de frecuencia es cuánto más precisas serán las medidas de calidad en relación al grado de complejidad, valorando así si la mejora es significativa o no. Por otra parte, disponer de dos modelos para un filtrado en canales por frecuencia y orientación hace posible que se puedan efectuar comparativas en cuanto a otra serie de parámetros: número de escalas, orientaciones, forma de los filtros, etc.

En general, todas las implementaciones realizadas van a efectuar una descomposición en bandas en el dominio de frecuencia como una multiplicación de la imagen de entrada por cada uno de los canales. Esta imagen de entrada va a variar según las opciones elegidas en cuanto al diseño del sistema de medida de calidad. Así, como se explicó en el apartado anterior el uso de la CSF determinará las distintas posibilidades que pueden presentarse como entrada al banco de filtros:

1. Imagen tras preprocesado, adaptación de luminancia y filtrado con la CSF, independientemente de los modelos empleados para la adaptación de luminancia (raíz cúbica, logaritmo, etc.) y la CSF (Rust, Daly, etc.). La descomposición en bandas sería, por tanto:

$$b_k(x, y) = \mathcal{F}^{-1}\{I_l(u, v) \cdot CSF(u, v) \cdot f_k(u, v)\} \quad (4.8)$$

$$b_{k,l}(x, y) = \mathcal{F}^{-1}\{I_l(u, v) \cdot CSF(u, v) \cdot f_{k,l}(u, v)\}$$

2. Imagen tras preprocesado y adaptación de luminancia. La CSF se implementa como una serie de pesos que se aplicarán posteriormente sobre cada canal. Las bandas resultado quedan expresadas como:

$$b_k(x, y) = \mathcal{F}^{-1}\{I_l(u, v) \cdot f_k(u, v)\} \quad (4.9)$$

$$b_{k,l}(x, y) = \mathcal{F}^{-1}\{I_l(u, v) \cdot f_{k,l}(u, v)\}$$

Tanto el conjunto de filtros obtenidos ($f_k(x, y)$ o $f_{k,l}(x, y)$), como la imagen descompuesta en canales o imagen en bandas ($b_k(x, y)$ o $b_{k,l}(x, y)$), van a guardarse en una matriz de 3D o 4D según el caso, donde las dos primeras dimensiones corresponden con el tamaño de filtro ($N \times N$), la tercera es el índice k (1 a N_f) que indica la escala y la cuarta, en caso de existir, es el índice l (1 a N_o) que indica la orientación. El paso bajo se almacena en $k = N_f + 1$, siendo el único filtro que contiene esta escala.

Finalmente, antes de pasar a describir en detalle las tres implementaciones realizadas habría que aclarar que la realización de los filtros *anillos* se incluirá dentro de la transformada *Cortex*, ya que, como se verá a continuación, forma parte de la misma.

4.4.1. Filtros de Gabor.

Las funciones de Gabor han adquirido creciente importancia dentro del ámbito de tratamiento digital de imagen, siendo utilizadas en diversos campos (análisis de texturas, detección de movimiento, compresión...) con resultados satisfactorios. El uso de transformadas de Gabor para la descomposición de la imagen en este y otros proyectos se fundamenta en las siguientes características:

- La representación de la señal visual es un problema a resolver desde el enfoque de tratamiento de imagen, ya que una representación de gran precisión en el dominio de frecuencia necesariamente lleva asociada baja resolución en el dominio espacial. Esto queda de manifiesto en el principio de incertidumbre de Heisenberg-Weil que plantea que la localización de una función está limitada y el producto de las varianzas de la función en el dominio espacial ($\Delta x, \Delta y$) y de frecuencia ($\Delta u, \Delta v$) no puede ser arbitrariamente pequeño. En concreto, para el caso de 2D esta relación de incertidumbre queda expresada como: $\Delta x \cdot \Delta y \cdot \Delta u \cdot \Delta v \geq 1/16\pi^2$.

En este contexto, Gabor demostró que se puede representar una señal mediante una serie de paquetes elementales de información que minimizan esta relación de incertidumbre proporcionando un empaquetamiento de la información óptimo [13]. Estos filtros 2D dan la máxima resolución en los dos dominios, tanto para la información de frecuencia espacial y orientación sobre la estructura de la imagen como para la información espacial sobre la posición.

- Otra característica importante es que estos paquetes elementales de información (“logons”) son adecuados en la descripción del procesado del sistema visual, que hace una representación de la imagen con un análisis bidimensional en el dominio de frecuencia sobre cada campo receptor retiniano (dominio espacial). Se ha demostrado que las funciones de Gabor pueden modelar el comportamiento de las células simples de la corteza visual.
- La implementación de las funciones de Gabor es sencilla desde el punto de vista computacional tanto en el dominio espacial (convolución) como en el de frecuencia a través de una simple operación de filtrado con los denominados filtros de Gabor.

Gabor se centró en el problema en 1D y Marcelja y Daugman [8, 20] ampliaron para el caso 2D. Los filtros de Gabor 2D en el dominio espacial y de frecuencia tienen la siguiente forma:

$$G(X) = e^{-\pi((X-X_0)\alpha)^2} e^{-2\pi i U_0(X-X_0)} \quad (4.10)$$

$$G(U) = e^{-\pi\left(\frac{(U-U_0)}{\alpha}\right)^2} e^{-2\pi i X_0(U-U_0)} \quad (4.11)$$

donde $X=(x,y)$, $U=(u,v)$, X_0 , U_0 , α y β son vectores en \mathbb{R}^2 . X_0 representa las coordenadas en el dominio espacial del centro de la envolvente Gaussiana, U_0 las coordenadas cartesianas de las frecuencias espaciales de sintonía de los filtros, α es el vector que va a determinar la anchura de banda radial y la forma de la envoltura Gaussiana.

La implementación utilizada de los filtros de Gabor y que se expone a continuación ha sido propuesta por Sylvain Fischer en el Instituto de Óptica del CSIC en la elaboración de su PFC [12]. Se va a obtener una familia de filtros repartidos en N_f escalas y N_o orientaciones, a los que se le añade el filtro paso bajo. Estos parámetros se pueden elegir, aunque en la práctica se ha optado por analizar los resultados sólo para algunos casos, demostrando con ello la influencia que tiene la elección del número de canales dentro de la métrica de calidad. En cuanto al diseño de estos canales se ha optado por una distribución de los mismos en octavas ya que de esta forma la energía en cada canal de frecuencia tiende a mantenerse constante.

Los filtros se van a construir según la fórmula (4.11) y sin más que se proporcionen las frecuencias centrales y las varianzas. Así, para el cálculo del centro de los filtros (definidos en coordenadas polares) habría que tener en cuenta que se ha elegido un intervalo entre filtros de un octavo en módulo de frecuencia y π/N_o en orientación (sólo se necesita cubrir la mitad del plano de Fourier ya que al ser imágenes reales sus transformadas serán simétricas). Por tanto, los centros van a quedar

expresados como: $centro_{k,l} = \left(\frac{\rho_{\max}}{2^{k-1}}, l \cdot \frac{\pi}{N_o} \right)$. La varianza podrá ser determinada

directamente en la construcción de los filtros o bien ser un parámetro ajustable. En la figura 4.10 se muestran los filtros de Gabor para 1D y 2D. En la implementación aquí mostrada (propuesta por Matthias Keil del Instituto de Óptica) se ha realizado un ajuste de las varianzas para que los filtros se solapen adecuadamente, produciéndose un decrecimiento en $1/\rho$, donde ρ es el módulo de la frecuencia central del filtro. Además, se han modificado los Gabores para que tengan respuesta nula en frecuencia cero. El filtro paso bajo tendrá centro en cero y su varianza se define como la frecuencia de corte (-3dB) de la suma de los filtros de escala.

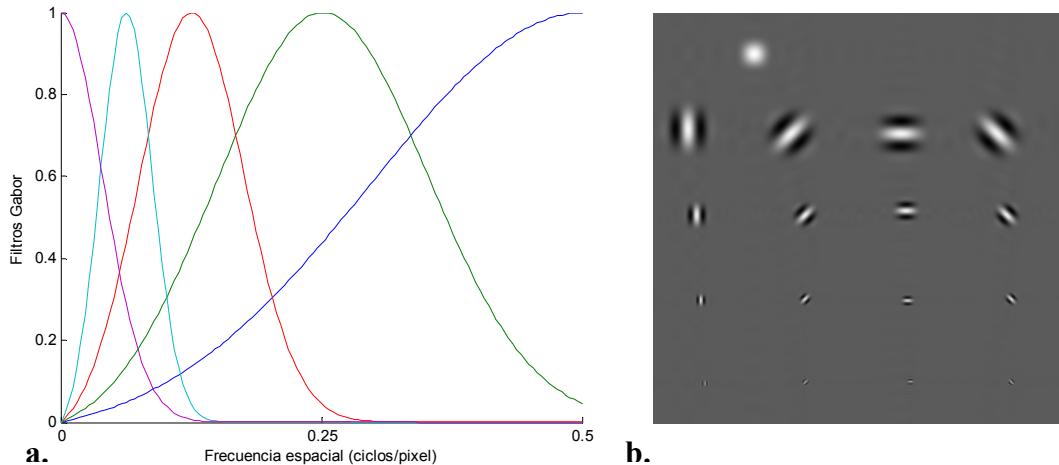


Figura 4.10: Filtros de Gabor para 4 escalas y 4 orientaciones más el paso bajo. **a.** En 1D en el dominio de frecuencia. **b.** En 2D en el dominio espacial, parte real. El residuo paso bajo aparece representado en la primera fila de la presente figura.

Partiendo de la familia de filtros que muestra la figura 4.10 conviene analizar una serie de factores que intervienen en el diseño de cualquier transformación y que se deben tener en cuenta para realizar las modificaciones que sean necesarias sobre los filtros Gabor implementados. Se estudiará, por tanto, la ortogonalidad y redundancia de coeficientes, la forma de los filtros en el dominio espacial (amplitud de lóbulos secundarios) y la reconstrucción.

La transformación es no ortogonal. Por ello, se va a tener un número de coeficientes en el espacio de la transformada igual al tamaño de la imagen por el número de canales ($N \cdot N \cdot (N_f \cdot N_o + 1)$). Esto representa una enorme redundancia de coeficientes por lo que se va a aplicar un submuestreo para reducirla en lo posible. Antes de aplicar este submuestreo los filtros deben ser modificados para que no se produzca aliasing, ya que, aunque su respuesta decrece rápidamente al alejarse del centro del filtro, ésta no es nula por lo que se tienen que poner a cero aquellos valores inferiores a un determinado umbral. En el presente trabajo se realizó un submuestreo disperso descrito en [12].

La reconstrucción será perfecta si se cumple: $\sum_n G_n(U)^2 = 1$, donde $G_n(U)$ es el filtro de Gabor en el dominio de frecuencia para el canal $n=(k,l)$. La suma al cuadrado de los filtros va a recibir el nombre de *roseta* y proporciona una representación gráfica de tipo de recubrimiento del plano de Fourier que lleva a cabo la

familia de filtros seleccionada. De la observación de la figura 4.10.a. se deduce claramente que esta condición no se verifica. En [12] se plantean una serie de correcciones para verificar en la medida de lo posible esta ecuación. Las propuestas para mejorar la reconstrucción fueron: ajuste de las escalas con pesos (de forma que la roseta valga uno en los centros de cada filtro), rotar los filtros de las escalas pares, dividir por la roseta en la reconstrucción y dividir los filtros por la raíz cuadrada de la roseta. Estas modificaciones pueden producir deformaciones en los filtros, por lo que se aplican sólo después de obtener una roseta lo más plana posible mediante el ajuste de los anchos de banda de los filtros. La figura 4.11 muestra las diferentes rosetas que pueden obtenerse al aplicar alguno de estos cambios, o la combinación de ellos. Para el uso de los filtros de Gabor en el sistema de medida de calidad se ha optado por la modificación que se muestra en la figura 4.11.e., es decir, escalas pares giradas y la división de los filtros por la raíz de la roseta.

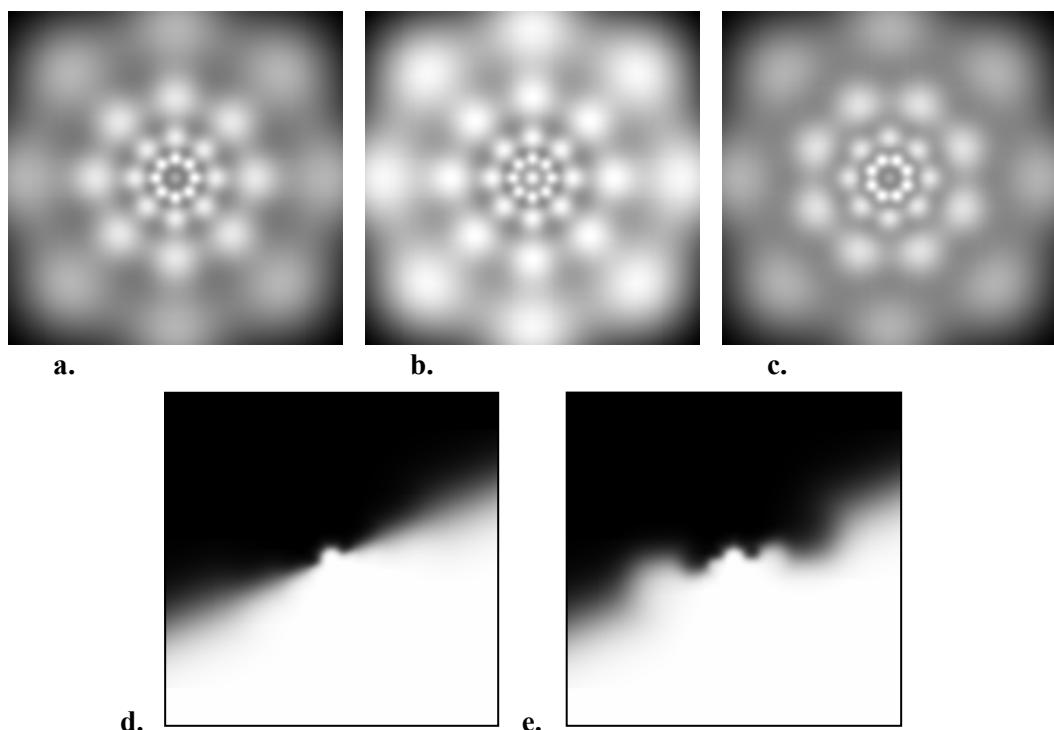


Figura 4.11: Posibles tratamientos de la roseta. **a.** Roseta para filtros sin modificar. **b.** Roseta con escalas ajustadas por pesos. **c.** Roseta con filtros escalas pares girados. **d.** División de los filtros por la raíz cuadrada de la roseta, se muestra solo la mitad del plano de frecuencia. **e.** Roseta con escalas pares giradas y filtros divididos por la raíz cuadrada de roseta (mitad del plano frecuencial).

En la práctica, el análisis o descomposición de la imagen de entrada mediante el banco de filtros se hará en el dominio de frecuencia como una multiplicación de los

filtros Gabor por la imagen transformada. En la figura 4.12 se muestra el procedimiento seguido para un canal en concreto ($k = 3$, $l = 2$), aunque la imagen de entrada no se corresponde con la que se tendría en realidad ya que no ha pasado por el preprocesado, ni por el módulo de adaptación de luminancia.

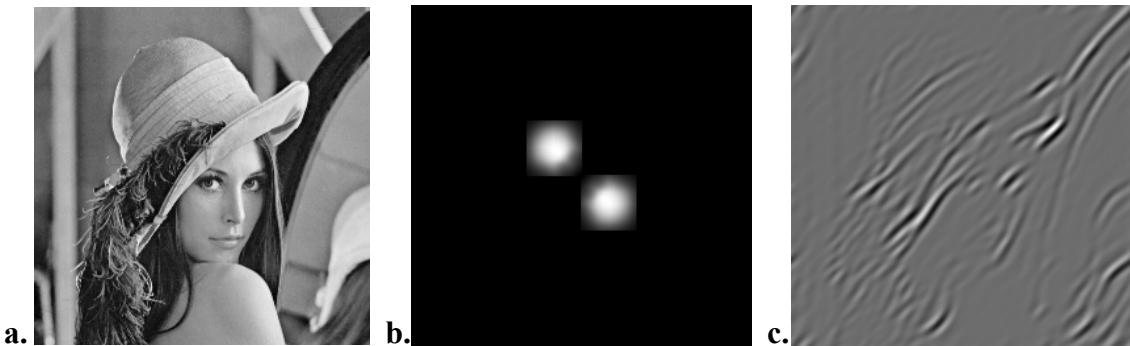


Figura 4.12: a. Imagen original Lena. b. Filtro Gabor de canal $n = (3,2)$ en frecuencia. c. Reconstrucción del canal (3,2)

En general, la reconstrucción de la imagen con filtros de Gabor es, como ya se ha señalado, no completa, principalmente porque se pierden las frecuencias espaciales altas (esquinas del plano de Fourier). Sin embargo, este problema podría solucionarse sin más que añadir al conjunto de filtros una escala más que cubriese estas frecuencias (el denominado residuo paso-alto). En la práctica, no va a ser necesario incluir este grado más de complejidad, ya que la información que se pierde debido a la no completitud es pequeña y no perceptible, más aún si tenemos en cuenta la forma de la CSF que atenúa bastante las altas frecuencias. Además, se debe tener en cuenta que el conservar toda la información visual no es el objetivo principal del SVH, sino el reconocimiento de objetos y escenas, por lo que algunos autores han considerado que la no completitud puede verse como una ventaja en cuanto a que supone una simplificación y se economiza tiempo de procesado.

En conclusión, se puede decir que las funciones de Gabor o paquetes de onda Gaussianos proporcionan una descomposición de la imagen en múltiples escalas y orientaciones en consonancia con el proceso que realiza el SVH. Estas funciones tienen algunas ventajas que ya se han comentado ampliamente y sus principales críticas se centran en la ausencia de ortogonalidad y en la no completitud de la reconstrucción de la imagen (síntesis) tras su descomposición (análisis). Es por ello que algunos autores han presentado otras alternativas como la que se muestra en el apartado siguiente.

4.4.2. Transformada Cortex. Filtros anillo.

La transformada Cortex fue propuesta por Watson [32] basándose en los descubrimientos y avances en el área de la visión y del funcionamiento del sistema visual. Sus principales objetivos consistían en encontrar una transformada que simulase la acción en las células de la corteza visual primaria y que además no tuviese una gran complejidad desde el punto de vista computacional. Para ello, enumera una serie de propiedades deseables en las que apoyarse para el posterior diseño.

En primer lugar, se buscan formas que se asemejen a las funciones Gabor en 2D ya que éstas presentan una gran similitud con los campos receptivos de las células de la corteza visual primaria, pero que no tengan los inconvenientes de la transformada Gabor que se han comentado en el apartado anterior. En segundo lugar, Watson busca una transformada en la que el número de canales que viene dado por la anchura de banda en frecuencia y en orientación, pueda seleccionarse. Otra de las propiedades que sería deseable es la de reconstrucción perfecta a partir de una suma de los canales. Por último, plantea como requisitos el que la transformada tenga estructura piramidal y que se pueda obtener mediante un algoritmo rápido.

La transformada Cortex que se ha implementado en este proyecto, no es la propuesta por Watson, sino una modificación de la misma que aparece dentro del VDP de Daly [7]. Los filtros *cortex* se obtienen de la multiplicación de una serie de filtros paso banda sintonizados en distintas frecuencias espaciales (filtros *anillo*) con otros que modelan la selectividad en orientación (filtros *abanco*). La familia de filtros obtenida tendrá un total de 31 canales que se corresponden con 5 filtros *anillo* ($N_f = 5$ escalas) combinados con 6 filtros *abanco* ($N_o = 6$ orientaciones), más un filtro paso bajo o filtro *base*. Los filtros *anillo*, como ya se ha comentado, van a cumplir dos funciones principalmente. Por una parte sirven como modelo de banco de filtros junto con un canal paso bajo y por otra, constituyen parte de la formación de la transformada Cortex.

$$\text{anillo}_k(\rho) = \begin{cases} \left. \text{mesa}(\rho) \right|_{\rho_{1/2}=2^{-(k-1)}} - \left. \text{mesa}(\rho) \right|_{\rho_{1/2}=2^{-k}} & k = 1..N_f - 1 \\ \left. \text{mesa}(\rho) \right|_{\rho_{1/2}=2^{-(k-1)}} - \left. \text{base}(\rho) \right|_{\rho_{1/2}=2^{-k}} & k = N_f \end{cases} \quad (4.12)$$

Los filtros *anillo* se van a calcular, por tanto, como diferencia de dos filtros *mesa*, que son filtros 2D paso bajo de tipo coseno alzado con la siguiente forma:

$$mesa(\rho) = \begin{cases} 1 & \rho < \rho_{1/2} - tw/2 \\ \frac{1}{2} \left(1 + \cos \left(\frac{\pi(\rho - \rho_{1/2} + tw/2)}{tw} \right) \right) & \rho_{1/2} - tw/2 < \rho < \rho_{1/2} + tw/2 \\ 0 & \rho > \rho_{1/2} + tw/2 \end{cases} \quad (4.13)$$

donde $\rho_{1/2} = 2^{-k}$ es la frecuencia para amplitud 1/2, $tw = \frac{2}{3}\rho_{1/2}$ es el valor para el

ancho de banda de transición y k es el índice que indica el canal de frecuencia radial. El filtro *base* que aparece en la ecuación (4.12) representa el canal paso bajo que se genera a partir de una Gaussiana para evitar en la medida de lo posible los efectos de “ringing”.

Presenta la siguiente forma:

$$base(\rho) = \begin{cases} e^{-(\rho^2/2\sigma^2)} & \rho < \rho_{1/2} + tw/2 \\ 0 & \rho \geq \rho_{1/2} + tw/2 \end{cases} \quad \sigma = \frac{1}{3} \left(\rho_{1/2} + \frac{tw}{2} \right) \text{ y } \rho_{1/2} = 2^{-(N_f+1)} \quad (4.14)$$

Los filtros descritos van a descomponer el plano de frecuencia en octavas, al igual que en el caso de los filtros Gabor. En la figura 4.13.a. se pueden observar estos filtros en 1D.

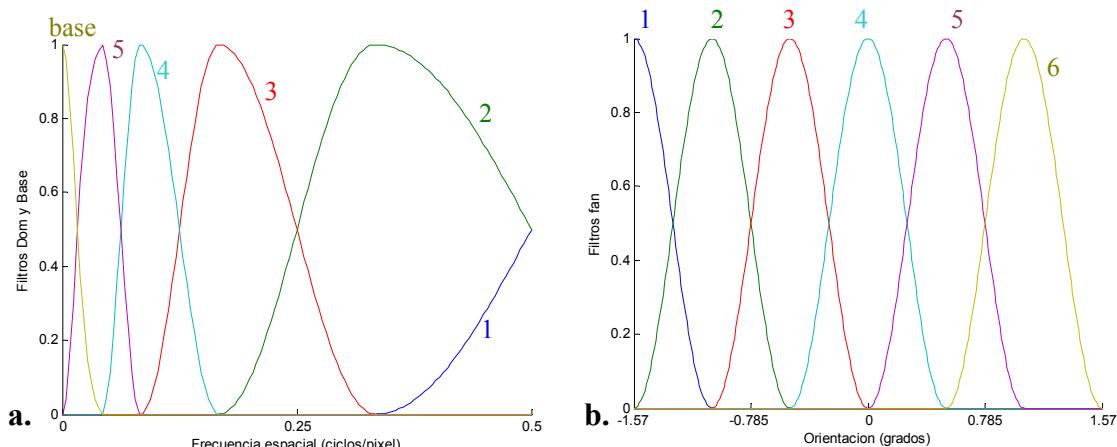


Figura 4.13: Filtros 1D de la transformada cortex. **a.** Filtros *anillo* para escalas k de 1 a 5 y *base*. **b.** Filtros *abanico* para orientaciones l de 1 a 6.

La selectividad en frecuencia se consigue con:

$$abanico_l(\theta) = \begin{cases} \frac{1}{2} \left[1 + \cos\left(\frac{\pi|\theta - \theta_c(l)|}{\theta_{tw}}\right) \right] & |\theta - \theta_c(l)| \leq \theta_{tw} \\ 0 & |\theta - \theta_c(l)| > \theta_{tw} \end{cases} \quad (4.15)$$

donde $\theta_c(l) = (l-1) \cdot \theta_{tw} - 90$ es la orientación del centro del filtro, $\theta_{tw} = \frac{180}{N_o}$ es la

anchura de banda de transición angular y l es el índice para indicar el canal de orientación. Estos filtros se muestran en la figura 4.13.b.

Finalmente, se obtiene el banco de filtros *cortex* de 31 canales que se muestra en la figura 4.14., mediante el producto de los filtros anillo con los filtros abanico.

$$cortex_{k,l}(\rho, \theta) = \begin{cases} anillo_k(\rho) \cdot abanico_l(\theta) & k = 1..N_f ; l = 1..N_o \\ base(\rho) & k = N_f + 1 \end{cases} \quad (4.16)$$

Esta implementación de la transformada Cortex presenta cuatro modificaciones básicas frente a la propuesta por Watson:

- Se utilizan funciones Hanning en las transiciones en lugar de Gaussianas para evitar el efecto de “ringing”.
- Las transiciones en orientación están definidas en grados.
- La utilización de una Gaussiana en el paso bajo para eliminar el “ringing” en estas frecuencias (filtro *base*).
- No se utiliza un filtro paso alto para recoger las altas frecuencias de forma equivalente a como se hace con el paso bajo. Las altas frecuencias también están divididas en orientaciones (como puede observarse en la figura 4.14 para $k = 1$).

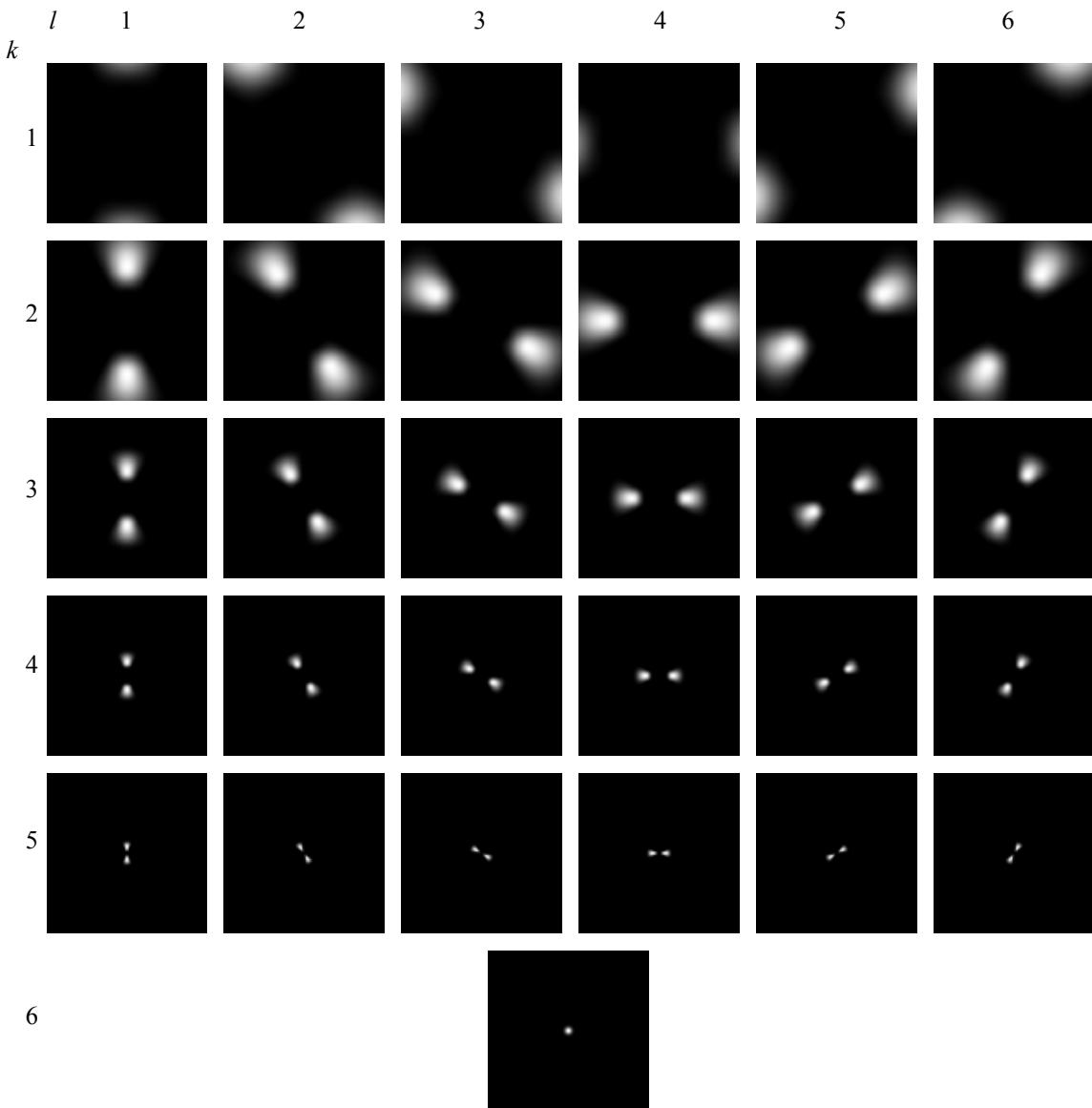


Figura 4.14: Transformada *cortex*. La figura de la esquina superior derecha se corresponde con $cortex_{1,6}(\rho, \Theta)$ y así sucesivamente. En la fila inferior se representa el paso bajo $cortex_6(\rho, \Theta) = base(\rho)$.

4.5. Cálculo de contrastes.

La conversión de la imagen original y la distorsionada a unidades de contraste es fundamental dentro del modelado del SVH y por tanto, también lo será en los sistemas de medida de calidad que se basan en estos modelos. El contraste percibido es un atributo básico de la imagen. En el procesamiento de imagen digital se han propuesto varias técnicas de medida y evaluación de contraste, sin que exista una definición única. En el Capítulo 2 se ha hecho ya un desglose de algunas de las fórmulas existentes para

el cálculo de contraste en relación con los mecanismos de la visión. Una vez realizado ese análisis lo que se pretende es posicionar la conversión a contraste como un elemento más en las técnicas de medida de calidad y comentar algunos de los problemas que se plantean en la práctica.

El cálculo de contrastes se va a realizar en el dominio espacial sobre el conjunto de imágenes obtenido de la salida del banco de filtros. Se tendrán, por tanto, valores de contraste por canales y píxel a píxel que muestran la variación del contraste en cada posición de la imagen. En el caso de una descomposición en orientación y frecuencia se tendrán $N_f N_o$ imágenes de contraste tanto para la imagen original como para la imagen distorsionada. Si el sistema realiza una descomposición únicamente en canales de frecuencia se obtendrán 2 conjuntos de N_f imágenes cada uno. En el entorno gráfico se ofrecen cuatro posibilidades para elegir el método de cálculo de contraste, adaptadas a los dos casos comentados (selectividad en orientación y frecuencia o descomposición sólo en frecuencia) y que se corresponden con las fórmulas de la siguiente tabla:

1.	$C_{Global_k}(x, y) = \frac{b_k(x, y)}{b_{Nf+1}}$	$C_{Global_{k,l}}(x, y) = \frac{b_{k,l}(x, y)}{b_{Nf+1}}$
2.	$C_{Local_k}(x, y) = \frac{b_k(x, y)}{b_{Nf+1}(x, y)}$	$C_{Local_{k,l}}(x, y) = \frac{b_{k,l}(x, y)}{b_{Nf+1}(x, y)}$
3.	$C_{Peli_k}(x, y) = \frac{b_k(x, y)}{\sum_{j=k+1}^{Nf+1} b_j(x, y)}$	$C_{Peli_{k,l}}(x, y) = \frac{b_{k,l}(x, y)}{b_{Nf+1}(x, y) + \sum_{j=k+1}^{Nf} b_{j,l}(x, y)}$
4.	$C_{Lubin_k}(x, y) = \frac{b_k(x, y)}{\sum_{j=k+2}^{Nf+1} b_j(x, y)}$	$C_{Lubin_{k,l}}(x, y) = \frac{b_{k,l}(x, y)}{b_{Nf+1}(x, y) + \sum_{j=k+2}^{Nf} b_{j,l}(x, y)}$

Tabla 4.1: Fórmulas para el cálculo de contraste en sistemas multicanal.

Se ha seguido la notación empleada en el diseño del banco de filtros, donde el canal paso bajo es b_{Nf+1} . Para una correcta interpretación de los sumatorios se debe

recordar que los filtros de escalas de frecuencia mayores se almacenan en índices de k menores.

Estas definiciones de contraste son las mismas que se plantearon en el Capítulo 2. Las dos primeras han sido propuestas por Daly [7] y calculan el contraste dividiendo en cada punto bien por un valor medio del paso bajo o bien por el valor exacto para esa posición. El C_{Peli} [23] y C_{Lubin} [26] son similares y utilizan en el denominador una suma de las escalas que están por debajo de la escala k en frecuencia (índices de k mayores) para una misma orientación. Como ya se comentó, el contraste de Peli se adapta bien al contraste “subjetivo” (el que percibiría un observador) y C_{Lubin} da buenos resultados para escenas naturales. Algunas fórmulas más para el cálculo de contraste pueden encontrarse en [9], donde se obtienen contrastes isótropos haciendo previamente un filtrado de la imagen de entrada direccional. En todas las definiciones mostradas en la tabla 4.1 subyace la idea común de una expresión del contraste como un radio adimensional, con un valor de amplitud local y un valor medio de luminancia, en sintonía con las propuestas de contraste de Michelson y Weber.

En la figura 4.15 puede verse una comparativa de las distintas opciones propuestas. En general, las fórmulas presentan pocas diferencias si se consideran las escalas de frecuencia menores (índices k mayores), ya que en estos casos tienen ecuaciones similares. Por eso, se ha escogido una escala de frecuencia $k = 3$ para que las imágenes mostradas no sean iguales. Además, observando la figura se aprecia claramente que la descomposición en canales elegida es para el banco de filtros *anillo*, puesto que en todos los ejemplos se observa que están presentes todas las orientaciones. Estos mapas de contraste se han obtenido sin tener en cuenta el efecto de la adaptación de luminancia ni de la CSF, ya que aquí, lo único que se pretende es mostrar la diferencia entre las distintas opciones de contraste propuestas. La diferencia de cualquiera de ellas con la $b_3(x,y)$ es notoria ya que se percibe el aumento de contraste sobre todo en aquellas zonas en las que hay un cambio brusco de nivel de luminancia, como por ejemplo los bordes del tejado, marco de las ventanas...

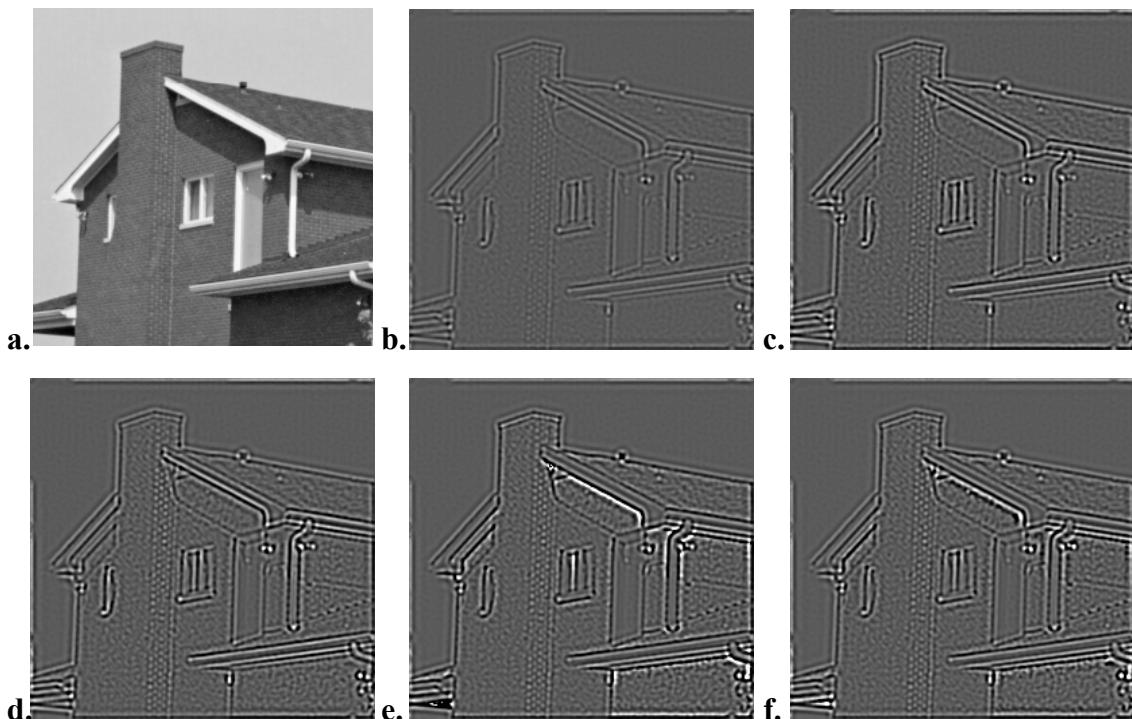


Figura 4.15: a. Imagen original *House*. b. Banda imagen para escala $k = 3$, $b_3(x,y)$.
c. $C_{Local_3}(x,y)$ d. $C_{Global_3}(x,y)$ e. $C_{Peli_3}(x,y)$ f. $C_{Lubin_3}(x,y)$.

Uno de los problemas principales que se encuentran en la práctica al incluir estas definiciones dentro de un sistema de medida de calidad viene derivado del uso de determinados modelos de CSF. La dificultad reside en el hecho de que la sensibilidad al contraste para bajas frecuencias varía de forma muy significativa de unos modelos a otros. Esta variación entre modelos es consecuencia de la complejidad de la medida de la CSF. Por una parte depende de gran variedad de parámetros y por otra, con los métodos de medida comentados es imposible obtener un valor en baja frecuencia. A todo esto hay que añadir que los autores escalan y modifican la CSF dependiendo de la aplicación. Así, por ejemplo, en algunas aplicaciones de compresión hacen que la CSF tenga una característica paso bajo para que no se perciba una degradación en la calidad cuando el observador aumenta la distancia de visualización (con lo que aumentarían las frecuencias espaciales). Por todo ello, se hace necesario modificar adecuadamente aquellos modelos de CSF en los que el valor del residuo paso bajo de la imagen es cero ya que esto supone que se tendrían valores de contraste indeterminados ($b_{Nf+1} = 0$ en el denominador).

4.6. Enmascaramiento espacial.

Como se vio en la sección 2.2.4. el enmascaramiento espacial es la reducción de la visibilidad de un componente en la imagen debido a la presencia de otro que lo oculta. Si no existiese el enmascaramiento la imagen original (su contenido) no influiría en la visibilidad del error, que en este caso es la diferencia entre la imagen de referencia y la distorsionada. Sin embargo, lo que sucede en realidad es que la imagen original sí afecta al umbral de visibilidad del error. Los efectos del enmascaramiento se modelan mediante funciones de elevación del umbral (ver figura 2.12), que describen el incremento que debe producirse en el valor de un estímulo para que resulte visible dentro la imagen. En definitiva, el propósito final será determinar si una distorsión es visible dentro de la imagen. Por tanto, el contenido de la imagen es la fuente del enmascaramiento y a efectos prácticos se considerará ruido, mientras que la distorsión es la señal cuya detectabilidad se quiere predecir.

La correcta incorporación del fenómeno de enmascaramiento en modelos de medida de calidad perceptible resulta difícil. La razón principal es, de nuevo, el diseño de los experimentos de los cuales se obtienen los resultados del enmascaramiento. Los umbrales de visibilidad varían significativamente dependiendo del ancho de banda, de la fase, la orientación y el conocimiento del observador de los estímulos utilizados. Por ello, definir un modelo general supone extrapolar estos resultados a imágenes de mayor complejidad.

Los primeros test realizados utilizaban estímulos sinusoidales que eran enmascarados también por sinusoides. Posteriormente se realizaron test con Gabores y sinusoides de diferentes frecuencias, fases y valores de contraste. Se descubrió que el enmascaramiento no dependía únicamente de la energía dentro de una banda, sino también entre bandas y distintas orientaciones. En [11] se presentan tres modelos básicos de enmascaramiento de diferente complejidad, según incorporen o no los efectos entre canales. La implementación llevada a cabo dentro del sistema de calidad de este proyecto se centrará en enmascaramiento dentro de un mismo canal, siendo una posible línea de investigación la mejora de este modelo.

Otra de las características que se pueden tener en cuenta y que ha sido también investigada está relacionada con el tipo de máscara utilizada y el aprendizaje del

observador [10]. Así, se descubrió que cuando se utilizaban ruidos aditivos de banda ancha para enmascarar, se necesitaban valores de umbral mucho más elevados para percibir el estímulo (normalmente una sinusoide). La situación además se complicaba con el efecto de aprendizaje. Los test mostraron que si se le daba suficiente tiempo al observador, éste se familiarizaba con la máscara de ruido y por tanto, podía ver el estímulo con un valor de umbral semejante al que se obtiene con máscaras sinusoidales. En conclusión, estos resultados revelaban que el grado de enmascaramiento dependía del conocimiento que el observador poseyese de la imagen que se le mostraba. Esto presenta un problema en cuanto al diseño de un modelo general puesto que no se puede saber a priori cual es el conocimiento que el observador tiene de la imagen que se le va a presentar. Sin embargo, sí se pueden sacar algunas conclusiones relativas al contenido de la imagen. Así, al ver una imagen, estructuras simples como bordes y zonas uniformes enmascaran menos que las texturas. Esto se debe a que el observador sabe o puede predecir fácilmente que apariencia deben tener los bordes y zonas uniformes, lo que no sucede generalmente con texturas.

Con las características comunes que se han comentado como base (modelo de elevación de umbral, enmascaramiento dentro de un mismo canal y distinción entre bordes, texturas y regiones uniformes) surgen dos implementaciones distintas para el enmascaramiento. Ambas han sido propuestas y utilizadas en el ámbito de sistemas de medida de calidad perceptible [7, 25, 26, 34] y están relacionadas de forma directa con el uso que se haga de la CSF dentro del sistema. Además, la aproximación común es modelar el enmascaramiento como un fenómeno localizado espacialmente de forma que la respuesta de un filtro a una determinada localización en el espacio es enmascarada sólo por la respuesta del mismo u otros filtros en la misma localización.

4.6.1. Enmascaramiento con CSF como filtro.

La función implementada fue propuesta por Daly [7]. Se establece un modelo para el enmascaramiento de contraste en bajas frecuencias similar al que se obtiene de experimentos con sinusoides (pendiente de la curva de umbral de elevación en figura 2.14 en torno a 0,7) y para frecuencias medias y altas el equivalente al conseguido al utilizar un ruido como máscara (pendiente de la curva en torno a 1). Este razonamiento va a suponer que las áreas de la imagen con contenido en frecuencias medias y altas van

a ser más complejas y por tanto menos familiares o predecibles para el observador. Sin embargo, las regiones que tengan sólo bajas frecuencias son bastante más sencillas o de fácil aprendizaje por lo que el aumento que debe producirse en la señal para que resulte visible en estas zonas es menor, es decir, enmascaran menos. En la función propuesta a continuación esto se traduce en un valor de $s = 0,7$ en baja frecuencia y $s = 1$ en el resto.

$$Te_n(x, y) = \left(1 + \left(k_1 (k_2 |b_n(x, y)|)^s \right)^4 \right)^{1/4} \quad (4.17)$$

Según la notación empleada hasta este momento $b_n(x, y)$ es la banda de la imagen en el canal n (escala k y orientación l , en caso de que exista) a la salida del banco de filtros y k_1 y k_2 son constantes. La CSF utilizada como filtro ha normalizado los valores de todos los canales. Por tanto, la función de elevación de umbral en caso de que no se haya producido enmascaramiento para un píxel concreto no va a alterar el valor de umbral de visibilidad dado por la CSF, tomando valor uno. Para los demás píxeles la magnitud de Te será siempre mayor que la unidad y dependerá del grado de enmascaramiento.

Hasta este momento la función Te planteada como modelo hace referencia exclusivamente a la imagen original (canales $b_n(x, y)$). Sin embargo, en la determinación de los umbrales de visibilidad y el enmascaramiento la imagen distorsionada también juega un papel importante. Este hecho se pone de manifiesto sin más que se analice el caso de una imagen distorsionada que presenta emborronamiento. Así, si el cálculo del mapa de elevación de umbral se basa únicamente en la imagen de referencia se van a obtener valores que sobreestiman el grado de enmascaramiento. Esto se debe a que se está calculando el enmascaramiento que posibilitan los bordes de la imagen original y estos no van a estar presentes en el caso de la imagen distorsionada. De igual forma sucede con los artefactos generados en esquemas de compresión en los que se produce una reducción en el contraste de las texturas con lo que el enmascaramiento visual es menor del que se estimaría considerando únicamente la imagen original. Se definen, por ello, umbrales de elevación para ambas imágenes siendo $Te_n(x, y)$ el de la imagen original y $\bar{Te}_n(x, y)$ el correspondiente a la imagen distorsionada en el canal n . El umbral final $Tem_n(x, y)$ se elige entre el mínimo de los dos y recibe el nombre de enmascaramiento mutuo [7].

$$Tem_n(x, y) = \min[Te_n(x, y), \overline{Te}_n(x, y)] \quad (4.18)$$

La figura 4.16 muestra mapas de elevación umbral para distintas bandas, siendo *Lena* la imagen de referencia y sometiéndola a una compresión con JPEG2000 con tasa 0,5 (PSNR = 30,46 dB) para formar la imagen distorsionada. Los píxeles de la figura de color negro son aquellos puntos de la imagen que para ese canal no presentan enmascaramiento y por tanto, tienen un valor igual a uno.

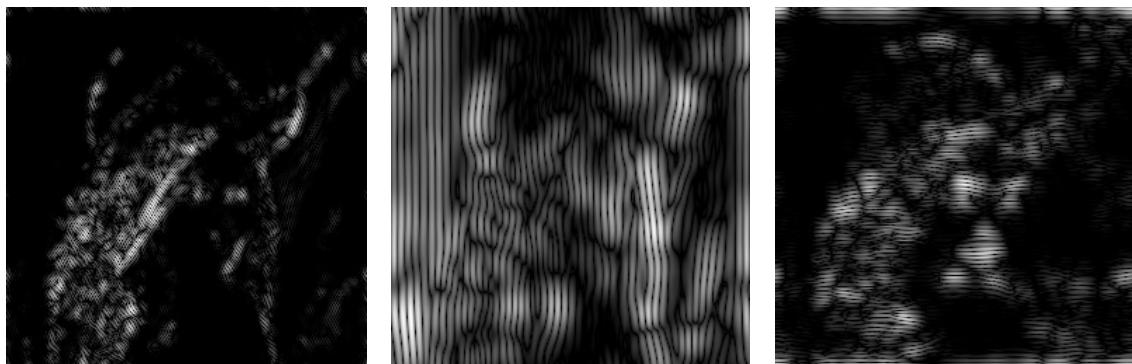


Figura 4.16: Mapas de umbral de enmascaramiento mutuo para imagen *Lena*. De izquierda a derecha $Tem_{2,3}(x, y)$, $Tem_{4,2}(x, y)$, $Tem_{3,1}(x, y)$.

4.6.2. Enmascaramiento con CSF por pesos.

Se modela de nuevo el enmascaramiento como un proceso de elevación de umbral para cada banda, que en este caso depende del contenido local de la imagen. La principal diferencia consiste en la forma en que se incluye aquí el hecho de que según sea el estímulo de predecible, las pendientes de la curva (figura 2.14) serán diferentes. En el caso anterior en función de la banda de frecuencia para la que se calculase el umbral, el valor del exponente variaba. En este caso, se va a realizar una clasificación de la imagen en distintas regiones sobre las que se aplicará posteriormente la elevación de umbral que corresponda. Por tanto, cada píxel de la imagen en cada canal va a pertenecer a una de las siguientes regiones:

- **Bordes:** se obtienen fácilmente mediante cualquier detector de bordes. En esta implementación se ha utilizado un detector tipo Sobel.
- **Regiones uniformes:** esta clasificación se va a llevar a cabo mediante la utilización de la CSF. Los puntos en cada canal o banda imagen cuyo valor de contraste

sea menor que el umbral que proporciona la CSF se considerarán regiones uniformes. Para calcular el umbral base de un canal se hace la media de los valores de sensibilidad de la CSF restringidos a la banda de frecuencias que ocupa dicho canal. El inverso de este valor es precisamente el umbral medio de visibilidad que se expresa como TH_O^n . Por tanto, si $C_n(x_1, y_1) < TH_O^n$ el píxel (x_1, y_1) pertenece a una región uniforme, donde $C_n(x_1, y_1)$ es el valor de contraste para (x_1, y_1) en el canal n .

- Texturas: los píxeles pertenecientes a texturas son aquellos que no forman parte de bordes, ni de regiones uniformes y en los que $C_n(x_1, y_1) > TH_O^n$.

Finalmente, el enmascaramiento se implementa como una curva de elevación de umbral:

$$Te_n(x, y) = \begin{cases} TH_O^n & , \text{si } (x, y) \in \text{region uniforme} \\ TH_O^n \left(\frac{C_n(x, y)}{TH_O^n} \right)^{0,7} & , \text{si } (x, y) \in \text{borde.} \\ TH_O^n \left(\frac{C_n(x, y)}{TH_O^n} \right)^1 & , \text{si } (x, y) \in \text{textura.} \end{cases} \quad (4.19)$$

Las regiones uniformes no van a generar apenas enmascaramiento por lo que no se produce elevación en el umbral de visibilidad y el valor de la función es directamente el dado por la CSF. Para los bordes y texturas el umbral aumenta gradualmente cuando el contraste se incrementa. La diferencia básica es el exponente, con valor 0,7 en el primer caso y 1 en el segundo. Esto supone una pendiente de la curva de umbral de elevación (figura 2.14) mayor para texturas, que al enmascarar más hacen que sea necesario un valor mayor de contraste para que el estímulo sea visible. En este caso se puede observar que la función Te_n base no va a ser uno ya que aquí aún no se ha tenido en cuenta la CSF, los canales b_n con los que se han calculado los contrastes C_n no están normalizados respecto de la CSF. Por tanto, un valor de 1 del modelo anterior de enmascaramiento equivale en este caso a TH_O^n . Además, el razonamiento hecho en el caso anterior para incorporar la influencia de la imagen distorsionada también se debe aplicar aquí en la forma que indica la ecuación (4.18).

La figura 4.17 muestra los umbrales de enmascaramiento conseguidos mediante este método. Se han obtenido para la imagen de referencia *Lena* y la CSF del modelo 1 (Rust). Los mapas de enmascaramiento son similares a los de la figura 4.16 con la salvedad de que en este caso no son los umbrales de enmascaramiento mutuo (no se ha considerado la imagen distorsionada en el ejemplo mostrado). Además, la gran diferencia respecto a las figuras 4.16 es que aquí el valor de los píxeles de color negro (los que no presentan enmascaramiento) es TH_O^n , no 1 como en el caso anterior.

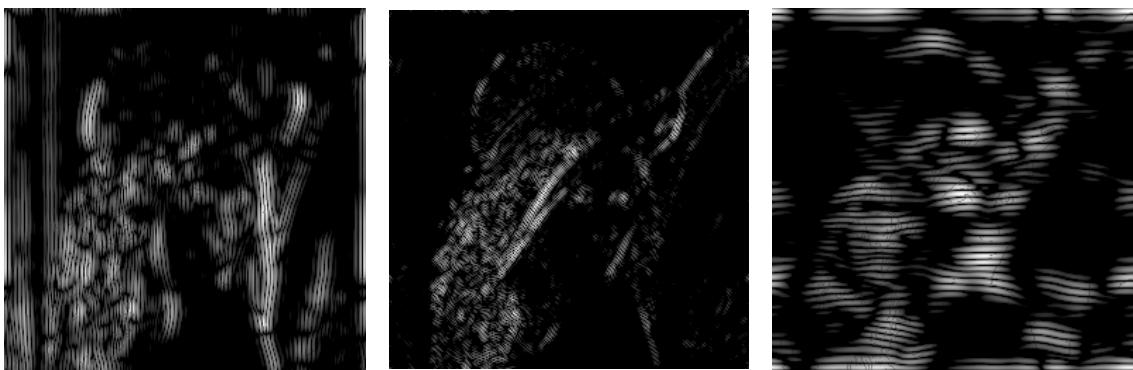


Figura 4.17: Mapas de umbral de enmascaramiento mutuo para imagen *Lena*. De izquierda a derecha $Te_{3,4}(x,y)$, $Te_{2,3}(x,y)$, $Te_{4,1}(x,y)$.

En conclusión, el enmascaramiento va a cambiar el umbral de detección del error en la imagen dependiendo del contenido local de la misma. Este nuevo umbral se usará posteriormente para determinar si las diferencias entre la imagen original y la distorsionada son visibles en cada punto.

4.7. Mapas e índices de error.

El paso final en un sistema de medida de calidad perceptible es la combinación de los mecanismos descritos anteriormente para obtener, o bien una imagen en la que cada píxel tiene un valor indicativo de la visibilidad del error en esa localización, o bien un valor numérico que describa la calidad global de la imagen, o ambos. Todo ello sin olvidar que tanto el mapa como el índice deben corresponderse con la valoración que se obtendría al mostrar la imagen de referencia y la distorsionada a un observador.

En el modelo descrito hasta el momento se ha asumido, además, una descomposición de la imagen de referencia y la distorsionada en canales. El conjunto de imágenes resultado de este filtrado varía en número según la elección que se haga del banco de filtros. Con todo, siempre se tendrá que integrar la información procedente de todos estos canales para obtener un resultado final. El procedimiento lógico para reducir la dimensión de esta información a 2D (mapa) o 1D (índice) consiste en sumar las respuestas de los distintos canales utilizando la métrica de Minkowski. Un paso previo a esta suma es el cálculo de la visibilidad del error en cada canal, que denominamos $V_n(x,y)$ y se expresa como:

$$V_n(x,y) = \frac{C_n(x,y) - \bar{C}_n(x,y)}{T_{em_n}(x,y)} \quad (4.20)$$

donde $C_n(x,y)$ y $\bar{C}_n(x,y)$ son respectivamente el contraste de la imagen original y la distorsionada en el canal n . $T_{em_n}(x,y)$ es el umbral de visibilidad definido en el apartado anterior.

Si se analiza el significado de estos mapas de visibilidad de error por canales se deben considerar dos interpretaciones distintas según la elección hecha para el uso de la CSF y el cálculo de los umbrales de enmascaramiento. Esto es:

- *Caso CSF como filtro:* el numerador $C_n(x,y) - \bar{C}_n(x,y)$ representa el contraste del error y $T_{em_n}(x,y)$ es el dado por la ecuación (4.18) con $T_{e_n}(x,y)$ descrito en (4.17). Así, un valor de 1 para el enmascaramiento dejaría en $V_n(x,y)$ el contraste del error. Sin embargo, hay que recordar que en este caso el contraste se ha calculado con bandas que han sido previamente filtradas con la CSF. Por tanto, el valor de $V_n(x,y)$ sería el contraste del error dividido por el umbral de visibilidad, ya que en esta expresión de contraste la CSF ya se ha utilizado como una normalización.
- *Caso CSF por pesos:* en este caso los $T_{e_n}(x,y)$ con los que se calcula el $T_{em_n}(x,y)$ vienen dados por la ecuación (4.19). El contraste de error se obtiene con la imagen en bandas. Por tanto, al igual que en el caso anterior, se va a tener el contraste dividido por el umbral de visibilidad, que en caso de que no exista enmascaramiento, es directamente el valor dado por la CSF (ecuación (4.19)).

Por consiguiente, aunque las interpretaciones son distintas el resultado es coherente en ambos casos. Así, por ejemplo, un valor de 1 en $V_n(x,y)$ indicaría que ese error en la banda n y en el píxel de posición (x,y) estaría justamente en el umbral de visibilidad. Una vez obtenidos los mapas de visibilidad de error se presentan dos opciones:

1. El cálculo final se obtiene directamente de los mapas $V_n(x,y)$ anteriores sin más que aplicar, como ya se ha mencionado, una métrica de Minkowski. Se va a obtener un mapa de visibilidad del error total, reduciendo la dimensionalidad mediante la siguiente fórmula:

$$V_T(x,y) = \left(\sum (V_n(x,y))^\beta \right)^{1/\beta} \quad (4.21)$$

donde β tiene un valor de 4. Se hace una suma de los canales primero en orientación (en caso de que exista) y después en frecuencia. La magnitud de $V_T(x,y)$ en cada punto va a representar cuántas veces por encima del umbral de visibilidad se sitúa el contraste del error.

Finalmente, para calcular el índice de calidad perceptible (*ICP*) de la imagen se utilizará de nuevo la métrica de Minkowski:

$$ICP = \left(\frac{1}{N \cdot M} \sum_{x=1}^N \sum_{y=1}^M (V_T(x,y))^\gamma \right)^{1/\gamma} \quad (4.22)$$

donde el valor de γ es 3. Se elige este valor siguiendo la recomendación hecha en otros modelos [25, 26], ya que es la que proporciona una mejor correlación con resultados obtenidos mediante test. Así, un valor de $\gamma = 1$ daría una calidad proporcional a la media de los errores en todos los píxeles, mientras que para valores mayores de γ la calidad sería proporcional al máximo. Para poder comparar este valor con los que se obtienen mediante el uso de técnicas DSCQS, DSIS, etc. es necesario que se den en un mismo margen. Es decir, en este caso se requiere un margen de 1 a 5 que se consigue mediante la siguiente ecuación:

$$ICP_{1-5} = \frac{5}{1 + p \cdot ICP} \quad (4.23)$$

donde p es una constante de escalado cuyo valor se obtiene al buscar una correlación máxima entre este índice y los resultados de test subjetivos. En [25, 26] se da un valor de $p = 0,8$.

2. Este segundo procedimiento [7] consiste en convertir los mapas de visibilidad del error en probabilidades de detección y posteriormente, reducir la dimensionalidad del problema. La conversión a probabilidades se hace mediante el uso de una función psicométrica como la que muestra la figura 4.18. Es una gráfica en la cual se obtiene la proporción de las respuestas afirmativas en una tarea de detección o discriminación en función del valor del estímulo. En este caso lo que muestra es la probabilidad de detección en función de $V_n(x,y)$, que es el contraste del error normalizado. La ecuación que describe esta gráfica es:

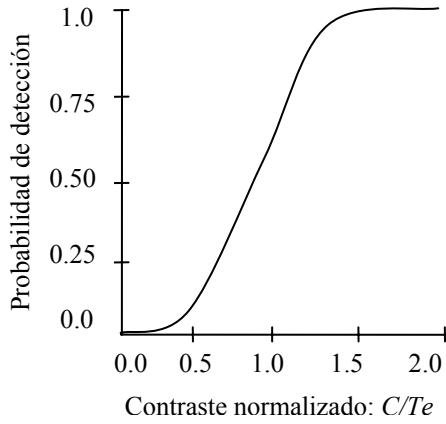


Figura 4.18: Función psicométrica.

$$P_d(C) = 1 - e^{-(C/T_e)^\alpha} \quad (4.24)$$

El parámetro α da la forma de la pendiente de la curva, que en la implementación realizada tiene un valor de 4. Para el cálculo de los mapas de probabilidad de detección en cada canal y para cada píxel, la fórmula anterior quedaría como:

$$P_n(x,y) = 1 - e^{-(V_n(x,y))^\alpha} \quad (4.25)$$

Finalmente, una vez que las probabilidades de detección se han calculado para cada canal, éstas se van a combinar en una única imagen en la que el valor de cada píxel representa la probabilidad de que el error sea detectado en esa posición. Esta imagen de probabilidad total se obtiene mediante la siguiente ecuación:

$$P_T(x,y) = 1 - \prod_n (1 - P_n(x,y)) \quad (4.26)$$

Los errores detectables y que están por consiguiente por encima del umbral se corresponderán con valores de $P_T(x,y)$ de uno. Este mapa de probabilidad no describe la apariencia del contraste del error, sino que simplemente da una indicación de la localización del mismo. Para pasar de este mapa de probabilidades a un índice de

calidad global en [7] se ha seleccionado simplemente el máximo de $P_T(x,y)$ (equivaldría a una métrica de Minkowski con un valor del exponente que tiende a infinito).

Con esto se obtendrán finalmente los mapas e índices que muestran la calidad que el observador percibe al visualizar una imagen de referencia con su correspondiente imagen distorsionada. En el capítulo 5 se verán los resultados obtenidos mediante la combinación de estos modelos y se mostrarán los mapas de visibilidad de error para algunas de las imágenes test utilizadas.

Comportamiento y comparativa de los modelos de medida de calidad.

5

En este capítulo se van a mostrar y analizar los resultados obtenidos para diferentes imágenes y distorsiones que son evaluadas con los sistemas de medida de calidad implementados (Capítulo 4) y disponibles en el interfaz de usuario (anexo B). Se debe tener en cuenta que la precisión del resultado se evalúa en función de los datos obtenidos mediante realización de test subjetivos, calculando la correlación entre estos datos y los proporcionados por el modelo de medida. Así, en el apartado 5.1 se va a detallar el proceso de medida seleccionado para la obtención de los datos subjetivos que se requieren para la valoración del modelo. Esto incluirá desde la descripción de las imágenes hasta la secuencia de tiempos utilizada en la visualización de las mismas. El resto de apartados se dedicarán a estudiar la influencia de los componentes específicos del modelo sobre la fidelidad general del resultado que proporciona la métrica. Se va a considerar el peso o importancia de los diversos factores que componen un modelo de medida (adaptación, CSF, contrastes...) incluyendo las diferentes opciones posibles para cada uno de ellos. Finalmente, se evaluarán los resultados obtenidos para algunas métricas de calidad no perceptible como las descritas en el Capítulo 3.

5.1. Validación de modelos de medida de calidad perceptible.

Una de las principales dificultades a la hora de trabajar con modelos de calidad perceptible es la complejidad y coste del proceso de validación. En primer lugar, se debe obtener un resultado global del modelo de medida integrando toda la información procedente de los distintos canales en un único número que, posteriormente, se comparará con datos obtenidos directamente de los observadores. Es necesario, por tanto, disponer de estas opiniones subjetivas y puesto que no hay una base de datos disponible será necesario realizar test con observadores para conseguirla. En el laboratorio (Instituto de Óptica del CSIC) se ha utilizado la recomendación ITU-R BT.500-10 (“Methodology for the subjective assessment of the quality of television pictures”), en la que se describen los modelos de medida de calidad de imagen subjetivos que se comentaron brevemente en el apartado 3.2 de este proyecto. En concreto, se ha seleccionado el método **DSCQS** (“Double-stimulus continuous quality-scale”) que permite medir la calidad de un sistema con relación a una referencia. Es un método cíclico en el que se presentan al observador la imagen original y la imagen test de forma aleatoria y éste debe puntuar la calidad de ambas. Además, de los métodos que aparecen en ITU-R BT.500-10, es el que presenta una mayor robustez minimizando los efectos contextuales (dar una puntuación excesivamente elevada a una imagen con un error que es perceptible después de haber mostrado imágenes con errores muy bajos).

A continuación, se va a enumerar algunas de las características más importantes del método DSCQS y su aplicación concreta en el laboratorio.

1. Condiciones de visualización en el laboratorio.

Las imágenes se visualizan en un monitor SONY F520 de 21" que ha sido calibrado previamente (ver apartado 4.1). La habitación en la que se realiza el proceso de medida carece de ventanas y la iluminación de la misma se controla para que se cumplan las especificaciones que se proporcionan en la recomendación. Se ha elegido una distancia de visualización de 60 cm., lo que combinado con la exigencia de la norma de que el ángulo de observación máximo sea de 30° permite realizar medidas con dos observadores a la vez.

2. Selección del material para el test.

Se han utilizado las cinco imágenes de referencia que se muestran en el Anexo A. Se ha realizado esta selección para permitir analizar posteriormente el comportamiento que va a tener el sistema frente a dos clases de imágenes claramente diferenciadas. Por una parte tendremos una muestra de tres imágenes médicas radiológicas MRI (figuras A.1, A.2 y A.3) y por otra, dos imágenes que son de uso frecuente en sistemas de compresión (figuras A.4 y A.5). Para la generación de imágenes distorsionadas a partir de las cinco anteriores se han utilizado técnicas de compresión, de marcado de agua y ruido gaussiano. Con ello, se conseguirá un conjunto de 30 imágenes (5 de referencia y por cada una de ellas otras 5 con distintos valores de PSNR, ver tabla A.1). Un conjunto mayor de imágenes plantearía un problema en cuanto a duración del test. Si se somete a los observadores a test demasiado largos se produciría un efecto de cansancio y perdida de atención con lo que los resultados no resultaría fiables. Esto podría solucionarse si se dividiese el conjunto de imágenes en dos sesiones por cada observador, pero esto también resultaría costoso en cuanto a requisitos temporales. Además, el conjunto de imágenes propuesto representa una muestra suficiente para el análisis que se pretende realizar.

3. Observadores.

La norma establece un mínimo de 15 observadores sobre los que se debe realizar el test. Estos, además, no podrán ser expertos, es decir, no deben estar habituados a trabajar con calidad de imagen en su entorno de trabajo ni estar familiarizados, en general, con tratamiento digital de imágenes. En el laboratorio se realizó el test a 18 sujetos con edades comprendidas entre los 20 y los 55 años.

4. Instrucciones sobre la realización del test. Estructura de la sesión.

Se debe proporcionar a los observadores detalles sobre la escala de medida, la secuencia y el tipo de imágenes de referencia e imágenes distorsionadas utilizadas. Para ello, previamente a la realización del test, se explica en detalle a los observadores el procedimiento que va a seguirse y se les muestran algunas imágenes originales y distorsionadas del mismo tipo que las que deben evaluar. Además, se dispone de una secuencia de demostración o entrenamiento para que

los sujetos se familiaricen con el proceso de medida y los tiempos empleados. Los resultados de esta secuencia no se computan y al finalizar la misma se responderá a posibles preguntas que surjan por parte de los observadores. A continuación, se procederá a realizar la sesión de evaluación. La norma establece un máximo de duración en torno a los treinta minutos. La sesión consta de las secuencias que muestra la figura 5.1. Una primera parte se dedica a secuencias de estabilización que permiten que la opinión del observador se centre y cuyos resultados no deben ser considerados. La segunda parte es la principal y de la que se obtendrán los resultados finales. En la figura también se muestra en detalle el diagrama temporal para una secuencia. Los tiempos de observación de cada imagen junto con los tiempos entre imágenes y el de evaluación dan un total de 54 segundos para cada secuencia. Puesto que disponemos de un conjunto de 30 pares de imágenes, la duración final del test será de 27 minutos, que se encuentra dentro de lo estipulado por la recomendación.

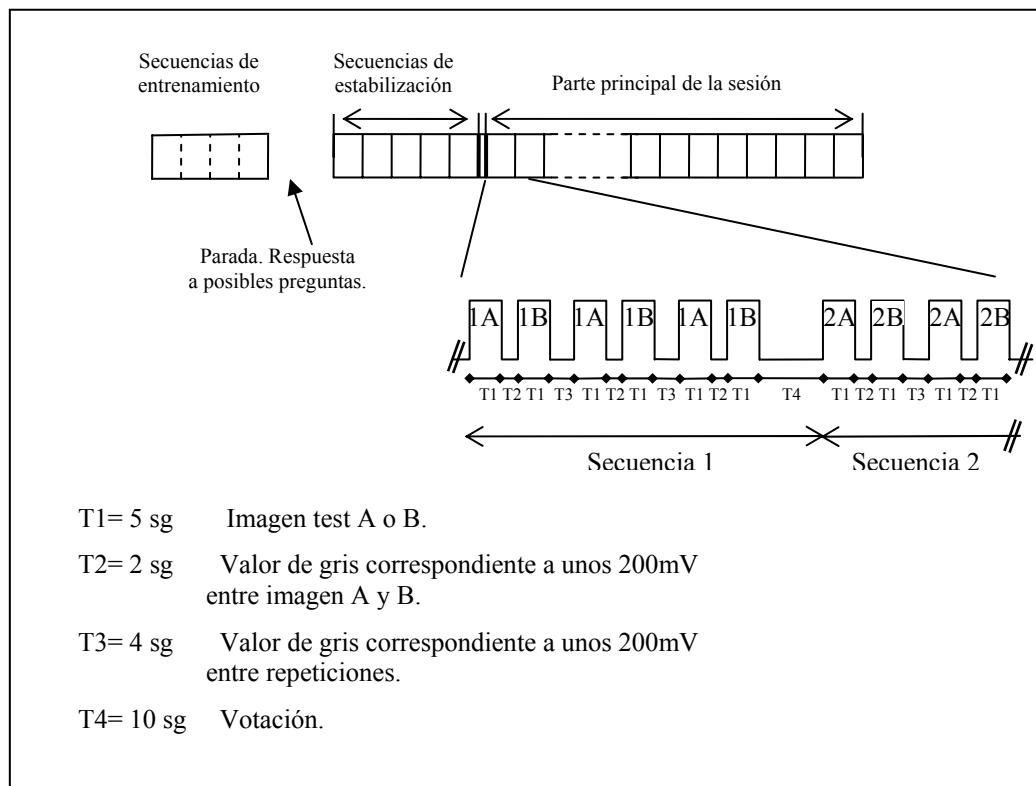


Figura 5.1: Estructura de la sesión test.

El orden de presentación debe ser aleatorio, es decir, en cada sesión el orden en el que aparecen las secuencias es diferente. Con ello se consigue que el efecto que podría tener sobre la valoración un posible aprendizaje del observador, así como el hecho de que el observador se encuentre más distraído o cansado al final de la sesión se minimicen, al no afectar siempre a las mismas secuencias.

5. Escala de puntuación.

Se requiere que el observador evalúe las imágenes A y B de la presentación sin saber cual de ellas corresponde a la imagen de referencia y cual a la imagen con error. Esto se consigue haciendo que durante el test la posición de la imagen de referencia vaya cambiando de forma pseudo-aleatoria. Así, en algunos casos, la imagen test A será la original y en otros sucederá al contrario, de forma que el observador no sabrá a priori si está viendo la imagen original o no.

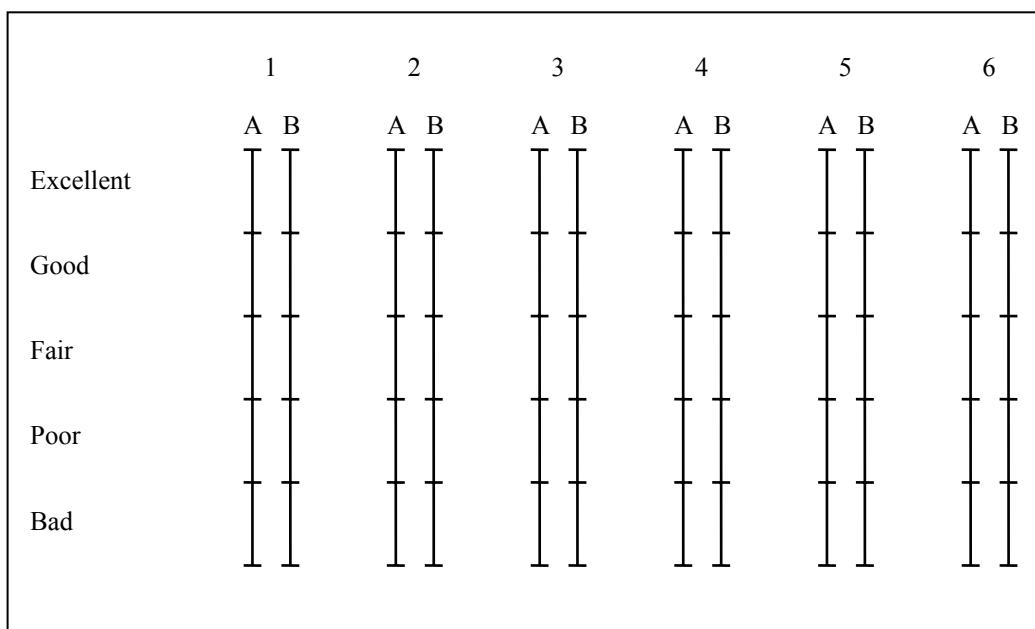


Figura 5.2: Escala de medida de DSCQS.

La figura 5.2 muestra una sección de las hojas utilizadas para la adquisición de los resultados. La escala continua va impresa en negro y la marca introducida por los observadores, de acuerdo a la calidad apreciada por ellos, en rojo. Con esto se consigue evitar posibles errores de confusión entre una marca de separación y una de evaluación. Los valores obtenidos mediante estas escalas se normalizan en valores enteros en un rango de 0 a 100. Es importante hacer comprender a los observadores, antes de realizar la sesión de test, que las

etiquetas que aparecen a la izquierda son meramente informativas. Dan una idea de la calidad correspondiente a un intervalo determinado sin que ello signifique que la escala proporciona cinco valores discretos. Se debe hacer por tanto hincapié en el hecho de que utilicen según su criterio todos los valores de los que dispone la escala.

6. Análisis e interpretación de los resultados subjetivos.

El conjunto de datos obtenidos para los distintos observadores que realizaron el test se muestran en la tabla A.1 del Anexo A. Para poder analizar estos resultados es necesario combinarlos mediante fórmulas estadísticas que proporcionen resultados gráficos y/o numéricos. En el Anexo 2 de la norma ITU-R BT.500-10 se muestran los pasos a seguir para el cálculo de la media y el intervalo de confianza. Al ejecutar estos cálculos sobre los valores de la tabla A.1 se obtienen los resultados MOS presentados en las figuras 5.3 a 5.7.

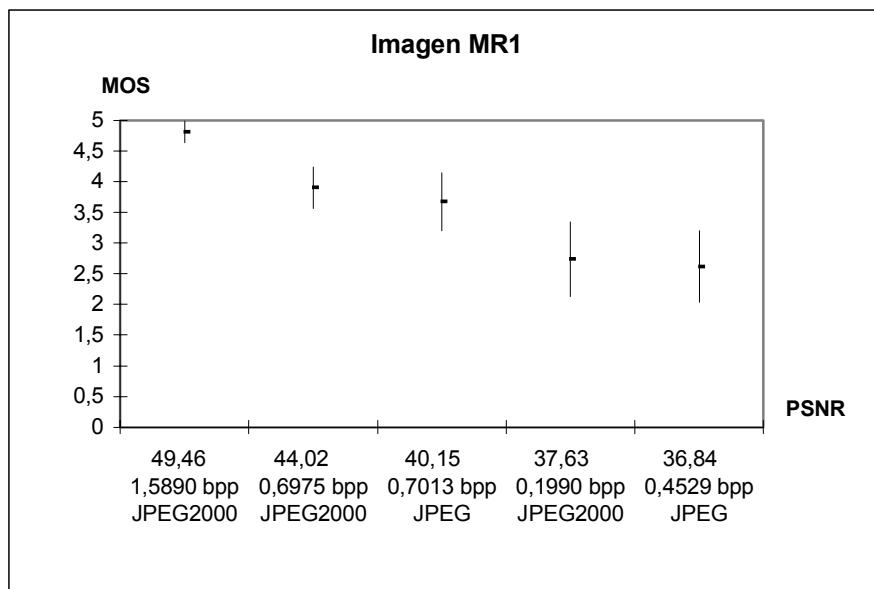


Figura 5.3: Resultados MOS e intervalo de confianza para MR1.

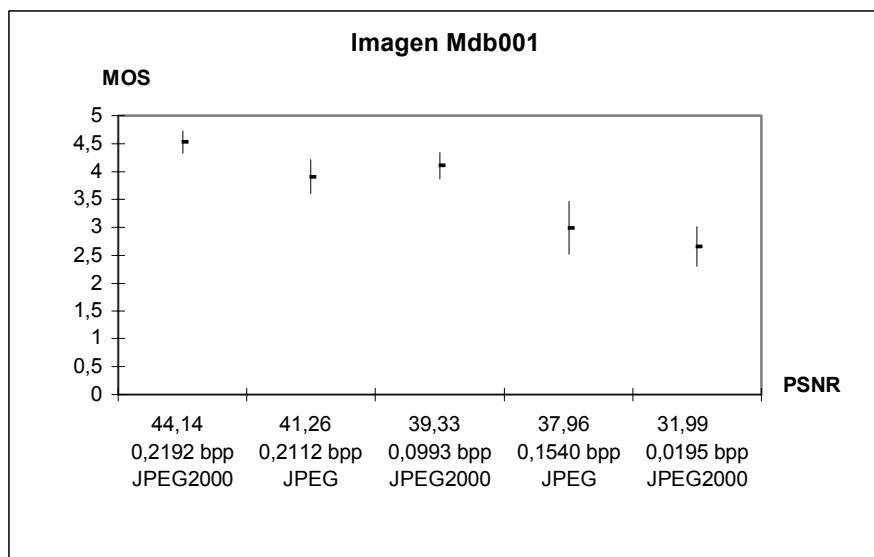


Figura 5.4: Resultados MOS e intervalo de confianza para Mdb001.

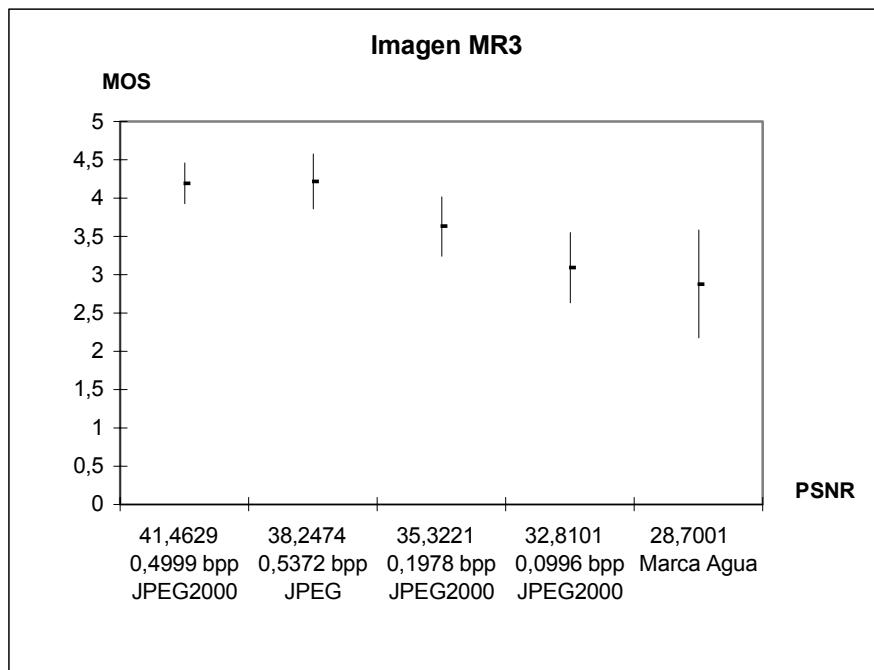


Figura 5.5: Resultados MOS e intervalo de confianza para MR3.

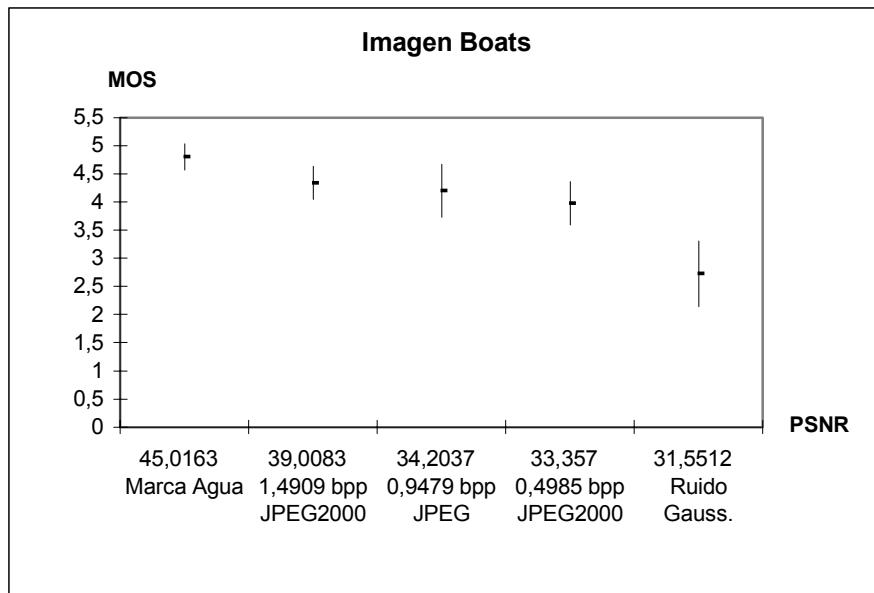


Figura 5.6: Resultados MOS e intervalo de confianza para Boats.

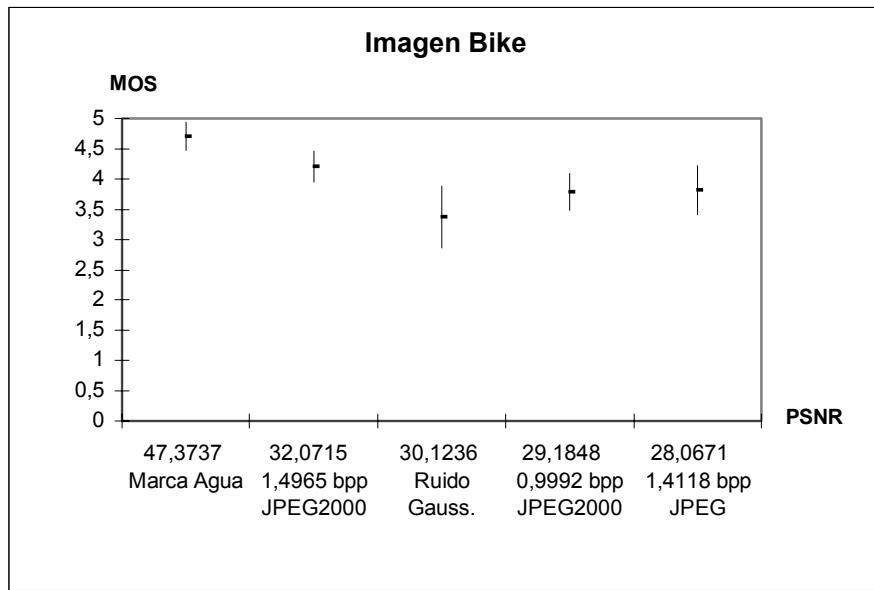


Figura 5.7: Resultados MOS e intervalo de confianza para Bike.

Como puede observarse los valores de PSNR van desde 28 a 50 dB aproximadamente para las distintas imágenes tratadas. La calidad percibida por los observadores abarca un rango de 2,6 a 4,8. Se debe recordar que el margen que se utiliza para comparar estos resultados con los índices obtenidos mediante la utilización de las métricas de calidad objetivas es [1..5], donde 5 se corresponde con la calidad máxima.

Un primer análisis de estos resultados puede llevar a la conclusión de que el rango de PSNR propuesto no es el más adecuado puesto que con él no se ha cubierto el rango subjetivo propuesto. Sin embargo, se debe tener también en consideración que estos valores de calidad han sido dados por no expertos. En el caso de las imágenes *Boats* y *Bike* esta imposición de la norma resulta lógica. En cambio, para imágenes MRI que están destinadas a visualización de expertos, probablemente, si se realizaran estos mismos test sobre sujetos que cuanto menos, hayan visto en alguna ocasión este tipo de imágenes se tendrían resultados más críticos. Se quiere con esto resaltar el hecho, ya comentado, de que los test subjetivos no son la panacea y hay que evaluar los resultados con cierta prudencia. Otro hecho que se ha observado en las diferentes sesiones realizadas es la tendencia que tienen los observadores a puntuar al alza y en general, sin utilizar valores extremos. Es decir, son reacios a la hora de evaluar una imagen con la peor puntuación. Una imagen que sea similar a la original la puntúan con valor cercano al máximo de calidad, sin embargo, una imagen con distorsión perceptible suelen situarla en la mitad del margen. Finalmente, destacar que la perfecta correlación con el SVH no puede lograrse debido a las variaciones naturales en la evaluación de la calidad subjetiva.

7. Relación entre los datos MOS y las medidas objetivas.

La norma ITU-R BT.500-10 proporciona en el Anexo 3 los pasos a seguir para hallar la relación entre medidas de calidad objetivas de la distorsión de la imagen y los datos subjetivos aportados por los observadores que han efectuado el test. En general la relación entre ambos es no lineal y en concreto la norma propone hacer el ajuste mediante funciones logísticas. En el análisis de todos los resultados que se muestran a continuación para los distintos modelos de medida de calidad propuestos se va a buscar esta relación. Así, mediante estas

regresiones no lineales se obtendrá un conjunto de valores (en función de las medidas objetivas) que son los que se comparan directamente con los MOS en términos de la correlación de Pearson.

Puesto que todos los métodos se van a evaluar de esta forma resulta conveniente mostrar los pasos seguidos y las ecuaciones matemáticas utilizadas. Se van a denominar con x los valores de distorsión de calidad medidos de forma objetiva, con s a los valores subjetivos MOS y con s' a los que se conseguirán al realizar el ajuste.

La función logística elegida tendrá la forma: $s' = f(x) = \frac{1}{1 + e^{(x-A)B}}$,

donde A y B son constantes a determinar. La relación anterior puede también ser expresada mediante una función lineal tal que:

$$y = (x - A) \cdot B, \quad \text{donde } y = \ln\left(\frac{1}{s'} - 1\right).$$

Así, el cálculo de A y B se puede hacer con una regresión lineal simple y con los valores A, B y x se obtienen los s' de forma inmediata. Para evaluar los distintos modelos de medida de calidad se calculará el coeficiente de correlación de Pearson entre los datos s y s' . Este coeficiente se va a obtener calculando la covarianza entre ambas variables, dividida por el producto de cada una de las desviaciones típicas de las variables, resultando así una medida de asociación adimensional. Una vez obtenido este coeficiente es necesario comprobar que es significativo estadísticamente, es decir, es preciso realizar un contraste de hipótesis. Para ello se plantean las siguientes hipótesis: $H_0 : \rho = 0$ y $H_1 : \rho \neq 0$

y se tendría que formar el estadístico de contraste: $t = r / \sqrt{\frac{1 - r^2}{n - 2}}$,

que sigue una distribución t-Student con $n-2$ grados de libertad. Valores elevados de t provienen de grandes diferencias entre la media observada y el valor que se contrasta, compatibles con la hipótesis alternativa. Para saber si el valor que se ha obtenido para t es suficientemente elevado se calcula la probabilidad de que bajo H_0 se presente un valor igual o superior al t obtenido, que viene dado por el área a la izquierda de menos el valor absoluto de t (caso de contrastes bilaterales). Este valor se conoce como p-valor. Para que t sea

considerado elevado su p-valor asociado tiene que ser pequeño. Se consideran valores pequeños para el p-valor aquellos inferiores a 0.05. Por tanto, en el análisis que se haga de los resultados se va a indicar, además del valor ρ de correlación su valor p asociado.

5.2. Sistema de medida de calidad con CSF como filtro.

Se va a evaluar el comportamiento que presenta un sistema de medida de calidad perceptible como el que muestra la figura 5.8. El sistema se compone de todos los elementos que ya se han descrito previamente en el Capítulo 4: preprocesado, adaptación de luminancia, CSF, etc. Además, se puede ver la correspondencia entre los distintos módulos y la nomenclatura asociada que se ha venido utilizando. La figura da una visión global de lo que es el modelo de calidad perceptible utilizando la CSF como un filtro. Se muestra como se relacionan y organizan dentro del modelo los distintos componentes del SVH. Es decir, en la figura 5.8 en realidad lo que se representa es el esqueleto a partir del cual se va a formar el modelo de calidad perceptible. Luego, dependiendo de la elección que se haga para los distintos componentes se van a implementar, sobre esta misma estructura, distintos modelos de calidad. En la interfaz gráfica detallada en el Anexo B este modelo se construye al pulsar el botón “ICP (CSF como filtro)” a partir de las opciones de CSF, filtros, contrastes, etc. que el usuario ha hecho previamente. En general, se podrían llegar a construir un total de $4 \times 4 \times 3 \times 4 = 192$ variaciones distintas combinando las implementaciones realizadas para los distintos componentes que se resumen en:

- (4) Adaptación de luminancia: ninguna, raíz cúbica, logaritmo y adaptación de Daly.
- (4) CSF: ninguna, modelo de Rust, modelo de Daly y modelo de Barten.
- (3) Banco de filtros: filtros Anillo, Cortex y Gabores.
- (4) Contrastos: modelos Local, Global, Peli y Lubin.

De hecho, en realidad, se podrían llegar a construir muchos más ya que para los filtros de Gabor es posible elegir el número de escalas de frecuencia y orientación con lo que, aunque se ha considerado como uno, se podrían llegar a formar n bancos con filtros de Gabor distintos.

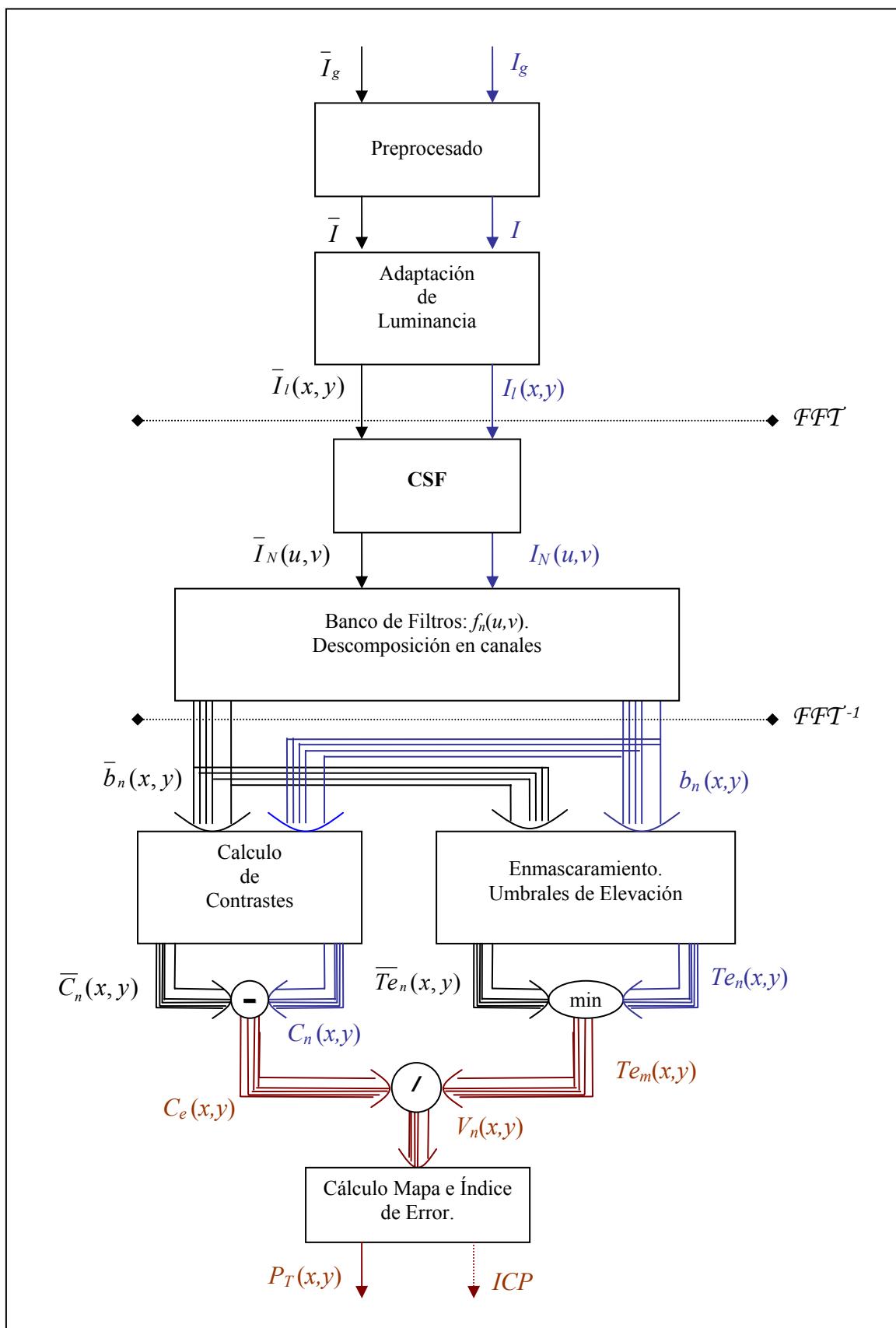


Figura 5.8: Sistema de medida de calidad perceptible con CSF como filtro.

Para el análisis de los resultados que se han obtenido mediante la implementación de distintas variantes sobre este modelo hay que tener presente que dado el gran abanico de posibilidades que se plantean, se va a conseguir un rango bastante amplio de valores de correlación. Esto dará una idea del peso de un componente visual en concreto dentro del modelo y para cada componente cuales son los modelos desarrollados que mejores ajustes con la calidad percibida proporcionan. Antes de empezar con esta comparativa es necesario explicar una serie de aspectos sobre el esquema general y algunas decisiones adoptadas en la implementación.

Selección de variantes y parámetros para el modelo de calidad perceptible.

En primer lugar, es necesario precisar de todo el conjunto de variantes potenciales cuales son las que se han utilizado para obtener los resultados. De las 192 mencionadas, en realidad, sólo se van a analizar 72 que surgen al excluir de las posibles combinaciones los siguientes modelos:

- Se ha eliminado el modelo con adaptación de luminancia mediante la función logaritmo ya que como se ha comentado (ver Capítulo 4) sobreestima la sensibilidad en algunos casos. Puesto que no aportaría gran información se ha optado por descartarla para no sobrecargar la exposición de resultados con información que no es relevante.
- Se excluyen también los modelos del contraste de Peli y Lubin ya que no se pueden adaptar al esquema de métrica perceptible con CSF como filtro como se explicará más adelante.

Por tanto, esto deja un total de $3 \times 4 \times 3 \times 2 = 72$ variantes con las que realizar las medidas del ICP sobre el conjunto de imágenes bajo estudio.

En segundo lugar, es preciso concretar algunas de las decisiones que se han tomado a la hora de implementar el modelo visual dentro de este sistema de calidad perceptible. Se van a enumerar los distintos componentes destacando en cada caso concreto las decisiones que se han tomado y que no se han explicado en detalle en el Capítulo 4 o bien porque han surgido directamente de la realización de las pruebas o bien porque son el resultado de considerar los componentes en conjunto y no de manera individual.

1. Adaptación de luminancia. Aquí merece especial atención el tratamiento que se les da a las imágenes médicas. Al incluir la adaptación dentro del modelo perceptual, el objetivo es conseguir que en las zonas en las que la luminancia local sea elevada, un mayor nivel de distorsión sea tolerado. Cualquiera de las funciones implementadas lo que hacen es someter a todos los píxeles de la imagen a una función no lineal que visualmente se vería como un aclarado de la imagen. Este “enmascaramiento de luminancia” tiene sentido implementarlo en las regiones de la imagen en las que exista señal que pueda enmascarar. Para el caso de imágenes comunes de entornos de compresión (ver Anexo A, figuras A4 y A5) u otras que son más o menos representaciones en 2D del mundo real, todo lo que la imagen muestra puede ser considerado como señal, incluyendo lo que se clasifica como fondo de imagen. En el caso de las imágenes médicas radiológicas esto no sucede así, sino que siempre se va a disponer de una estructura (un hueso, rodilla, etc.) claramente diferenciada del fondo y además, éste no aporta ninguna información para la evaluación de la calidad o para el diagnóstico. Si se implementase la adaptación en imágenes médicas para todos los puntos de la imagen, por una parte, en la zona de la estructura se produciría el efecto deseado de enmascaramiento de distorsión y por otra, en el fondo estaríamos provocando el efecto contrario considerando la distorsión más visible de lo que en realidad es y subestimando la calidad percibida. Para evitar este efecto en el caso de las imágenes MR1, Mbd001 y MR3 (ver Anexo A, figuras A1, A2 y A3) la adaptación de luminancia se va a efectuar sólo en las zonas de señal (sobre la estructura de la imagen y no en el fondo) que se identifican mediante una máscara. Por tanto, para las tres imágenes mencionadas, antes de la adaptación se calcula esta máscara (ver Anexo B, figura B.6). Así, todos los resultados de las imágenes médicas se han obtenido teniendo en cuenta este procedimiento al implementar la adaptación.

2. Función de sensibilidad al contraste. En cuanto a la CSF y su implementación dentro del modelo como un filtro, ya se ha explicado en el Capítulo 4. Aquí, simplemente se resaltan algunas de las implicaciones que tiene este uso de la CSF para que se tengan presentes, ya que éste es uno de los componentes de mayor importancia dentro del modelo. De hecho, la CSF condiciona el esquema entero de la métrica. El utilizarla como filtro tiene una serie de consecuencias a la hora de efectuar el cálculo de los contrastes y el de los umbrales de enmascaramiento. Para el cálculo de los contrastes hay que tener presente el problema que plantea la medida de la CSF en frecuencia cero

(ver Capítulo 4) y que afecta al contraste puesto que éste se define dividiendo por el canal paso bajo. Al haber filtrado por la CSF, los valores que ésta tenga para bajas frecuencias van a determinar los contrastes obtenidos. En el caso de los modelos analíticos de Daly y Rust que presentan cero a frecuencia cero es necesario conservar la banda base. En cuanto al cálculo de los umbrales de enmascaramiento, se recuerda que éstos son los obtenidos mediante la ecuaciones 4.17 y 4.18. Finalmente, las funciones de sensibilidad que se han empleado para la obtención de resultados se muestran en la figura 5.9.

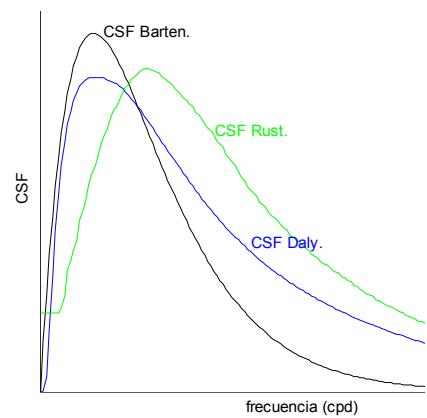


Figura 5.9: CSF's utilizadas.

3. Banco de filtros. Las pruebas se han realizado para los tres bancos de filtros descritos. En el caso de los filtros Cortex y Anillo no se ofrece ninguna posible variación, pero para los Gabores es posible elegir el número de escalas y orientaciones con los que se trabaja. Los resultados, en el caso concreto de los filtros Gabor, se han obtenido utilizando $N_o = 6$ (número de orientaciones) y $N_f = 5$ (número de escalas) más el paso bajo. Se ha realizado esta selección para poderlo comparar en términos de igualdad con el banco de filtros Cortex, que tiene estos mismo canales.

4. Cálculo de contrastes. Todas las definiciones propuestas coinciden al expresar el contraste como un cociente adimensional entre un valor local y un valor medio de luminancia. Sin embargo, tanto el contraste de Peli como el de Lubin calculan el valor del denominador (la luminancia media) como la suma de canales de frecuencia que están por debajo del canal para el que se está calculando el contraste (ver tabla 4.1). En principio, aunque la definición dada es del todo válida, en la aplicación concreta dentro de este modelo no tiene sentido ya que las bandas con las que se calculan los contrastes han sido normalizadas previamente por la CSF, es decir, si hay dos componentes con la misma amplitud, éstos van a tener la misma visibilidad. Por tanto, si al hacer la división se incluyen varias bandas en el denominador se estará distorsionando esta normalización. Por consiguiente, en la realización práctica de las pruebas para la

obtención de resultados se han omitido estas dos definiciones de contraste puesto que no tienen sentido dentro de este modelo.

Análisis de resultados del modelo.

Una vez realizadas estas consideraciones sobre la puesta en práctica del modelo sólo queda exponer los resultados obtenidos mediante las 72 variantes del sistema de medida mencionadas. Para ello, nuevamente, se va a desglosar el análisis por medio de la efectividad de los componentes dentro del modelo. Destacaremos las conclusiones generales obtenidas y mostraremos gráficamente aquellas variantes con mejor comportamiento.

El modelo de **adaptación de luminancia** para el que se obtienen mejores resultados en general es la raíz cúbica. Aunque, en realidad, los resultados obtenidos son prácticamente los mismos que con el modelo de Daly ya que las funciones son muy similares para el rango de luminancias con el que se trabaja. Esta pequeña mejoría se debe fundamentalmente a que la función de Daly está algo por encima de la raíz cúbica (ver figura 4.4) y eleva un poco más el nivel de luminancia (con lo que los errores se percibirían menos) lo que se traduce en valores algo mayores de calidad percibida que los que proporciona la raíz cúbica. Es decir, con el modelo de Daly se está sobreestimando el valor de calidad percibida. En cualquier caso hay que resaltar el hecho de que esta diferencia entre ambos modelos es mínima y que proporcionan valores de correlación prácticamente iguales para el resto de componentes fijos.

Especial atención merece el caso de la no implementación de un modelo de adaptación de luminancia. En general, el no implementarla empeora los resultados ligeramente, pero hay algunas salvedades. En el caso particular de las imágenes médicas evaluadas en las que la calidad percibida es baja, al implementar un modelo de adaptación el ICP va a estar por encima del valor MOS. La explicación de este fenómeno es la misma que para el caso de Daly y la raíz cúbica, es decir, al implementar el modelo de adaptación los errores en la estructura de la imagen médica se perciben menos, pero este salto cuantitativo no es suficientemente sensible ya que cuando los errores son perceptibles y están por encima del umbral de visibilidad aunque la luminancia local sea elevada estas distorsiones no están ocultas y al realizar la adaptación se va a elevar aún más este nivel de luminancia local ocultando errores que

en realidad son visibles. Por tanto, en algunos casos, no implementar la adaptación puede resultar provechoso.

En cuanto a los posibles modelos de la **función de sensibilidad al contraste**, el que mejores resultados proporciona sobre el modelo de calidad perceptible con CSF como filtro es el de Daly, aunque con algunas puntuaciones ya que la CSF de Barten da también resultados muy similares, quedándose la de Rust por debajo de estas dos. Esto se debe a que la función de Rust es mucho más suave. Es la que más conserva las altas y bajas frecuencias y tiene el pico de sensibilidad algo más desplazado a medias frecuencias (ver figura 5.9). En resumen, la CSF del modelo de Rust es la de mayor sensibilidad, lo que se traduce en que el error va a ser visible con un valor de contraste menor. Proporciona, por tanto, resultados ICP menores: subestima la calidad. Un ejemplo de esto se encuentra en la figura 5.10 donde se han presentado los mapas de visibilidad obtenidos con CSF de Rust y Daly. El valor MOS dado por los observadores para la imagen mostrada es de 4,7 y los valores ICP obtenidos con la CSF de Daly y Rust son respectivamente 4,52 y 4,39. Los mapas de visibilidad que muestra la figura han sido modificados para poder llegar a ver algo en la representación, ya que la calidad de la imagen es muy buena y el mapa real en ambos casos es prácticamente uniforme en su totalidad. Sobre ambos mapas se ha establecido un umbral del 40% de forma que todos los píxeles por encima de este valor se muestran en blanco. El mapa obtenido con la CSF de Daly sigue siendo prácticamente uniforme, mientras que en el de Rust sí se perciben más los errores al provocar esta saturación en la imagen.



Figura 5.10: Mapas de visibilidad de error para imagen Bike con marcado de agua PNSR = 47,37. De izquierda a derecha: imagen distorsionada, $V_n(x,y)$ con CSF Daly y $V_n(x,y)$ con CSF Rust (ambos mapas modificados para visualización y con mejora de contraste).

En el caso de los modelos de Barten y Daly el comportamiento similar se debe básicamente a que la CSF de Barten es más aguda. Es decir, es menos sensible para altas frecuencias, mientras que en el resto, que es donde realmente el SVH es sensible, está por encima de la de Daly. Sin embargo, el principal problema del modelo de Barten es que es una función isótropa y no está teniendo en cuenta la menor sensibilidad en frecuencias oblicuas, mientras que la de CSF de Daly sí que lo tiene en cuenta, con lo que los resultados obtenidos mediante ambas funciones al final resultan similares.

Por último, en el caso de no implementar ningún modelo de CSF los valores de ICP obtenidos van a estar por encima de los MOS, es decir, los resultados de correlación son mucho peores. Este comportamiento es del todo lógico, ya que la CSF es la pieza fundamental del modelo. El no incluir el filtrado por la CSF implica que no se tiene ninguna indicación de la sensibilidad del SVH y por tanto, de si el error es perceptible o no.

De los tres **bancos de filtros** empleados, los Cortex y los Gabores son los que proporcionan los resultados ICP que se corresponden más con los datos MOS. Sin embargo, la diferencia respecto a los filtros anillo no es muy elevada y estos últimos presentan como ventaja que el cálculo del ICP es mucho más rápido al haber reducido la dimensionalidad del problema. Estos resultados también parecen bastante lógicos ya que la descomposición en frecuencia y orientación no se explota convenientemente. Es decir, no se está implementando un enmascaramiento entre canales. Por tanto, la diferencia entre un banco de filtros con descomposición en orientaciones de uno que no lo tenga está en que el cálculo de contrastes y umbrales es más acotado, más sensible a las diferencias y más acorde con el funcionamiento del SVH. Si esta propiedad no se adapta con todas sus posibilidades al modelo, al final, los valores ICP calculados no van a ser muy diferentes.

Finalmente, para el cálculo de los **contrastos**, se comprueba que los resultados que ofrece el contraste Global están muy por encima de los conseguidos con el contraste Local. Esto en parte se debe a la colocación del módulo de contraste dentro del sistema ya que cuando llega a realizarse este cálculo ya se ha realizado la adaptación y se ha filtrado por la CSF con lo que la información original se ha perdido en parte y no puede modelarse, como se pretende, que el ojo se puede adaptar a pequeñas áreas en la imagen [7]. Un estudio más a fondo de los distintos modelos de contraste se verá en el siguiente

apartado, ya que para la métrica con CSF por pesos, las bandas utilizadas para el cálculo de contrastes no han sido modificadas por la CSF y puede estudiarse mejor la influencia real que tienen.

En conclusión, las variantes que muestran unos resultados más acordes con los percibidos por el observador son aquellas basadas en el modelo de adaptación de luminancia raíz cúbica, la CSF de Daly, los filtros Cortex o Gabores y un cálculo de contraste Global. Estos resultados se muestran en la tabla 5.1, donde se indican los valores de correlación obtenidos y sus valores p asociados, que según lo explicado en el apartado 5.1 de este capítulo van a indicar si las variables son significativas estadísticamente. En la figura 5.11 se muestran gráficamente los resultados obtenidos para estas variantes. Los valores ICP se corresponden con los símbolos \square y las curvas son los ajustes logísticos conseguidos según la recomendación de la norma ITU-R BT.500-10.

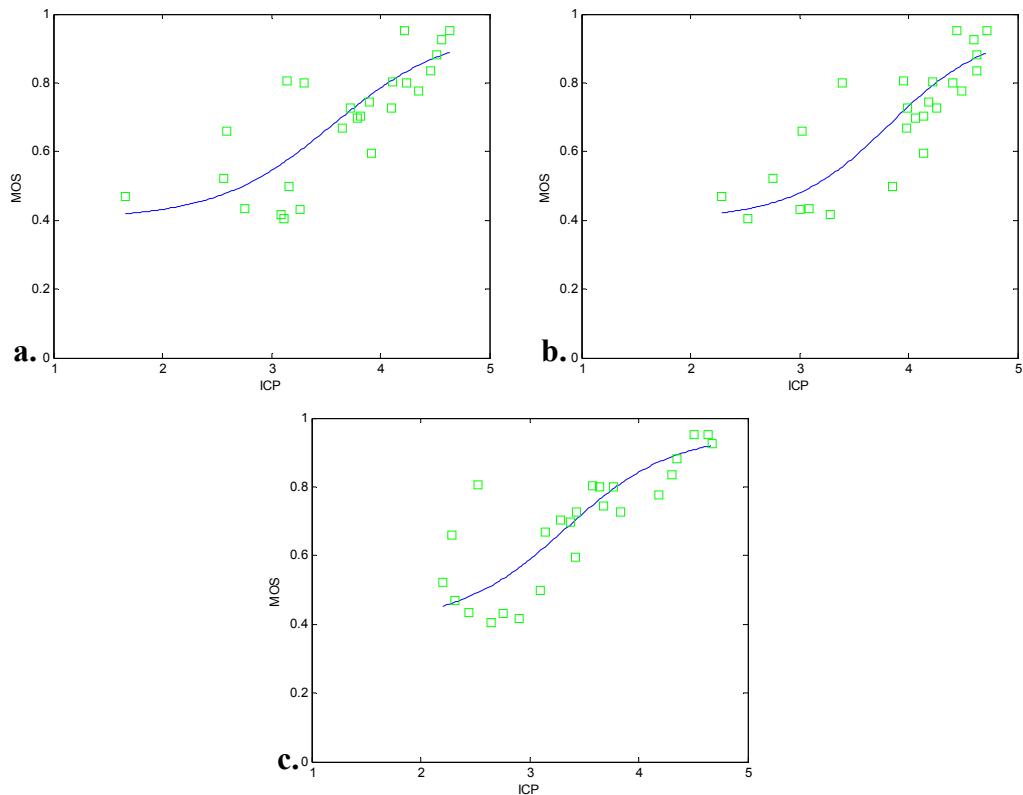


Figura 5.11: Sistema de calidad con CSF como filtro, valores ICP vs. MOS. **a.** Para el modelo con adaptación de luminancia raíz cúbica, CSF de Barten, filtros Anillo y contraste Global. **b.** Para el modelo con adaptación de luminancia raíz cúbica, CSF de Daly, filtros Gabor y contraste Global. **c.** Igual que b. con filtros Cortex.

Variante {	Adaptación CSF Banco filtros Contraste.	Raíz cúbica Modelo Barten Anillo Global	Raíz cúbica Modelo Daly Gabores Global	Raíz cúbica Modelo Daly Cortex Global
Correlación (ρ)	0,7855	0,8276	0,8223	
Valor p asociado	0,00003	0,000003	0,000004	

Tabla 5.1: Valores de correlación para las variantes significativas.

En general, los valores obtenidos de correlación superan los proporcionados por la PSNR ($\rho = 0,6$). Además, la ventaja añadida de este modelo es que aporta mapas de visibilidad del error, con lo que se puede comprobar de forma visual la correspondencia entre lo que se percibe y lo que proporciona la métrica. No se debe olvidar que las condiciones de visibilidad juegan un papel muy importante y que para que tenga efecto esta comprobación visual los parámetros introducidos en la etapa de preprocesado deben ser los que realmente existen en la práctica (valor gamma del monitor, distancia de visualización, luminancia máxima, etc.). Las figuras 5.12 a 5.15 muestran algunos de los mapas de visibilidad obtenidos así como los valores ICP asociados.

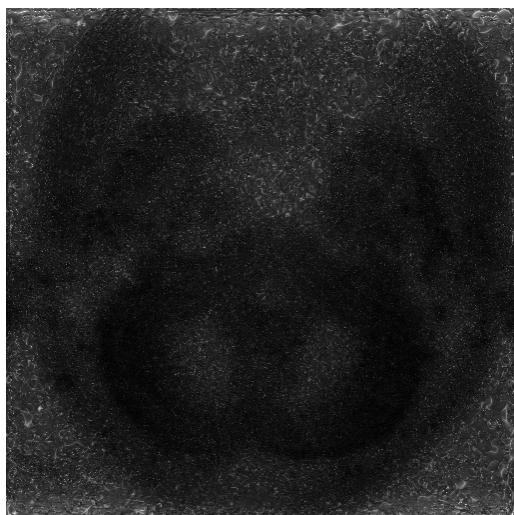


Figura 5.12: Mapa de visibilidad (con realce de contraste) para:

- Imagen original: MR1.
- Imagen distorsionada: JPEG2000 0,69bpp.
- MOS: 3,90.
- PSNR: 44,02 dB.
- ICP: 3,6.

Figura 5.13: Mapa de visibilidad (con realce de contraste) para:

- Imagen original: Mbd001.
- Imagen distorsionada: JPEG2000 0,22bpp.
- MOS: 4,52.
- PSNR: 44,14 dB.
- ICP: 4,3.

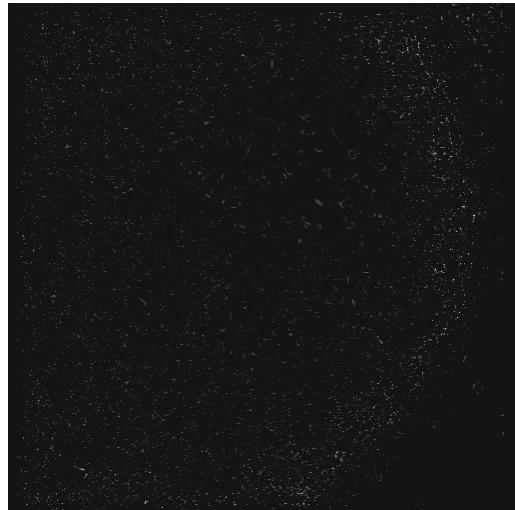


Figura 5.14: Mapa de visibilidad (con realce de contraste) para:

- Imagen original: MR3.
- Imagen distorsionada: JPEG 0,53bpp.
- MOS: 4,21.
- PSNR: 38,24 dB.
- ICP: 3,8.

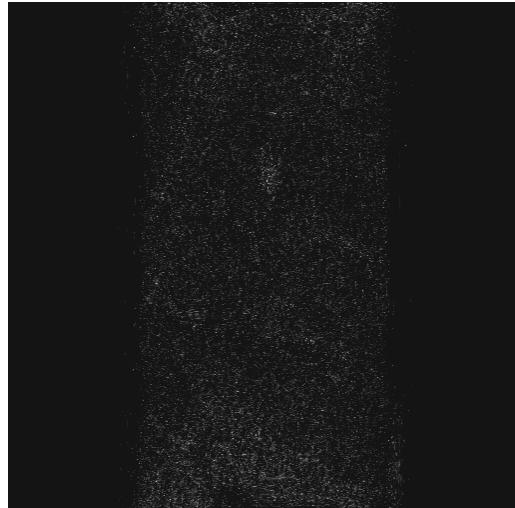


Figura 5.15: Mapa de visibilidad (con realce de contraste) para:

- Imagen original: Boats.
- Imagen distorsionada: JPEG2000 0,5bpp.
- MOS: 3,97.
- PSNR: 33,35 dB.
- ICP: 3,89.



Las imágenes originales se muestran en el Anexo A. No se han incluido las imágenes distorsionadas ya que en general son imágenes de alta calidad y al mostrarlas o imprimirlas en un tamaño menor y sin las condiciones de visualización apropiadas, no aportarían casi información. Además, todos los mapas de visibilidad mostrados han sido modificados. Es necesario realizar este realce ya que sino, el lector no percibiría nada. En la imagen del cerebro (figura 5.12), la distorsión es perceptible sobre el ruido de fondo de la imagen (esquinas y parte superior central), que se ha suavizado. Para el caso de la mamografía y la rodilla (figura 5.13 y 5.14), la distorsión es apenas perceptible en los bordes y son las que mayor puntuación MOS tienen. Finalmente, para la figura 5.15, la distorsión se percibe sobre todo en la parte inferior de los barcos y alrededor de los mástiles.

5.3. Sistema de medida de calidad con CSF por pesos.

En este apartado lo que se pretende es evaluar el comportamiento que presenta un sistema de medida de calidad perceptible como el que muestra la figura 5.16. en el que la CSF se va a implementar como una serie de pesos a la hora de calcular los umbrales de enmascaramiento. La figura, por tanto, presenta el esqueleto del sistema sobre el que se implementarán los distintos modelos de los componentes del SVH. Como en el caso anterior (modelo con CSF como filtro) el sistema se compone de todos los elementos descritos en el Capítulo 4. Se ha seguido la nomenclatura allí empleada al incluirlos en la figura 5.16. Para la ejecución de este modelo dentro del entorno gráfico construido y que se detalla en el Anexo B, bastaría con hacer las selecciones de los componentes que se deseen y pulsar el botón “ICP (CSF por pesos)”.

En el análisis del comportamiento de este sistema de calidad se van a seguir los mismos pasos que en el apartado anterior. Por consiguiente, el primer paso consiste en exponer cuales han sido los modelos del SVH sobre los que se evalúa el sistema para, posteriormente, sacar las pertinentes conclusiones del análisis de los resultados.

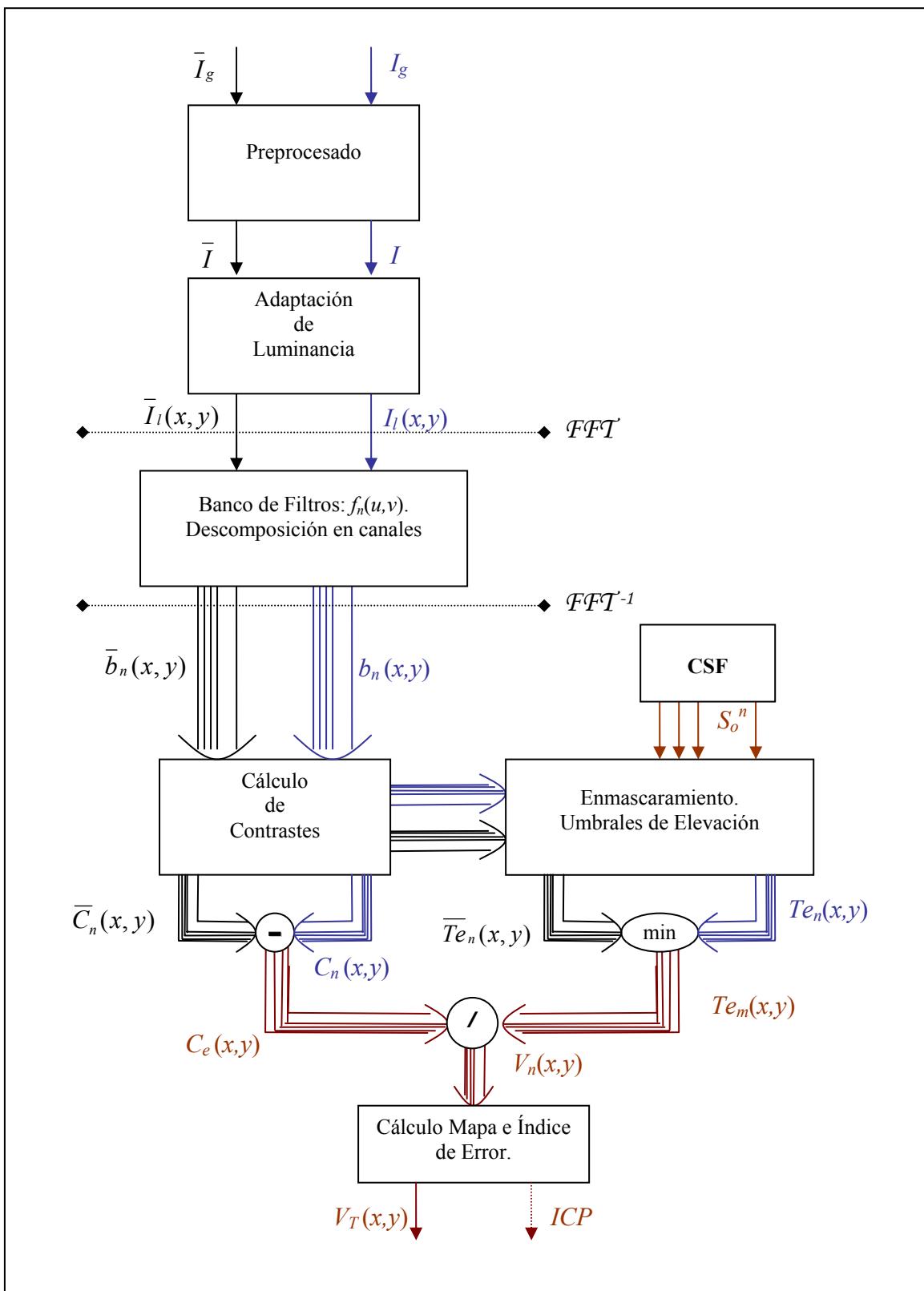


Figura 5.16: Sistema de calidad perceptible con CSF por pesos.

Selección de variantes y parámetros para el modelo de calidad perceptible.

En primer lugar, se recuerda que una variante no es más que una determinada combinación de modelos por cada componente del SVH (adaptación, CSF, filtros y contrastes) que se enlazan siguiendo las pautas que traza el sistema de calidad perceptible. Al igual que en el caso del modelo con CSF como filtro, aquí se ha descartado la adaptación de luminancia implementada con logaritmo por los motivos ya comentados. En principio, no es necesario hacer ninguna otra restricción por lo que para este caso tendríamos un total de $3 \times 4 \times 3 \times 4 = 144$ variantes posibles. La evaluación de resultados indicará cuales de éstos son adecuadas en la práctica.

En segundo lugar, en cuanto a las decisiones tomadas sobre parámetros para la implementación de los distintos componentes, se seguirán las indicaciones realizadas en el apartado anterior y que se resumen en:

1. Adaptación de luminancia para las imágenes médicas mediante máscaras.
2. Las funciones de sensibilidad al contraste implementadas son las que se muestran en la figura 5.9. De nuevo, se recuerda que la CSF es la que define por completo la estructura de la métrica, influyendo en este caso, en el cálculo de los umbrales de enmascaramiento, que se debe efectuar según las ecuaciones 4.18 y 4.19.
3. Banco de filtros Gabor con $N_o = 6$ (número de orientaciones) y $N_f = 5$ (número de escalas) más el paso bajo.

Análisis de resultados del modelo.

Se va a seguir el mismo esquema que en el apartado anterior, estudiando la influencia de los componentes dentro de este sistema de calidad por separado e indicando en cada caso los que obtienen resultados más acordes con la calidad percibida por el observador. Así, puesto que los modelos de los componentes son siempre los mismos, en muchos casos el efecto o la relevancia que tienen en la métrica va a coincidir con lo explicado en el apartado anterior.

En la implementación de la **adaptación de luminancia** sucede precisamente lo mismo que en el caso de la CSF como filtro. Puesto que las curvas son las mismas y además su colocación dentro del esquema de la métrica es también similar, resulta lógico que el efecto sea equivalente. De hecho, el módulo de adaptación de luminancia se sitúa en ambos casos en la fase inicial del procesamiento y hasta ese punto ambos

esquemas (figuras 5.8 y 5.16) coinciden. Por tanto, la adaptación que resulta más conveniente es la raíz cúbica ya que la función propuesta por Daly sobreestima la calidad percibida ligeramente y en algunos casos especiales, el modelo lineal (no implementación de adaptación) ofrece mejores resultados.

En cuanto al comportamiento de las distintas CSF propuestas (figura 5.9), aquí, la principal diferencia con el caso anterior es que no se tendrá un valor de CSF en cada punto de frecuencia. Se calcula para cada banda de frecuencia un valor o peso de la CSF como la media en ese canal. Aún así, las consideraciones hechas en el apartado anterior siguen siendo del todo válidas. La CSF de Rust es la que proporciona peores resultados ya que presenta en general umbrales de detección menores (mayor sensibilidad o pesos mayores) y está subestimando la calidad percibida. Los modelos de Daly y Barten proporcionan resultados similares. La CSF de Barten por su forma se adapta mejor a los datos MOS pero no modela la anisotropía característica del SVH, mientras que la CSF de Daly para altas frecuencias está subestimando la calidad percibida, pero compensa al proporcionar valores o pesos para las frecuencias oblicuas menores, es decir, se necesitan umbrales de detección mayores para que la distorsión sea percibida, lo que se adecua al comportamiento del SVH.

En el caso especial de que no exista un modelo de CSF implementado, de nuevo, los resultados obtenidos indican que los ICP no se corresponden con los MOS en absoluto. Con esto lo que se pretende resaltar es la importancia que tiene el modelo de CSF dentro de la métrica perceptible, tanto en este caso como en el anterior. La CSF es la que proporciona la sensibilidad del SVH y es en la que se basan los fenómenos de enmascaramiento. Un ejemplo de esto se muestra en la figura 5.17. En el centro se muestra el mapa de visibilidad de error sin CSF para la imagen Boats distorsionada mediante ruido gaussiano. Se puede ver claramente, que el no incluir un modelo de CSF lo que implica es que no se dispone de umbrales de enmascaramiento. El mapa obtenido es igual para todas las zonas, aunque la textura del suelo, por ejemplo, sí enmascare. Este efecto, sin embargo sí está presente al incluir CSF, como se observa en la figura situada a la derecha.

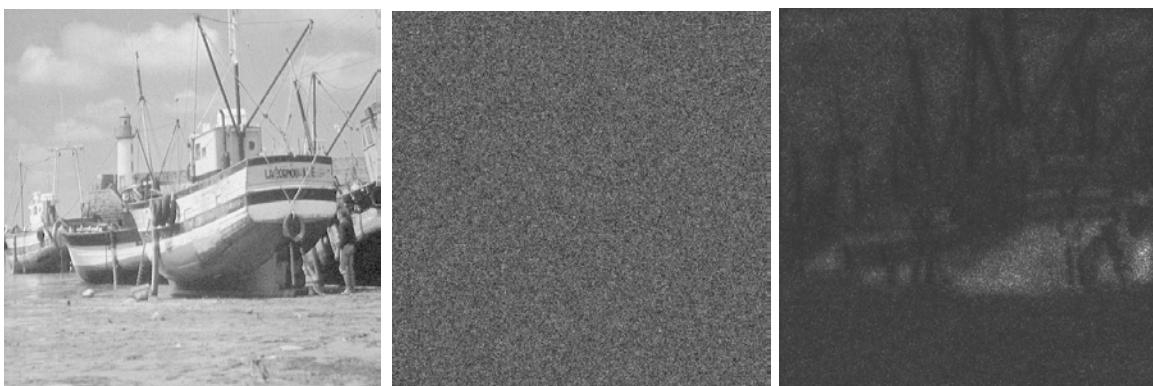


Figura 5.17: Mapas de visibilidad de error con implementación de CSF (imagen situada a la derecha) y sin implementación (imagen central) para la imagen distorsionada Boats.

Para la elección del **banco de filtros** que mejor se ajuste a los resultados MOS obtenidos hay que tener en cuenta que en esta métrica el banco de filtros va a definir el número y tamaño de los canales para los que luego se van a calcular los pesos de la CSF. Aquí, de nuevo, los bancos con selectividad en frecuencia y orientación (filtros Gabor y Cortex) van a aportar resultados mejores y muy similares. El caso de los filtros Anillo merece algunas puntuaciones, ya que, puesto que los pesos CSF dependen del banco de filtros seleccionados, en este caso se calcularán sólo 6 pesos, que no distinguen la selectividad en orientación. Es decir, aunque la CSF utilizada fuese la de Daly o Rust al calcular un peso por toda una banda de frecuencia la información de la sensibilidad en orientación queda diluida en la media. En cambio, para los casos Cortex y Gabor este efecto sí que se tiene en cuenta y los umbrales para orientaciones oblicuas serán mayores. Por tanto, el comportamiento de los filtros Anillo va a ser mejor con la CSF de Barten que con la de Daly ya que la principal ventaja de esta última (anisotropía) se está perdiendo en el cálculo de 1 peso por escala frecuencial. Esto no sucede en la métrica con CSF como filtro, ya que allí, la CSF filtra la imagen independientemente del banco de filtros utilizado.

En general, se ha observado que los filtros Anillo van a proporcionar un peor ajuste, sobre todo para distorsiones leves que empiezan a ser percibidas (valores elevados de calidad). Lo que está sucediendo es que se está subestimando la calidad que percibe el observador. La razón de este fenómeno se puede explicar considerando la ecuación 4.20 que define la visibilidad del error. En el denominador se sitúan los umbrales de enmascaramiento que están determinados por la CSF empleada y en el

numerador el contraste del error. Para el caso que se analiza, los umbrales, aunque varían, no lo harán de forma significativa. La CSF es una función simétrica y aunque al utilizar filtros Anillo se pierda la selectividad en orientación los valores de umbrales obtenidos (como una media) están en sintonía con los que se obtienen con filtros Cortex o Gabor. La diferencia principal con estos dos consiste en el contraste del error, ya que aunque los umbrales aumentan, no lo harán en igual medida que el contraste del error con lo que la visibilidad del error va a aumentar. Al disminuir el número de canales, en una banda de frecuencia se va a aglutinar el cálculo para los errores de todas las orientaciones dando un valor mayor, menos ajustado.

Los resultados obtenidos con las distintas ecuaciones de **contraste** muestran, de nuevo, que el que mejor se adapta a los resultados MOS es el contraste Global. En general, los otros tres van a subestimar la calidad percibida. El contraste Local ofrece, igual que en el caso anterior, resultados de correlación muy bajos. El contraste de Lubin proporciona resultados muy parecidos a los del contraste Local (las definiciones para algunas bandas son iguales) y con el contraste de Peli se obtienen valores ICP aún menores. La razón básica de este comportamiento y su principal diferencia con el contraste Global es que dan un cálculo de contraste píxel a píxel y son inestables sobre todo en el caso de las imágenes médicas. En concreto, para las imágenes MR1, MR3 y Mbd001 estudiadas se obtienen valores de calidad muy bajos, aún incluso para imágenes con bajo nivel de distorsión. Esto sucede porque al realizar el cálculo del contraste punto a punto habrá valores del denominador cercanos a cero sobre todo en la zona del fondo de la imagen, donde no hay señal, y por tanto esto provoca que haya posiciones en las que el contraste calculado sea muy elevado, lo que no se corresponde con la realidad.

Finalmente, en la tabla 5.2 se muestran los resultados obtenidos con algunas de las mejores variantes para este modelo, que coinciden con las de la tabla 5.1. Estos valores son ligeramente superiores a los que proporciona la métrica con la CSF como filtro, aunque la diferencia no llega a ser significativa. En cualquier caso, sí son mayores a los de la PSNR ($\rho = 0,6$) y proporcionan mapas de visibilidad equivalentes a los mostrados en el apartado anterior. La figura 5.18, muestra los resultados ICP vs. MOS para estas variantes y las curvas logísticas correspondientes.

Variante {	Adaptación CSF Banco filtros Contraste.	Raíz cúbica Modelo Barten Anillo Global	Raíz cúbica Modelo Daly Gabores Global	Raíz cúbica Modelo Daly Cortex Global
Correlación (ρ)		0,8060	0,8482	0,8510
Valor p asociado		$6,1 \cdot 10^{-6}$	$8,5 \cdot 10^{-8}$	$6,9 \cdot 10^{-8}$

Tabla 5.2: Correlación para las variantes significativas en modelo CSF por pesos.

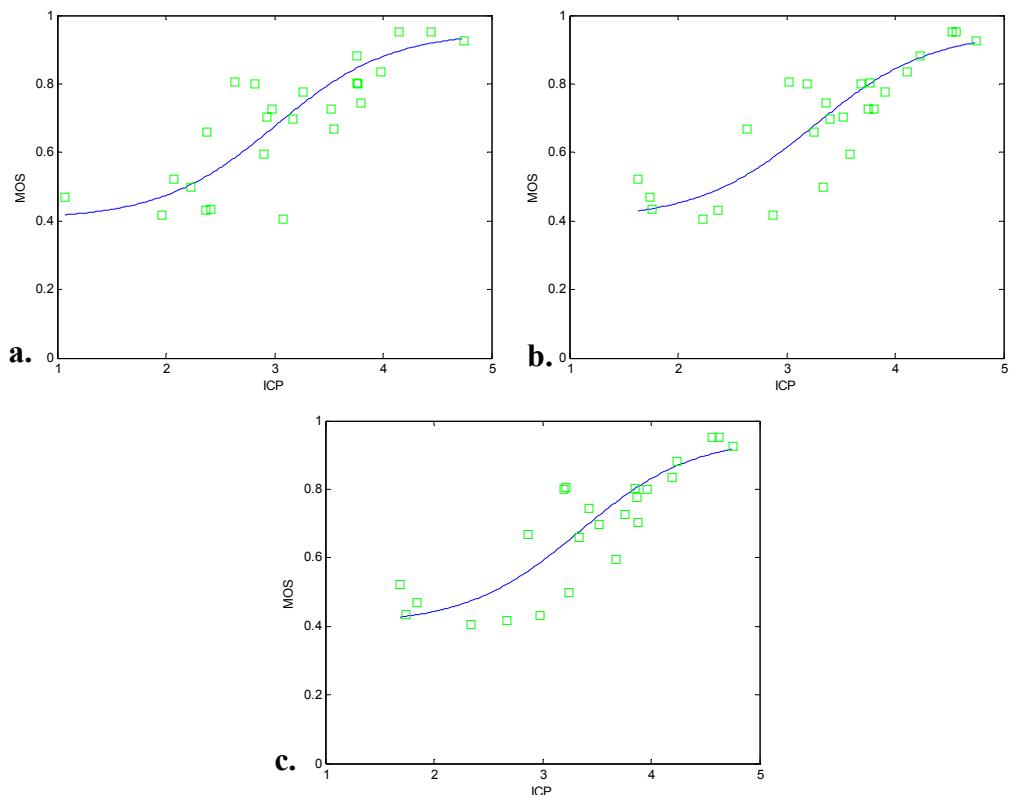


Figura 5.18: Sistema de calidad con CSF por pesos, valores ICP vs. MOS **a.** Para el modelo con adaptación de luminancia raíz cúbica, CSF de Barten, filtros Anillo y contraste Global. **b.** Para el modelo con adaptación de luminancia raíz cúbica, CSF de Daly, filtros Gabor y contraste Global. **c.** Igual que b. con filtros Cortex.

5.4. Otros modelos.

En este apartado se va a evaluar el comportamiento que presentan otras métricas de medida de calidad. En concreto, las presentadas anteriormente en el apartado 3.3.3 que abordan el problema del cálculo de un índice de calidad sin tener en cuenta características del SVH pero considerando otras cualidades propias de la información visual. En primer lugar, se analizarán los resultados obtenidos mediante la medida de calidad basada en la información de la estructura de la imagen [30, 31]. En segundo lugar, se estudiarán los valores obtenidos mediante la utilización de distribuciones Wigner [4].

5.4.1. Medida de calidad basada en información estructural (SSIM).

Este modelo propuesto por Z. Wang y A. C. Bovik [30, 31] se ha descrito ya en el Capítulo 3. Dentro de la clasificación de métricas de calidad que se ha realizado en ese capítulo se considera esta métrica como un modelo híbrido, ya que si bien su definición es plenamente matemática (sin que ningún modelo de SVH sea empleado de forma explícita), va a tener en cuenta nuevos aspectos de la imagen como la información de la estructura de los objetos del campo visual.

El punto de partida de esta medida de calidad consiste en modelar la degradación de la imagen basándose no en el error, sino en la distorsión estructural. Esta nueva filosofía propuesta resulta muy interesante ya que en realidad el SVH está bastante acostumbrado a extraer la información de la estructura del campo visual. Por tanto, como ya se ha comentado, aunque no incorpora un modelo del SVH sí que está teniendo en cuenta de modo implícito algunas de sus características más significativas. El SSIM va a dar una idea de la distorsión que se ha producido en la estructura de la imagen y esto es lo que considera como la medida de la calidad que percibiría el observador. En la figura 5.19. se muestran los mapas SSIM para dos imágenes del conjunto bajo estudio. Los puntos del mapa SSIM en blanco son aquellos en los que no se percibe la degradación de la calidad. Puede apreciarse, por ejemplo, que para el caso

de la imagen Boats las zonas en las que el ruido gaussiano es más visible son efectivamente aquellas que en el mapa SSIM se muestran más oscuras.

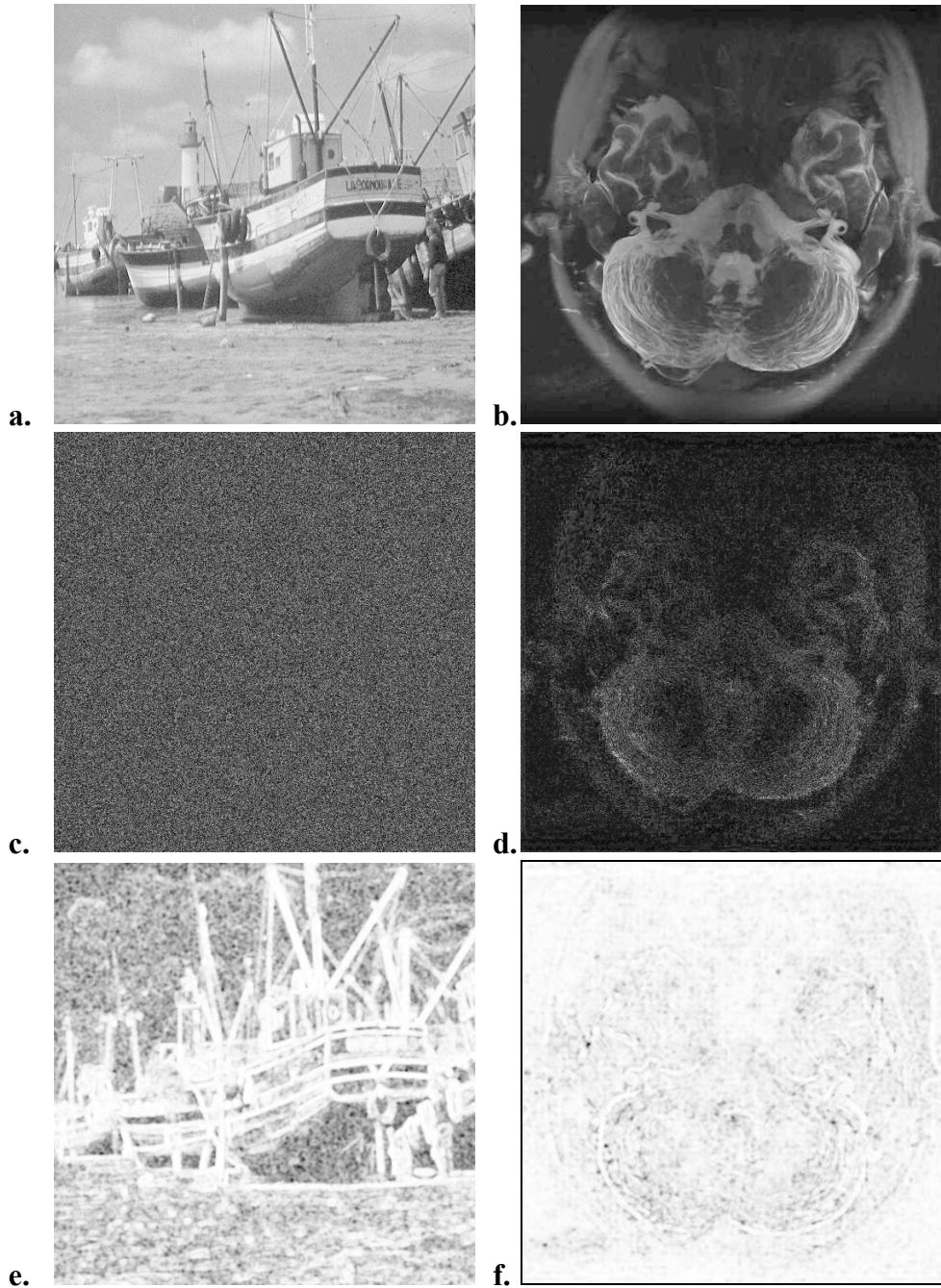


Figura 5.19: Mapas SSIM para dos imágenes del conjunto de muestra Boats y MR1. **a.** Imagen Boats con ruido gaussiano y valor PSNR = 31,55. **b.** Imagen MR1 comprimida con JPEG con PSNR = 36,84. **c.** y **d.** Mapas de error absoluto para las imágenes anteriores. **e.** y **f.** Mapas SSIM.

En el caso de la imagen MR1, sin embargo, el pixelado producido por la compresión JPEG es claramente visible en la zona superior central y en los bordes del cráneo en la parte inferior, lo que no se muestra en la figura 5.19.f. En este caso, el mapa SSIM es prácticamente blanco y el valor del índice es de 0,95 ya que la información de la estructura no se ha visto afectada. Por tanto, para este caso concreto, ni el índice ni el mapa dan información relevante de lo que percibe el observador.

En la figura 5.21 se muestran los resultados obtenidos para el conjunto de imágenes de test entre las que se encuentran los dos ejemplos analizados con anterioridad. Como puede verse, aunque el SSIM (“Structural SIMilarity”) es un valor de calidad entre 0 y 1, el margen obtenido para las imágenes estudiadas es aproximadamente de 0,5 a 1. Un valor menor o cercano a cero es difícil de conseguir ya que eso significaría que la información de la estructura se ha perdido casi por completo lo que no sucede aunque se distorsione la imagen original en gran medida. Esto se muestra en la figura 5.20. El valor del índice SSIM para este ejemplo es de 0,657. Se ve claramente que la imagen distorsionada es de muy mala calidad pero aun así, la degradación que se ha producido en la estructura de la imagen con respecto a la original MR1 no es tanta. Por ello, el índice SSIM y el mapa correspondiente dan valores mayores de los que cabría esperar, puesto que lo que están midiendo no es la degradación de la calidad que el observador percibe sino la degradación de la estructura de la imagen.

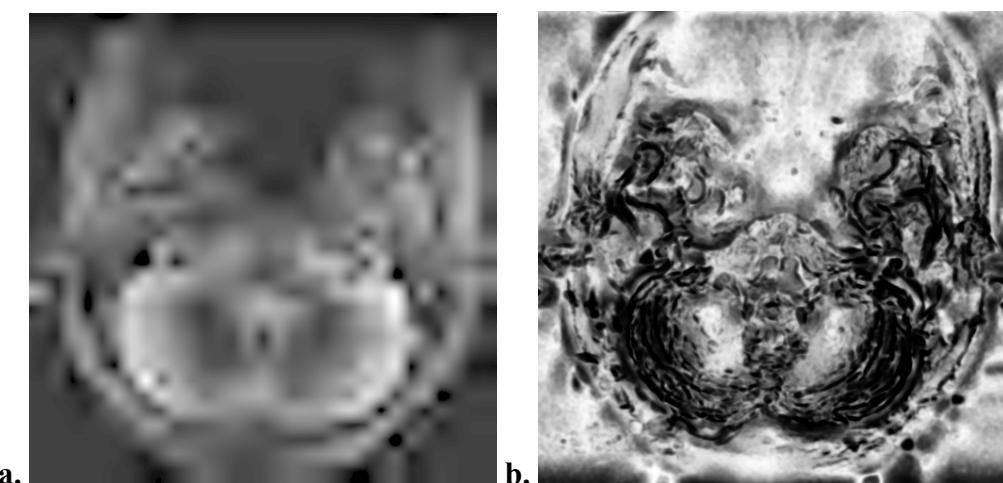


Figura 5.20: a. Imagen test MR1 comprimida con JPEG2000 y valor PSNR = 20,7.
b. Mapa SSIM correspondiente con valor SSIM = 0,657.

Tanto los valores de SSIM que se muestran en la figura 5.21, como los mapas SSIM mostrados en los ejemplos anteriores se han obtenido directamente de la aplicación de los programas Matlab disponibles en [36] sobre el conjunto de imágenes bajo estudio. Para la evaluación de los resultados obtenidos se utiliza el coeficiente de correlación de Pearson y su valor p asociado que en este caso son de:

$$\rho = 0,5467 \quad \text{y} \quad p\text{-valor} = 0,0046.$$

Este valor de correlación es el obtenido después de realizar el ajuste no lineal entre los datos objetivos y los MOS según las curvas logísticas mostradas en el apartado 5.1. El análisis estadístico de estos datos indica que aunque las variables guardan relación entre sí, esta relación no es muy fuerte. Es decir, la variable dependiente debe ser explicada por la independiente más otros factores o variables que no se tuvieron en cuenta.

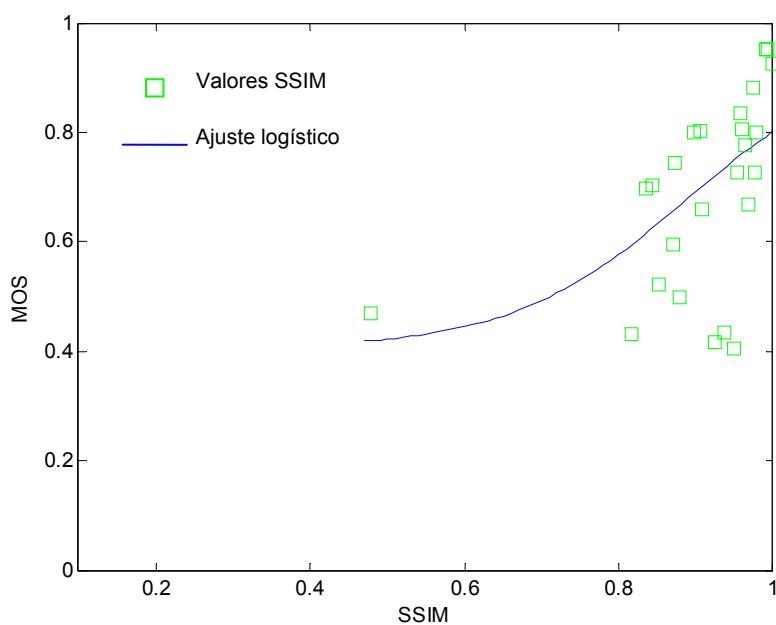


Figura 5.21: Resultados MOS vs. SSIM.

Comparando el valor de correlación obtenido en este estudio con otros en los que se ha utilizado el SSIM se observa que este es uno de los casos en el que esta medida de calidad presenta peores resultados. Así, en [6] dan un valor después de ajustar entre 0,45 y 0,72 dependiendo del tipo de distorsión introducido y en [31] de 0,96. Esta variación de resultados lo que viene a demostrar es el grado de dependencia de esta métrica con el tipo de imágenes utilizadas. De hecho, si del conjunto de imágenes bajo test se consideraran los resultados parciales que se obtendrían por un

lado para las imágenes médicas y por otro para las imágenes típicas de entornos de compresión se obtiene en el primer caso un $\rho = 0,45$ y en el segundo $\rho = 0,87$. Por tanto, el índice SSIM sí que presenta mejores resultados en comparación con medidas matemáticas simples como la PSNR ($\rho = 0,60$) si no se consideran las imágenes médicas.

En definitiva, no se debe olvidar el hecho de que el SSIM es una medida de calidad que surge al asumir que el SVH presenta como característica propia la extracción de la información de estructura del campo visual y además está adaptado para ello. Definen la información estructural de una imagen como aquellos atributos que representan la estructura de los objetos de la escena independientemente de la luminancia media y el contraste. En resumen, deducen que la medida del cambio producido sobre esta información estructural puede dar una buena aproximación de la distorsión percibida en la imagen. El enfoque de este tipo de métrica y las utilizadas anteriormente es diferente en cuanto que la degradación de la imagen se estima en un caso como el error percibido y en otro como la variación estructural que se percibe. Por tanto, lo que cabría preguntarse es en qué casos la calidad percibida coincide con la degradación de la estructura, ya que será precisamente en éstos en los que se consiga una mayor correlación con los datos MOS. La medida de calidad dada por el valor SSIM en el fondo es simplemente la combinación de tres factores (luminancia media, contraste y correlación) calculados entre la imagen original y la distorsionada. En la mayoría de los casos el cambio relativo de estos factores es pequeño con lo que, por una parte la medida que proporciona no va a ser sensible en algunos casos a la variación de calidad percibida (sobre todo para los márgenes más bajos de calidad) y por otra, variaciones del índice SSIM pequeños se deben corresponder con grandes saltos de calidad para que exista correlación con los datos MOS. Este hecho queda más claro si se observa de forma gráfica en las funciones de ajuste para esta métrica, tanto en la figura 5.21, como en la mostrada en [31]. Así, si se considera el margen de calidades con las que evalúa un observador entre 0 y 100, valores del SSIM entre 0 y 0,9 cubren un margen de calidad entre 0 y 50. En esta zona, la calidad percibida por el observador puede ser prácticamente la misma para valores SSIM, por ejemplo, de 0,6 o 0,8. Sin embargo, para valores SSIM de 0,9 a 1, pequeñas diferencias en el SSIM se van a corresponder con valores diferenciados de calidad percibida.

En el contexto de la radiología y en concreto para las imágenes médicas analizadas esta medida no proporciona buenos resultados ya que la calidad de este tipo de imagen reside en como de útil es para conseguir un diagnóstico certero y preciso. En estas imágenes las características especiales que influyen en la calidad son el contraste, el ruido y la resolución espacial. El contraste es la diferencia entre el fondo de la imagen y su estructura. Por ejemplo, en una radiografía la estructura está relacionada con la densidad óptica. El ruido produce un efecto de granulado en la imagen, siendo el más común de ellos el denominado ‘ruido cuántico’, que se debe al número limitado de fotones involucrados en el proceso de formación de la imagen. La presencia de ruido en las imágenes es inevitable en la mayoría de las ocasiones, ya que por ejemplo en imágenes de rayos X para obtener una mejor resolución habría que aumentar la dosis de radiación, lo cual obviamente no puede llevarse a cabo. En cuanto a la resolución espacial se trata de una medida de la bondad de un sistema para producir imágenes de objetos muy pequeños. Por tanto, en este caso la luminancia media, el contraste o la información de la estructura que define el SSIM no van a variar en gran medida pero aquí si que lo hará la calidad percibida ya que el test de estas imágenes debe servir para detectar incluso pequeñas anomalías.

5.4.2. Medida de calidad basada en distribuciones Wigner.

Este nuevo modelo de medida de calidad [4] se basa en las distribuciones de Wigner-Ville (WVD). Al igual que en el caso del índice SSIM, esta métrica se ha clasificado dentro del grupo de las híbridas, ya que si bien no considera la implementación de ningún modelo visual, sí va a tener en cuenta otras características propias del SVH. De hecho, la WVD se usa como base de esta métrica ya que es un hecho aceptado que en las primeras etapas del procesado visual se genera una representación conjunta espacio-frecuencia espacial, que es precisamente lo que proporciona esta distribución. Además, también tiene en cuenta de forma muy básica los mecanismos no lineales de inhibición en la corteza visual primaria al incluir la función máximo en la definición del modelo. La descripción de la medida SNR_W que se obtiene mediante esta métrica ya se ha realizado en el Capítulo 3 de este proyecto, por tanto, a continuación sólo se van a analizar los resultados que proporciona y se van a comentar algunas de sus características más relevantes.

La medida SNR_W, al igual que en el caso anterior, también tiene en cuenta la información de la estructura de la imagen, pero la diferencia principal estriba en que ésta no es una medida de la degradación de la estructura sino de la visibilidad del error basándose en que éste sea o no estructurado. Es decir, la idea básica de esta métrica consiste en explotar el hecho de que las distorsiones estructuradas, que normalmente son las más molestas (el SVH está adaptado para extraer la información de la estructura del campo visual) se encuentran concentradas en el dominio espacio-frecuencia espacial. En general, entre dos distorsiones que posean la misma energía aplicadas sobre la misma imagen, el observador va a percibir de forma más clara la que tenga un pico de energía localizado en espacio y frecuencia, es decir, la estructurada. Por tanto, puntuará con valores menores de calidad imágenes con distorsiones estructuradas que son más molestas al ser más perceptibles. En definitiva, este es el concepto sobre el que se asienta esta métrica y la definición dada para la SNR_W. En resumen se pueden destacar tres propiedades fundamentales de esta definición:

1. La utilización de WVD permite localizar las estructuras tanto en la imagen original como en la distorsionada. Entonces, en aquellos puntos de la imagen distorsionada en los que la estructura del ruido se haga patente (tenga un pico) la diferencia resultante en el denominador será mayor dando valores SNR_W menores.
2. El ruido introducido en la imagen no va a afectar únicamente en aquellos puntos donde la WVD tenga coeficientes para el ruido distintos de cero, sino que también introduce términos de interferencia, con lo que cuanto mayores sean los coeficientes WVD de ruido mayores serán las diferencias entre la WVD de la imagen original y la distorsionada.
3. Se utiliza el máximo para retener el componente WVD de mayor energía para cada posición simulando de esta forma un enmascaramiento básico.

Estas consideraciones a la hora de calcular una medida de calidad hacen que los resultados obtenidos sean mejores que medidas simples como la PSNR. Así, en la figura 5.22 se muestran con una cruz (+) los valores SNR_W para el conjunto de imágenes bajo estudio y su curva de ajuste. Con estos datos el coeficiente de correlación de Pearson y su valor *p* asociado son de:

$$\rho = 0,7474 \quad \text{y} \quad p\text{-valor} = 0,00002.$$

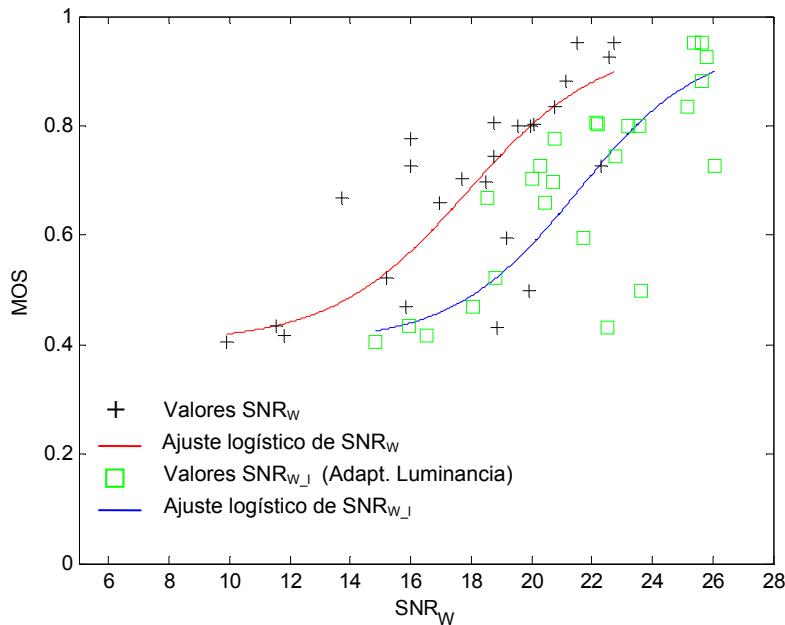


Figura 5.22: Resultados MOS vs. SNR_Wigner.

El desarrollo de esta métrica está motivado por la necesidad de una medida de calidad simple que se corresponda con la percepción visual humana en mayor medida que otras métricas como la PSNR, SNR o RMSE. Este objetivo se ha alcanzado en parte, ya que el valor de correlación conseguido presenta una mejora en relación al de la PNSR ($\rho = 0,6037$). Sin embargo, habría que plantearse si esta mejora compensa ya que el método desarrollado es complejo y costoso desde el punto de vista computacional. En el Capítulo 3 se vieron algunos de estos inconvenientes. Por ejemplo, la WVD en 2D puede presentar dispersión espectral “aliasing” y términos de interferencia que pueden dificultar el proceso de análisis. El uso de sobremuestreo o de suavizamiento puede ser necesario para reducir este aliasing y para los términos de interferencia se hace imprescindible la utilización de filtros en la definición de la WVD o mediante la utilización de la señal analítica. Se debe buscar un compromiso entre términos de interferencia y resolución, lo que no es sencillo. Además, otra dificultad añadida es la alta dimensionalidad (4D) de la 2D-WVD. No es factible utilizar la WVD como tal, sino una pseudo distribución de Wigner para la que hay que buscar la señal analítica (2D). Por todo ello, si se utilizan este tipo de distribuciones convendría poder explotar mejor sus propiedades e incluirlas en un modelo más complejo de sistema visual humano. Una primera aproximación en esta dirección consistiría en incluir la adaptación de luminancia dentro de la métrica descrita como muestra la figura 5.23.

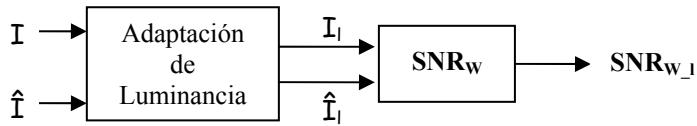


Figura 5.23: Adaptación de luminancia para la medida SNR_W .

Para la implementación de la adaptación de luminancia en el dominio espacial valdría con utilizar cualquiera de las funciones descritas en el apartado 4.4. Las entradas serían, como siempre, la imagen original y la distorsionada sólo que en este caso es necesario que estén en cd/m^2 . En el caso de calcular sólo la SNR_W valdría con que se hubiesen considerando las opciones de visualización, es decir, para un monitor calibrado no sería necesario hacer nada puesto que el paso de niveles de gris a cd/m^2 es una función lineal, pero si no lo está hay que considerar que la función gamma del monitor está modificando la visualización, oscureciendo la imagen. Finalmente, las entradas (I e \hat{I}) sobre las que se va a aplicar la SNR_W , dando como resultado la medida de calidad denominada SNR_{W_1} , son I e \hat{I} sometidas a la función no lineal seleccionada para modelar la adaptación. Con esto lo que se consigue reflejar es que si la luminancia local aumenta, un mayor nivel de distorsión puede ser tolerado. Los resultados obtenidos al aplicar esta modificación sobre la métrica original se muestran en la figura 5.22 mediante cuadrados (\square), con su curva de ajuste correspondiente. Se ha utilizando la función raíz cúbica como modelo de adaptación sobre el conjunto de imágenes bajo estudio para la adquisición de estos datos. El valor de correlación obtenido para la medida SNR_{W_1} y su valor p asociado son:

$$\rho = 0,7704 \quad \text{y} \quad p\text{-valor} = 0,000006.$$

Aunque este valor representa una mejora frente al cálculo anterior, ésta no es muy significativa ya que como se ha manifestado la implementación que se haga para la adaptación de luminancia no tiene gran impacto en los modelos de medida de calidad. Su relevancia dependerá de si las imágenes presentan un gran contraste o si la localización del enmascaramiento de luminancia es significativa. Además, la función máximo en la definición de la SNR_W ya está considerando en cierta forma este enmascaramiento producido por la luminancia por lo que tampoco se puede llegar a esperar un gran cambio en la correlación obtenida. Aún así, modelar la adaptación de

luminancia resulta realmente sencillo y aunque la mejora que proporciona en los resultados sea pequeña no deja de ser un avance hacia una medida que se corresponda con lo que el observador percibe. Lo ideal sería buscar la forma de incluir de alguna manera dentro de esta medida la CSF, que es la piedra angular de los modelos visuales. Para ello se tendría que hacer un cálculo de contrastes que se ajustara con la representación en el plano espacio-frecuencia espacial que proporciona la WVD. Queda, por tanto, esta línea de investigación abierta a futuras implementaciones.

Otra posible mejora en el uso de distribuciones de Wigner en medida de calidad de imagen sería que se proporcionaran mapas de visibilidad del error además del valor SNR_W como índice global. Aunque no se dispone de esta información que permita reconocer los puntos en los que el error es perceptible, en el entorno gráfico diseñado (ver Anexo B) se ha optado por mostrar el denominador de la definición de la SNR_W : $\max_{u,v} \left(|PWD_I(i, j, u, v) - PWD_{\hat{I}}(i, j, u, v)| \right)$, que da una idea de la energía y estructura de la distorsión.

En general, el valor de SNR_W no presenta ninguna dependencia de la imagen de entrada, es decir, la correlación obtenida se adapta igual para las imágenes médicas que para las que no lo son. A pesar de ello, el valor SNR_W obtenido no es equivalente a la calidad percibida por el observador, no sólo por los resultados de correlación obtenidos, sino en parte, por la propia definición o concepto en el que se basa el índice. Así, como ya se ha comentado, básicamente, el SNR_W es sensible a las variaciones en la estructura del error. Las distorsiones estructuradas son más visibles porque el SVH está adaptado para extraer la información estructural. Aunque este hecho es claro, es sólo uno de los factores que influyen en la percepción, existiendo muchos otros y de gran importancia: sensibilidad al contraste, enmascaramiento, procesos de atención visual, etc. Por consiguiente, aunque se obtengan buenos resultados en algunos casos, la definición de este índice no es general. Esto se ve ilustrado en la figura 5.24. Se ha tomado como ejemplo la imagen original Bike sometida compresión con JPEG, JPEG2000 y un último caso en el que se ha añadido ruido gaussiano. Se muestra sólo una sección de la imagen donde, para los tres casos, se percibe más claramente la distorsión, en el resto, el contenido de la imagen enmascara más y no es tan evidente. Así, la PSNR de estas imágenes es de 30, 29 y 28dB, siendo la mejor la distorsionada con ruido gaussiano y la peor la JPEG. El valor del índice SNR_W para estas imágenes es 19,15 (gaussiana), 18,46

(JPEG2000) y 17,71(JPEG). Los valores MOS dados por los observadores puntúan justamente al contrario las imágenes. La de peor calidad es la sometida a ruido gaussiano ya que este se percibe claramente como un granulado sobre todo en la zona interior de las frutas que es uniforme. A continuación, iría la imagen comprimida con JPEG2000 ya que en este caso lo que está sucediendo es que se pierden los bordes y esto resulta bastante molesto para el observador que tiene una sensación general de emborronamiento al ver la imagen. El tercer caso es el mejor ya que aunque la compresión con JPEG produce artefactos también en los bordes, estos están enmascarando en parte esta distorsión con lo que se distingue en menor medida. Esto es un poco controvertido, ya que cuando el efecto de blocking es muy pronunciado la sensación de pérdida de calidad en JPEG debería ser mayor que con ruido gaussiano, pero también hay que tener en cuenta el carácter de no-expertos de los observadores.

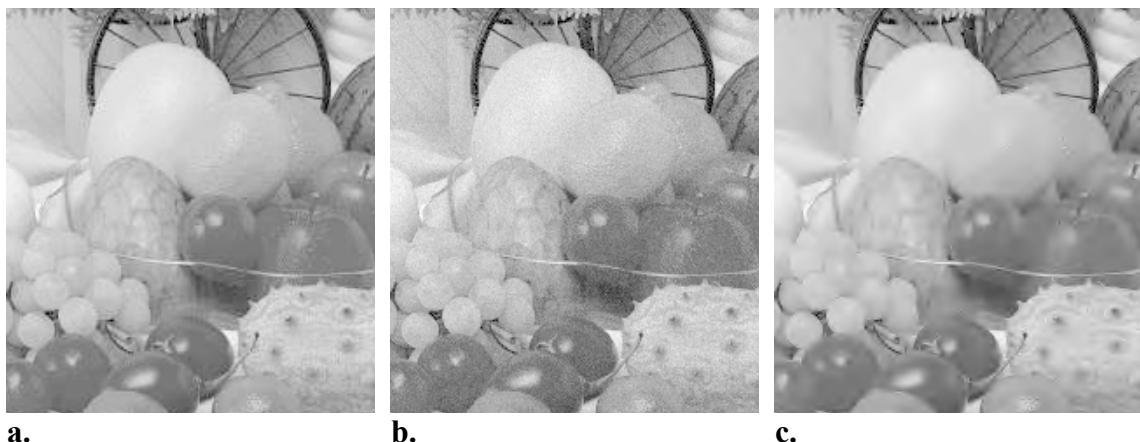


Figura 5.24: Sección de la imagen *Bike* sometida a distintas distorsiones.
a. Compresión con JPEG con 1,42 bpp. y PSNR = 28,06. **b.** Ruido gaussiano con PSNR = 30,12. **c.** Compresión con JPEG2000 con 0,99 bpp y PSNR = 29,18.

En conclusión, el uso de WVD proporciona de forma natural algunas de las características o propiedades del SVH, pero éstas no son suficientes para que la métrica resulte eficiente desde el punto de vista de los resultados obtenidos en función del coste que implica llevar a cabo la PWD. Por ello, ya en [3] se han planteado nuevos modelos con ondículas que posean las características de la WVD pero que no sean tan costosas. Aún así, las WVD y el SNR_w son un punto de partida para la investigación de nuevas métricas, buscando un modelo que aproveche mejor las propiedades de las WVD y que incluya un modelo más completo del SVH.

6

Conclusiones.

En este capítulo se pretende dar la visión general de los objetivos alcanzados mediante la realización del presente proyecto, así como de las posibles ampliaciones que se podrían llevar a cabo tomando como base el trabajo realizado. Por tanto, en primer lugar se expondrán las conclusiones generales que se han obtenido de la implementación de las distintas métricas de calidad. De esta manera se va a disponer de un enfoque global que resume las principales características de los modelos de medida de calidad de imagen y en concreto de las métricas de calidad perceptible. En segundo lugar, se van a indicar posibles recomendaciones para la optimización e investigación futuras.

6.1. Conclusiones.

El objetivo fundamental de este proyecto ha sido la investigación de los atributos visuales perceptivos de mayor relevancia en la evaluación de la calidad de imagen y como interactúan entre sí. En realidad, este propósito se ha logrado mediante la implementación de los modelos visuales (Capítulo 4) y el análisis de resultados obtenidos en su incorporación dentro de una métrica (Capítulo 5). Además, la creación del entorno gráfico detallado en el Anexo B proporciona la herramienta perfecta para la investigación y prueba de las distintas combinaciones de modelos visuales. Por tanto, la realización del proyecto como tal, ha concluido y lo que queda es exponer las conclusiones generales que se han extraído en el desarrollo del mismo.

Merece especial atención la definición que se hace para la calidad de una imagen. Aunque en principio este hecho parece bastante básico, en la aplicación práctica las métricas se construyen en base a la definición que proponen para la calidad. Por tanto, un concepto equivocado o no muy completo va a dar como resultado métricas de calidad parciales. En concreto, este es un fallo que presentan en mayor o menor medida todas las medidas de calidad que se han desarrollado en este proyecto y que ya se ha hecho notar en capítulos anteriores. Así y de forma simplificada, el SSIM define la calidad percibida como la degradación de la estructura, el SNR_W la define en función de la estructura del error y los ICP se basan en la visibilidad del error. Si bien está claro que los tres factores mencionados influyen en la calidad que el observador percibe, no son los únicos y por tanto proporcionarán buenos resultados para aquellas imágenes en las que éstas sean las características predominantes. Puesto que ninguna de las tres medidas anteriores tiene en consideración las definiciones de las otras dos está claro que son incompletas, todo esto si añadir otros factores que también influyen y que ninguna de las tres tienen en cuenta, como por ejemplo los procesos de atención visual. Aún así, no todas las características mencionadas en las que se basa la calidad percibida tienen la misma relevancia. Así, la implementación del SVH dentro del modelo es fundamental y hace que las métricas obtengan en general resultados mejores en un rango más amplio de imágenes y distorsiones.

Los modelos de medida de calidad perceptible, precisamente, incorporan este modelado del SVH. De hecho, lo que se ha pretendido es disponer de estructuras para la métrica en las que luego se insertan los modelos de los distintos componentes del SVH. Esto proporciona un gran abanico de posibilidades y hace que la extracción de conclusiones sea en ocasiones complicada, sobre todo, porque algunas de ellas no se han explorado. Aún así en el Capítulo 5 se ha llegado a las siguientes:

La adaptación de luminancia es un componente que va a depender de la imagen de entrada y de la distorsión a la que ha sido sometida, siendo en algunos casos más recomendable no implementarla. Aún así la función raíz cúbica tiene un buen comportamiento para la mayoría de las imágenes en general.

La CSF que mejores resultados va a proporcionar será la que posea una función paso banda más aguda, eliminando bastante las altas frecuencias.

Los bancos de filtros selectivos en orientación y frecuencia proporcionan resultados más acordes con la calidad percibida. Sin embargo, el inconveniente es que son modelos mucho más lentos pues tienen que procesar un mayor conjunto de datos. Aquí, el comportamiento del modelo de calidad perceptible con CSF por pesos y CSF como filtro presenta resultados distintos. Así, en el primero, la diferencia entre la correlación obtenida para un banco de filtros selectivo en orientación y uno que no lo sea, es mayor que en el caso del modelo de calidad perceptible con CSF como filtro. Esto se debe principalmente a que al implementarse la CSF como un filtrado previo a la descomposición en canales el modelo es más independiente al banco de filtros utilizado, mientras que en el caso del cálculo de la CSF por pesos, estos pesos se van a definir precisamente en los canales a la salida del banco de filtros.

En cuanto a las distintas definiciones de contraste, la que muestra unos resultados globales mejores es el contraste Global. Sobre todo porque el resto de contrastes facilitados no se adecuan a las características de las imágenes médicas que se han incluido en el estudio. Sin embargo, en el caso de las imágenes naturales las demás definiciones dan buenos resultados.

Finalmente, aunque los valores de correlación obtenidos para el modelo de calidad perceptible con CSF por pesos sean algo más elevados, ambas métricas presentan resultados muy similares y esta variación no es realmente significativa.

6.2. Posibles ampliaciones.

En la elaboración de un sistema de medida de calidad perceptible es fundamental el modelado que se haga del SVH. Por tanto, las posibles mejoras y campos de investigación futura se basan, precisamente, en la implementación de nuevos mecanismos que formen parte del proceso de percepción o en la mejora de los ya existentes. Como ya se ha expuesto con anterioridad, el SVH es de una gran complejidad y en este sentido hay abiertas dos posibilidades en cuanto a su simulación. Por un parte se analizan en detalle las características visuales y se intenta reproducir su comportamiento mediante el uso de funciones que van a resultar complejas. Por otra, el SVH se trata como una especie de caja negra y aunque se estudia también el comportamiento de los distintos procesos, de lo que se trata es de encontrar sistemas de medida alternativos que proporcionen medidas cercanas a las subjetivas pero sin tener en cuenta la gran complejidad del SVH. El primero de los métodos es el que se sigue habitualmente y en el que se han basado la mayoría de las métricas y también, el sistema de medida ICP propuesto. Un ejemplo de la segunda podría ser el SSIM [30,31].

Con todo esto lo que se pretende es dar una idea del gran campo de análisis y estudio del que se dispone a la hora de tratar con la subjetividad que plantea el SVH. Por tanto, las ampliaciones que podrían hacerse para este proyecto pueden seguir diferentes líneas:

1. Ampliaciones y/o mejoras para los mecanismos de visión estudiados.

- En cuanto a la adaptación de luminancia se podrían plantear nuevas funciones como las sigmoidales. También se podría tratar de implementar un método que trate la adaptación de luminancia en lugar de en el espacio en el dominio frecuencial.
- La función CSF es clave dentro del modelo. Comparar los resultados que se obtendrían si se dispusiese de un sistema de medida para la misma en el lugar que se emplea para visualizar las imágenes cuya calidad se quiere evaluar, con los que se obtienen mediante el uso de las CSF propuestas sería interesante. De hecho, como ya se ha comentado, las funciones analíticas presentadas plantean varias inconvenientes. Un modelo de CSF como el de

Rust, pero medido en el lugar de trabajo probablemente mejoraría la métrica puesto que se ajustaría mejor. El problema de esto es que la medida de la CSF, si se pretende hacer con cierta rigurosidad, es bastante complicada. En cualquier caso, esta comparativa resultaría interesante para ver realmente si con la forma de la CSF es suficiente o tener un modelo lo más cercano posible al real mejora la medida objetiva de la calidad.

- En cuanto al banco de filtros podrían realizarse medidas para otras N_f y N_o , e incluso, implementar otros bancos de filtros. Sin embargo, esto último no es lo más recomendable puesto que no va a aportar cambios significativos.
- El cálculo de contrastes también podría ampliarse considerando nuevas definiciones ya existentes, como la propuesta en [35].
- La implementación del enmascaramiento también puede mejorarse, por ejemplo, considerando el efecto del enmascaramiento entre canales, que además, existe en el SVH.
- El cálculo de mapas e índices de error se ha basado principalmente en la aplicación de la métrica de Minkowski. Los sumatorios efectuados pueden evaluarse dando valores distintos a los exponentes y también realizando una ponderación previa de los canales añadiendo probabilidades de ocurrencia del error, etc.
- Se podría implementar un modelo para los procesos de atención.

2. Otras métricas.

- Las conclusiones y el análisis en conjunto y por separado de los distintos componentes del SVH pueden ser utilizados para la inserción de los mismos en otras métricas. Un ejemplo muy sencillo de esto sería el SNR_{WI} calculado.
- En el entorno gráfico construido se pueden incluir las métricas de calidad que se consideren oportunas a fin de abarcar una mayor área de estudio y comparación. Por ejemplo, las nuevas definiciones de calidad realizadas en [3] y que siguen la línea de la SNR_W .

7

ANEXOS.

ANEXO A: Imágenes y resultados del test DSCQS.

A continuación se muestran las imágenes escogidas como referencia en la realización del test DSCQS (“Double-stimulus continuous quality-scale”). Las tres primeras son imágenes médicas. A.1 y A.2 se corresponden con imágenes radiológicas MRI de un cerebro y una rodilla respectivamente. A.3 es una mamografía. Las dos imágenes restantes, A.4 y A.5, son de utilización común en entornos de compresión.

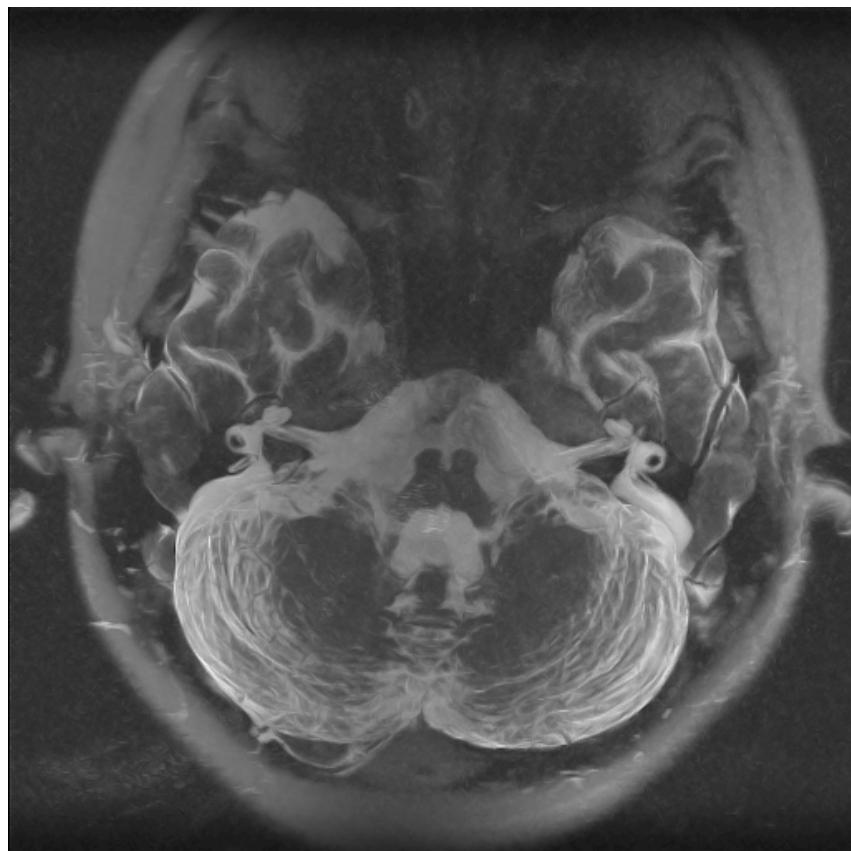


Figura A.1: Imagen de referencia MR1.



Figura A.2: Imagen de referencia MR3.

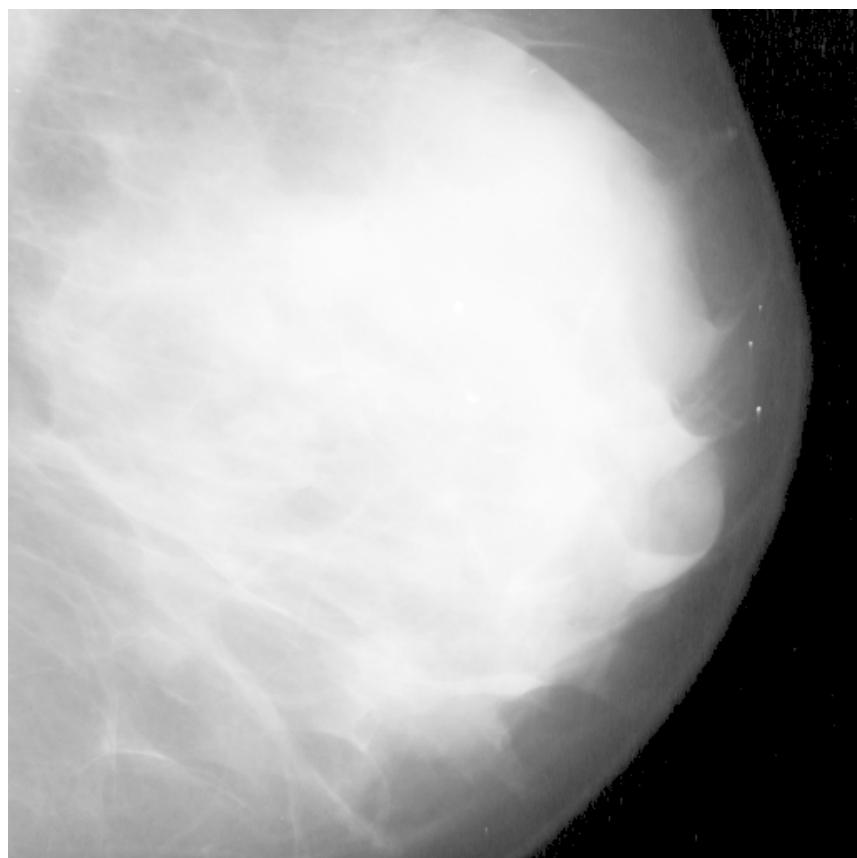


Figura A.3: Imagen de referencia Mbd001.



Figura A.4: Imagen de referencia Boats.



Figura A.5: Imagen de referencia Bike.

En la siguiente tabla se van a mostrar los resultados que se han obtenido al realizar el test DSCQS con 18 observadores. Para cada imagen de referencia se tienen cinco imágenes distorsionadas que se generan al someter a la original a una compresión, marcado de agua o introducción de algún tipo de ruido. La tabla muestra, para cada una de ellas, la PSNR y el valor entre 1 y 5 (computado al restar la puntuación dada a la imagen de referencia menos la puntuación de la distorsionada y llevada al rango [1,5] indicado). Aquellas medidas que sean mayores que 5 indican que el observador ha puntuado con un valor mayor de calidad la imagen distorsionada que la de referencia. Con estos valores se obtienen las medidas MOS que se muestran en las tablas y figuras del Capítulo 5.

Imágenes	Tipo de distorsión, (tasa)	PSNR	Observadores					
			1	2	3	4	5	6
MR1	JPEG2000 1,5890 bpp.	49,46	5,00	5,73	5,00	4,93	4,93	4,78
	JPEG2000 0,6975 bpp.	44,02	3,40	2,75	3,53	5,26	4,46	3,13
	JPEG 0,7013 bpp.	40,15	3,35	4,55	1,82	4,95	4,95	3,75
	JPEG2000 0,1990 bpp.	37,63	2,04	1,91	2,18	4,17	4,24	2,97
	JPEG 0,4529 bpp.	36,84	3,95	0,42	0,42	3,25	3,32	2,29
Mbd001	JPEG 0,2192 bpp.	44,14	5,00	4,38	4,24	4,27	4,56	4,27
	JPEG 0,2112 bpp.	41,26	4,16	4,31	3,22	4,56	4,42	3,91
	JPEG2000 0,0993 bpp.	39,33	3,47	4,82	3,25	4,64	4,49	4,13
	JPEG 0,1540 bpp.	37,96	3,15	1,82	0,56	4,08	4,48	2,75
	JPEG2000 0,0119 bpp.	31,99	3,42	3,42	2,09	2,79	3,04	2,79

Imágenes	Tipo de distorsión, (tasa)	PSNR	Observadores					
			1	2	3	4	5	6
MR3	JPEG2000 0,4999 bpp.	41,46	3,40	2,75	4,20	5,00	4,49	4,56
	JPEG 0,5372 bpp.	38,25	3,46	4,59	4,92	4,66	4,99	3,79
	JPEG2000 0,1978 bpp.	35,32	4,99	3,39	3,66	4,79	3,59	3,53
	JPEG2000 0,0996 bpp.	32,81	2,20	0,86	3,53	3,93	4,52	2,86
	Marcado de agua.	28,70	4,84	0,98	2,24	3,51	3,24	1,64
Boats	Marcado de agua.	45,02	5,22	5,00	5,00	5,00	5,18	4,35
	JPEG2000 1,4909 bpp.	39,01	4,92	4,92	4,86	4,99	4,99	3,26
	JPEG 0,9479 bpp.	34,20	4,99	4,99	4,99	4,46	4,46	4,19
	JPEG2000 0,4985 bpp.	33,36	4,99	4,99	4,99	4,19	4,72	4,59
	Ruido gaussiano	31,55	2,36	2,36	0,62	3,82	3,62	1,82
Bike	Marcado de agua.	47,37	5,00	4,85	4,20	4,85	4,93	3,91
	JPEG2000 1,4965 bpp.	32,07	3,39	4,72	3,39	4,59	4,66	4,99
	Ruido gaussiano	30,12	3,73	4,59	3,46	3,26	3,39	2,59
	JPEG2000 0,9992 bpp.	29,19	3,33	4,86	3,66	3,53	3,86	4,32
	JPEG 1,4118 bpp.	28,07	3,53	3,26	3,39	4,06	4,39	3,79

Imágenes	Tipo de distorsión, (tasa)	PSNR	Observadores					
			7	8	9	10	11	12
MR1	JPEG2000 1,5890 bpp.	49,46	4,20	4,78	5,00	5,22	4,71	4,71
	JPEG2000 0,6975 bpp.	44,02	4,06	3,66	4,99	3,66	3,53	3,53
	JPEG 0,7013 bpp.	40,15	4,02	4,68	4,55	2,02	3,89	3,62
	JPEG2000 0,1990 bpp.	37,63	3,97	4,44	3,77	3,51	2,97	4,17
	JPEG 0,4529 bpp.	36,84	3,09	3,05	3,25	3,35	3,89	2,72
Mbd001	JPEG 0,2192 bpp.	44,14	4,56	4,64	5,00	5,73	4,20	4,49
	JPEG 0,2112 bpp.	41,26	4,56	3,98	4,56	3,47	4,20	4,42
	JPEG2000 0,0993 bpp.	39,33	3,98	4,49	4,71	3,47	4,49	4,05
	JPEG 0,1540 bpp.	37,96	3,82	3,29	4,35	3,15	2,49	3,29
	JPEG2000 0,0119 bpp.	31,99	3,30	3,55	2,79	2,15	2,40	2,53
MR3	JPEG2000 0,4999 bpp.	41,46	4,64	4,56	4,71	4,13	3,76	4,20
	JPEG 0,5372 bpp.	38,25	3,66	4,32	4,99	3,53	4,99	4,46
	JPEG2000 0,1978 bpp.	35,32	4,12	4,46	4,06	3,79	4,19	4,19
	JPEG2000 0,0996 bpp.	32,81	4,06	3,46	3,66	2,06	4,19	4,19
	Marcado de agua.	28,70	4,44	4,10	4,30	3,51	4,04	4,17

Imágenes	Tipo de distorsión, (tasa)	PSNR	Observadores					
			7	8	9	10	11	12
Boats	Marcado de agua.	45,02	4,20	4,53	5,15	5,00	5,00	5,07
	JPEG2000 1,4909 bpp.	39,01	4,06	4,19	4,86	4,99	4,32	4,19
	JPEG 0,9479 bpp.	34,20	4,99	4,99	4,99	4,99	4,19	4,99
	JPEG2000 0,4985 bpp.	33,36	4,06	4,32	4,72	3,26	3,93	3,53
	Ruido gaussiano	31,55	2,49	3,95	3,95	2,36	4,62	4,22
Bike	Marcado de agua.	47,37	4,56	5,00	5,07	5,00	4,49	5,00
	JPEG2000 1,4965 bpp.	32,07	4,99	4,06	4,06	4,99	4,59	4,06
	Ruido gaussiano	30,12	3,73	3,39	4,72	2,20	2,06	4,06
	JPEG2000 0,9992 bpp.	29,19	2,86	4,19	4,72	4,86	3,93	3,53
	JPEG 1,4118 bpp.	28,07	2,20	4,19	4,86	4,86	3,79	4,59

Imágenes	Tipo de distorsión, (tasa)	PSNR	Observadores					
			13	14	15	16	17	18
MR1	JPEG2000 1,5890 bpp.	49,46	4,64	5,00	4,42	4,35	4,20	5,00
	JPEG2000 0,6975 bpp.	44,02	3,79	3,26	3,79	3,53	4,99	4,99
	JPEG 0,7013 bpp.	40,15	2,69	2,02	2,95	4,28	3,35	4,68
	JPEG2000 0,1990 bpp.	37,63	3,37	2,31	0,58	0,84	0,71	1,11
	JPEG 0,4529 bpp.	36,84	3,75	0,56	2,29	3,62	3,35	0,56
Mbd001	JPEG 0,2192 bpp.	44,14	4,56	4,05	4,20	4,13	4,20	5,00
	JPEG 0,2112 bpp.	41,26	4,56	2,67	3,18	4,13	3,33	2,67
	JPEG2000 0,0993 bpp.	39,33	4,64	3,25	4,05	3,62	4,20	4,13
	JPEG 0,1540 bpp.	37,96	4,08	2,75	2,22	2,89	1,56	3,15
	JPEG2000 0,0119 bpp.	31,99	3,04	0,37	2,02	2,28	2,40	3,55
MR3	JPEG2000 0,4999 bpp.	41,46	4,20	3,40	4,13	5,00	4,13	4,27
	JPEG 0,5372 bpp.	38,25	4,59	2,06	3,93	4,99	3,53	4,46
	JPEG2000 0,1978 bpp.	35,32	2,86	2,06	3,39	2,99	1,80	3,53
	JPEG2000 0,0996 bpp.	32,81	3,39	2,06	2,73	3,53	2,33	2,06
	Marcado de agua.	28,70	3,37	0,44	3,37	2,44	1,78	-0,62

Imágenes	Tipo de distorsión, (tasa)	PSNR	Observadores					
			13	14	15	16	17	18
Boats	Marcado de agua.	45,02	4,49	3,18	5,22	4,93	5,00	4,93
	JPEG2000 1,4909 bpp.	39,01	4,32	2,99	4,46	3,79	3,39	4,59
	JPEG 0,9479 bpp.	34,20	2,86	3,53	1,13	3,66	3,53	3,66
	JPEG2000 0,4985 bpp.	33,36	3,66	2,06	3,53	3,53	4,06	2,46
	Ruido gaussiano	31,55	2,89	0,49	2,16	4,09	2,36	0,89
Bike	Marcado de agua.	47,37	4,71	3,40	5,51	4,20	5,00	5,00
	JPEG2000 1,4965 bpp.	32,07	4,32	3,39	4,06	3,66	3,66	4,19
	Ruido gaussiano	30,12	3,79	0,60	4,46	4,99	3,53	2,20
	JPEG2000 0,9992 bpp.	29,19	4,06	3,39	3,93	3,53	3,39	2,33
	JPEG 1,4118 bpp.	28,07	4,46	3,66	4,06	2,59	2,06	4,99

ANEXO B: Entorno gráfico.

Para facilitar el manejo de las herramientas de medida de calidad y la implementación de modelos de forma dinámica se ha desarrollado la interfaz de usuario que se muestra en la figura B.1. En él están contenidas todas las herramientas para el diseño y comparación de los distintos sistemas de medida de calidad que se han descrito en los diferentes capítulos de este proyecto. Esta interfaz se abre ejecutando el programa *interfaz_usuario.m*.

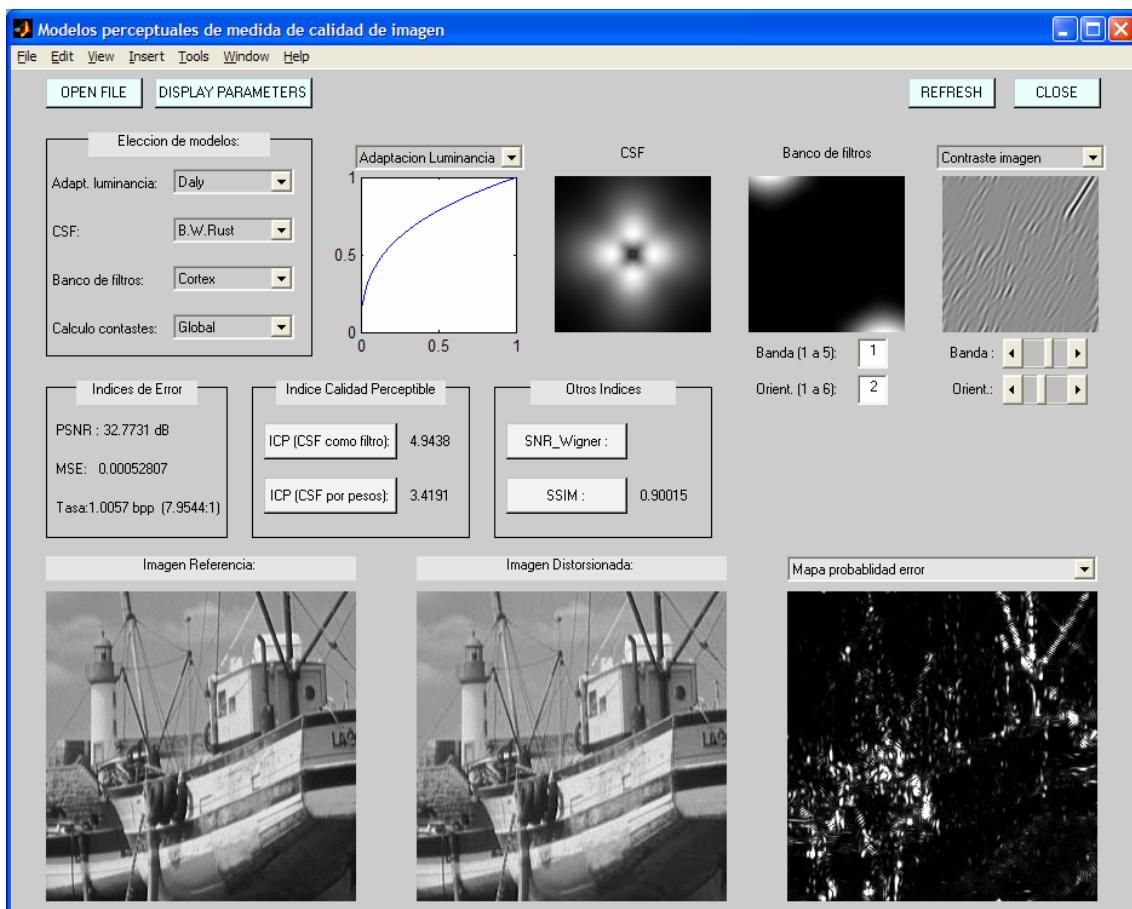


Figura B.1: Interfaz de usuario para modelos de medida de calidad.

El funcionamiento es bastante sencillo aunque hay que seguir una serie de pasos básicos para que el sistema se comporte según lo previsto. A continuación se va a explicar brevemente como se debe utilizar este entorno y en qué consisten las diferentes opciones que se facilitan, de forma que se disponga así de un manual de usuario básico.

1.- Calibrado y carga de imágenes.

- Lo primero que habría que hacer para poder trabajar de forma correcta es introducir los **parámetros de calibrado** que se muestran en la figura B.2. Si el monitor está calibrado no sería necesario conocer el valor de gamma, pero sí todos los demás. Estos datos se van a guardar en C:\Matlab\work\disp_param.mat y serán los que se tomen por defecto a partir de entonces y mientras que no se introduzcan otros. Para cambiar el directorio donde se guardan habría que modificar *display_parameters.m* (ver anexo C).

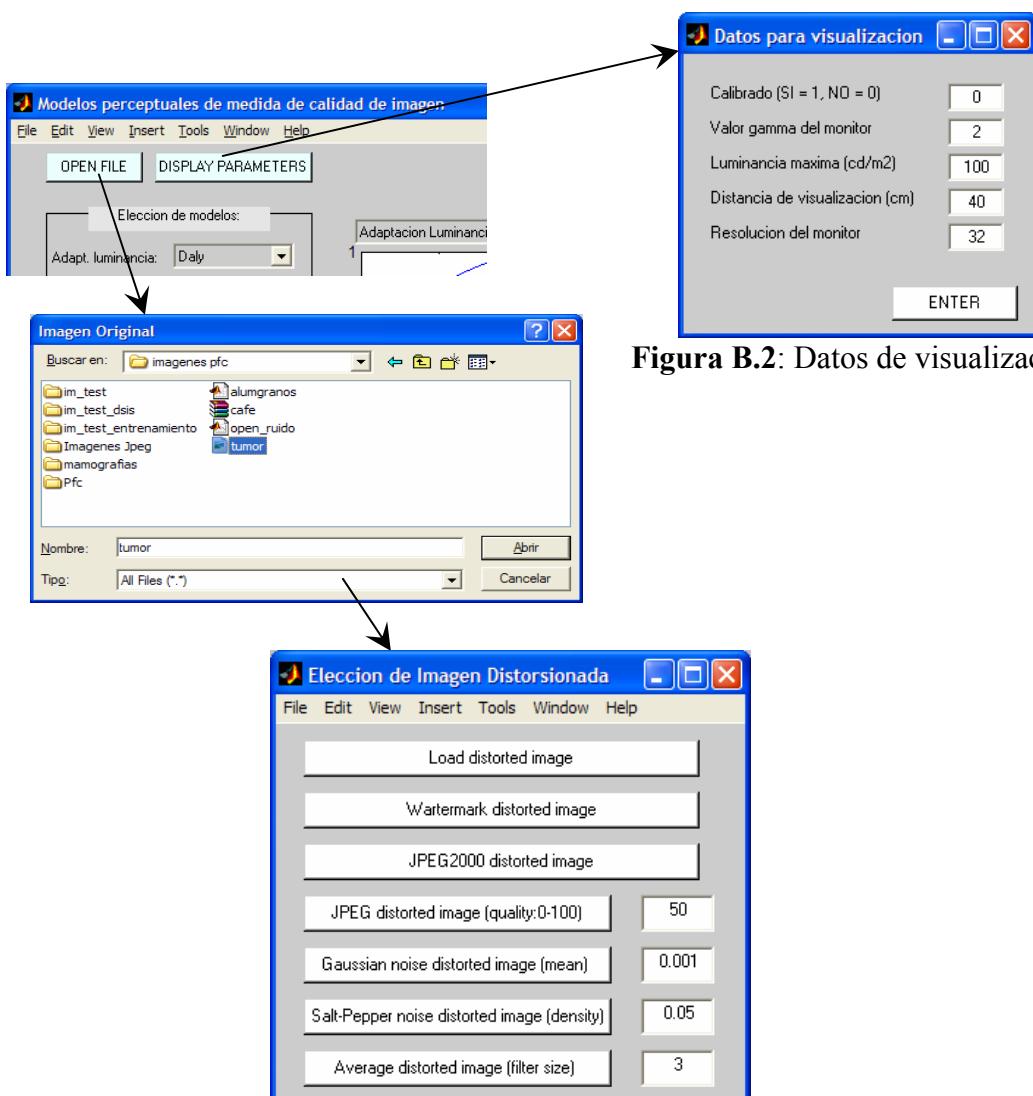


Figura B.2: Datos de visualización.

Figura B.3: Elección imagen distorsionada.

- La **carga de imágenes** es la que se muestra en la figura anterior. Merecen especial mención las diferentes opciones que se proponen para la imagen distorsionada (figura B.3). La primera consiste en cargar una imagen de error que ya existe para la imagen original seleccionada previamente. Las demás permiten construir en ese momento la imagen distorsionada. Para los cuatro últimos botones de la figura sólo habrá que introducir un parámetro que en cada caso es indicativo de la distorsión a la que se somete a la imagen original (calidad de imagen entre 0 y 100, media, densidad y tamaño del filtro). En el caso de que se quiera realizar una marca de agua sobre la imagen original o una compresión con JPEG2000 se van a abrir las ventanas de la figura B.4 y B.5 respectivamente. Para el marcado de agua se dispone del material proporcionado por el Prof. Fernando Pérez de la ETS de Ing. De Telecomunicación de Vigo y por tanto, solo se ha tenido que construir esta ventana para la introducción de los parámetros. La imagen original se guarda en el fichero ‘im.tiff’ siempre que se utilice dentro del interfaz que se está describiendo.

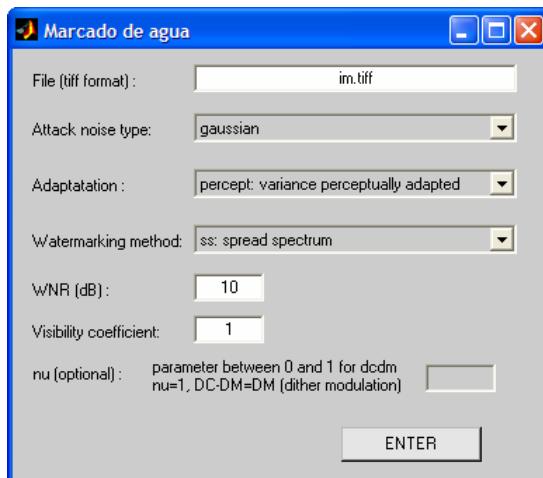


Figura B.4: Marcado de agua en la imagen distorsionada.

En la compresión con JPEG2000 se utiliza el entorno gráfico que muestra la figura B.5 y que ha sido proporcionado por Rafael Redondo del Instituto de Óptica del CSIC. Para cargar la imagen original se selecciona el menú imagen → nueva imagen y se abre el fichero ‘im.pgm’. Una vez se escogen los parámetros de compresión deseados y se aplican, al cerrar la ventana la imagen distorsionada se visualiza al lado de la de referencia como se mostró en la figura B.1. Para que este módulo y el anterior funcionen en la forma en la que se ha explicado es necesario que se especifiquen los

directorios donde se deja la imagen original. En el primer caso, la ruta del fichero ‘im.tiff’ se puede modificar el programa *im_watermark.m*. En el caso de JPEG2000 el directorio en el que se guarda la ‘im.pgm’ se incluye en el programa *im_jpeg2000.m*.

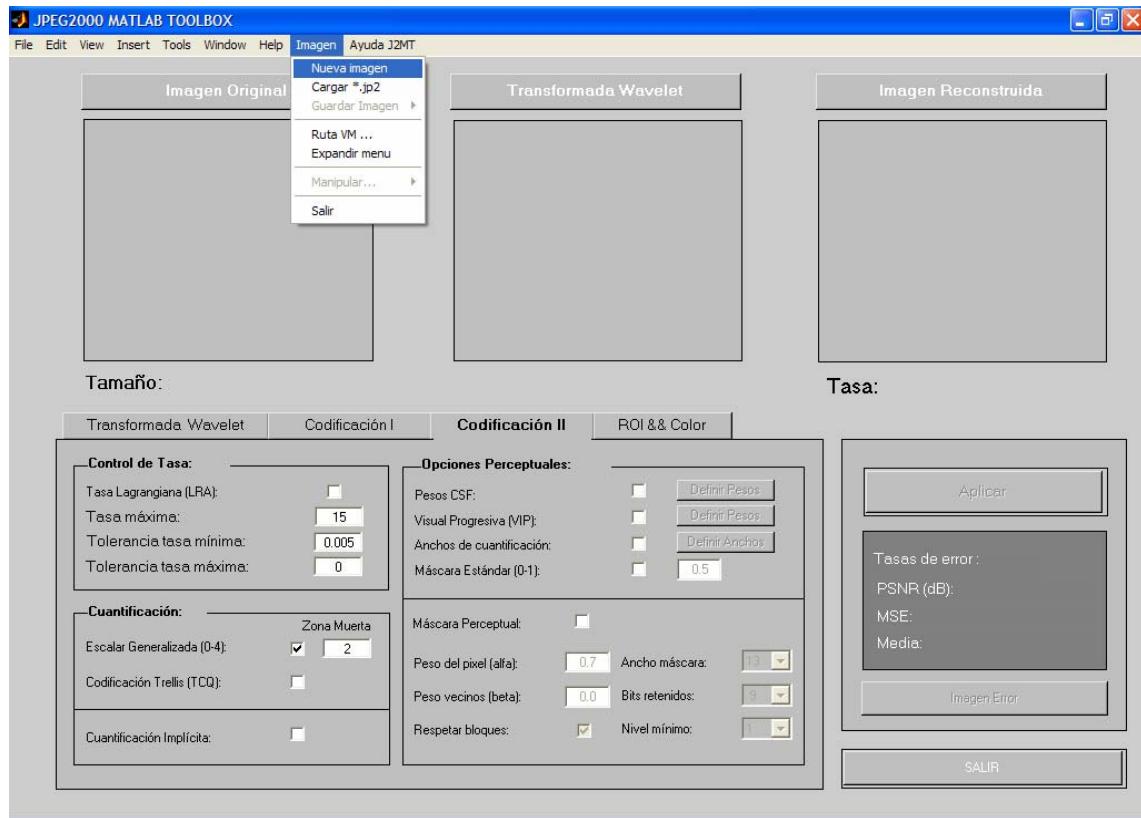


Figura B.5: JPEG2000 MATLAB TOOLBOX para creación imagen distorsionada.

2.- Elección de los distintos componentes del modelo de medida de calidad.

El marco con el título “Elección de modelos” proporciona todas las opciones de implementación de los distintos componentes del modelo visual: adaptación de luminancia, CSF, banco de filtros para la descomposición de la imagen y cálculo de contrastes. Las figuras a la derecha de este marco aportan información visual de la selección hecha para cada componente y en el último caso, en el desplegable “Opciones Visualización” se incluyen las imágenes de bandas, contrastes y umbrales de detección.

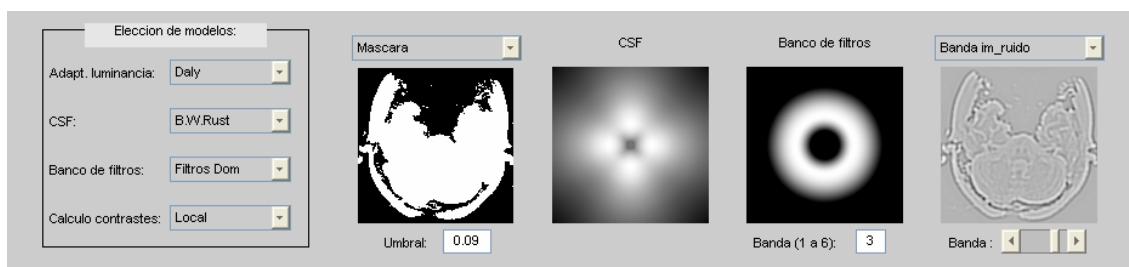
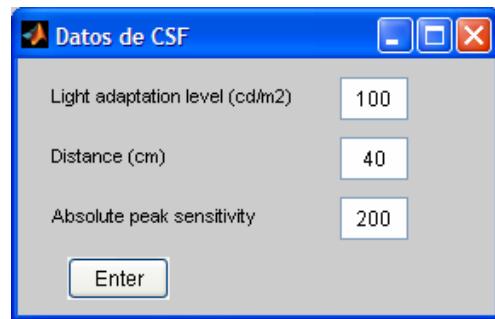


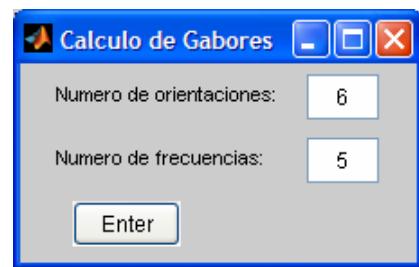
Figura B.6: Elección de los componentes del modelo.

- En el caso de la **adaptación de luminancia**, al elegir un modelo concreto (ninguna, Daly, raíz cúbica o logaritmo) lo que se va a mostrar en la figura es la curva que se está empleando para la conversión. Para ver el efecto visual sobre la imagen original o sobre la distorsionada se debe seleccionar la opción en el desplegable que hay sobre la figura. Además, hay una posibilidad más dentro de la adaptación de luminancia que es la que se ha mostrado en la figura B.6. Se puede utilizar una máscara para que la adaptación de luminancia se efectúe sólo sobre aquellas zonas en las que haya señal. Variando el valor del umbral situado bajo la figura se obtienen distintos mapas de umbral hasta que se consigue el deseado. De esta forma la adaptación se lleva a cabo sólo en las regiones en blanco del mapa que se ha logrado mediante este proceso. El orden para que esto funcione según se ha explicado consiste en seleccionar primero la adaptación de luminancia y luego la máscara. Si después de obtener la máscara se volviese a seleccionar la adaptación de luminancia el cálculo de la máscara dejaría de tener efecto.

- Para la construcción de la CSF, además de elegir el modelo deseado, será necesario introducir una serie de parámetros que definen la forma de la misma (ver apartado 4.3). Así, al seleccionar tanto el modelo de Daly, como el de Barten saldrá la figura B.7 para la introducción de los datos. En el caso de la CSF de Rust sólo será necesario introducir la distancia.

**Figura B.7:** Datos de CSF

- La elección del **banco de filtros** contempla tres opciones (ver apartado 4.4). En el caso de seleccionar un banco de filtro con Gabores se va a pedir que se indique el número de canales de orientación y frecuencia con los que se desea construir el mismo, figura B.8. Para los filtros Cortex y los filtros Anillo el número de canales es fijo. En todos ellos el canal que se muestra en la figura es el que viene indicado bajo la misma. Al ir cambiando el número de la banda y orientación (en caso de que exista) se van visualizando todos los canales posibles. Números mayores en banda indican frecuencias menores.

**Figura B.8:** Datos filtros Gabor.

- Finalmente, como ya se ha comentado, bajo el desplegable “**Opciones Visualización**” se han situado las opciones de visualización de las bandas de imagen y las bandas de contraste tanto para la imagen original como para la distorsionada. También es posible mostrar los mapas de elevación del umbral tanto para el enmascaramiento con CSF como filtro o por pesos (ver apartado 4.6). Para visualizar dentro de las diferentes opciones mencionadas los distintos canales basta con mover las barras situadas bajo la imagen tanto en banda como en orientación (en caso de que exista).

3.- Mapas e índices de medida de calidad.

- En la interfaz están representados, además, los diferentes modelos de imagen objetivos según la clasificación hecha en el Capítulo 3. Los **índices de error** se calculan al cargarse en la interfaz las imágenes de referencia y distorsionada. Los índices de medida de calidad perceptible se calculan en base a las opciones elegidas dentro de la “Elección de modelos”, con los dos esquemas básicos desarrollados según la utilización de la CSF (ver figuras 5.3, 5.4, C.1 y C.2). Finalmente, se pueden obtener los índices para los modelos basados en la distribución de Wigner y los basados en la denominada “similaridad estructural” que comparan simultáneamente la luminancia, el contraste y la estructura.

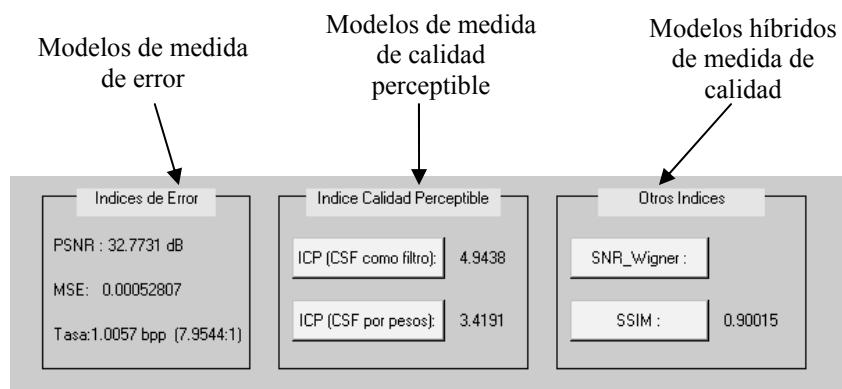


Figura B.9: Índices.

- Los diferentes **mapas de error** se sitúan en la imagen de la esquina inferior derecha en la figura B.1. En total son seis los mapas que pueden visualizarse. Al pulsar el botón “ICP (CSF como filtro)” el mapa que se mostraría sería el de probabilidad de error, pero también se calcula el de visibilidad del error (ver apartado 4.7). Para la “ICP (CSF por pesos)” sucede de igual forma. Los dos restantes se corresponden con los botones de “SSIM” y “SNR_Wigner”. En la figura B.10 se muestran las distintas posibilidades que se han comentado.

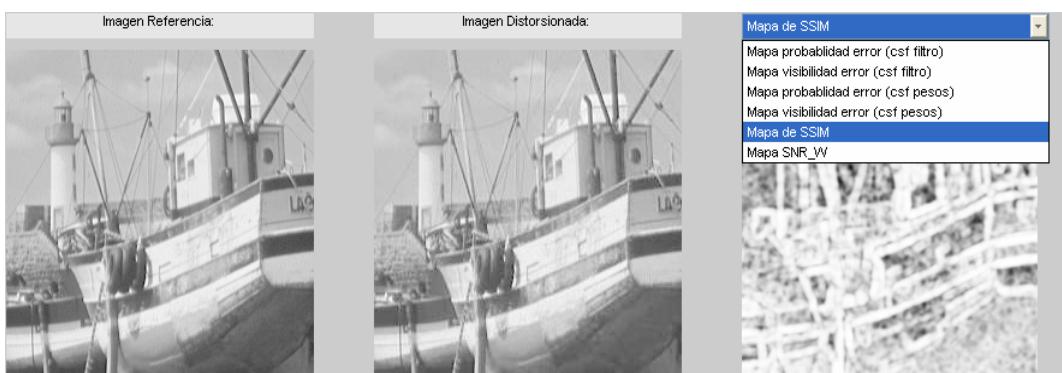


Figura B.10: Mapas.

4.- Otras funcionalidades.

Una vez efectuados los cálculos deseados para un modelo e imagen concretos siguiendo los pasos que se han descrito se plantean tres posibilidades.

- Se puede cerrar el entorno de usuario mediante el botón “Close”, con lo que se cerrarían todas las figuras y se limpiarían todas las variables.
- Se puede volver a diseñar un nuevo modelo de calidad (adaptación de luminancia, CSF, etc.) para las mismas imágenes pulsando “Refresh”.
- Se pueden cargar nuevas imágenes mediante el botón “Open”, comenzando de nuevo el diseño del modelo para nuevas imágenes.

Por último, destacar que al pulsar sobre cualquier figura de la interfaz con el botón derecho del ratón se ofrece la posibilidad de ver esa imagen en una nueva figura que se abre con el tamaño real de la misma, como se observa en B.11. Esto resulta de gran utilidad si se quieren comparar visualmente, por ejemplo, la CSF que se obtiene con el modelo de Daly y la que se obtiene con Barten. Como ambas tienen la misma ubicación dentro del interfaz, al calcular una después de la otra la primera se eliminaría. Esta opción permite conservar la primera en una figura independiente y hacer lo mismo con la segunda para compararlas en su tamaño real y así, con cualquier figura del entorno gráfico que resulte de interés.

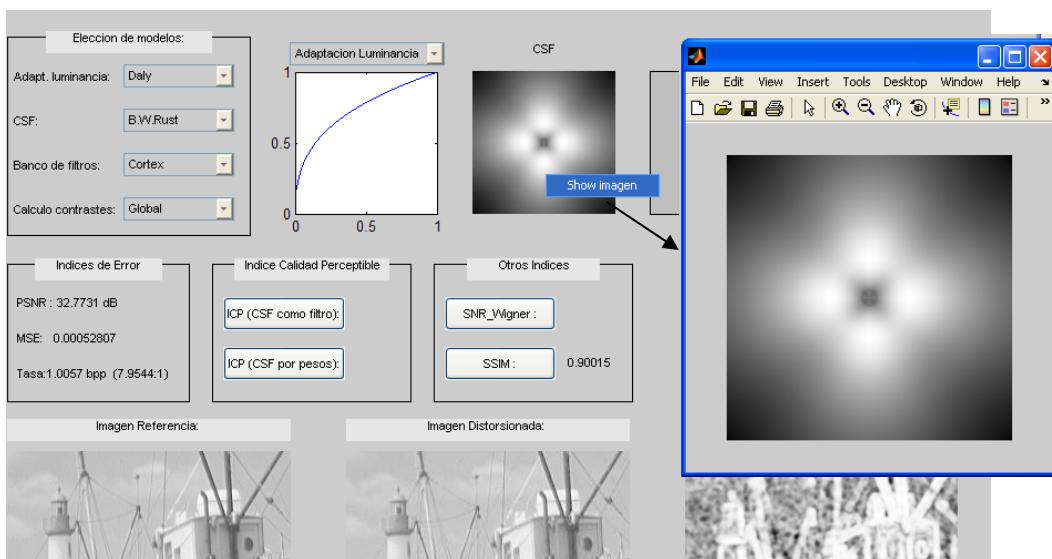


Figura B.11: Mostrar imagen.

ANEXO C: Manual de programas básicos.

A continuación se muestran las funciones MATLAB principales sobre las que se soporta la implementación de modelos visuales y que se corresponden con la descripción que se ha desarrollado previamente en el Capítulo 4.

➤ *display_parameters*.

Ventana para la carga de los parámetros del monitor con el que se trabaja: calibrado del monitor, gamma, resolución del monitor en pix/cm., luminancia máxima en cd/m² y distancia de visualización en cm. Si no se introducen valores de entrada va a guardar por defecto los que se han calculado para el monitor del laboratorio, que son los que se muestran. El directorio en el que se cargan se define dentro de este programa y es “C:\MATLAB7\work\disp_param”.

➤ *out_values = calibrar (in_values)*.

Función gamma de calibrado para el monitor. Se utiliza siempre que se van a mostrar imágenes de cualquier tipo en la interfaz de usuario. Los valores de entrada (*in_values*) son los niveles de gris de la imagen y los de salida (*out_values*) son éstos corregidos para obtener una función lineal en la visualización de la imagen. Esta función utiliza internamente los parámetros globales *calibrado* (igual a 1 la función es transparente), *lmax* (luminancia máxima) y *gamma*. Estos valores a introducir según el monitor con el que se trabaje se cargan mediante *display_parameters* y de no ser así, se toman por defecto los del monitor del laboratorio.

➤ *adap_lum = adap_luminancia (imag, modelo, máscara)*.

Función para implementar los distintos modelos de adaptación de luminancia según el valor de entrada de *modelo*: 1 no hay adaptación de luminancia, 2 de Daly, 3 raíz cúbica y 4 logarítmica. La imagen de entrada *imag* se convierte internamente a valores de luminancia para cada píxel. Si se utiliza una máscara la adaptación de luminancia se efectuará para aquellas zonas en el que el valor de la máscara es uno.

➤ $\text{CSFunction} = \text{CSF} (\text{tipo}, \text{tam}, r0, \text{lum}, d)$.

Genera la función de sensibilidad al contraste en 2D para los parámetros de entrada: *tipo* (1: Rust, 2: Daly, 3: Barten), tamaño de la imagen sobre la que se aplica la CSF, resolución del monitor, luminancia y distancia de visualización. Los tres últimos son opcionales ya que lo normal va a ser que se tomen los definidos globalmente en *display_parameters*. Además, para el cálculo de CSF se harán llamadas internamente a las siguientes funciones: *csfor* (para el tipo 1, Rust) y *sens* (para el 2 y 3, Daly y Barten).

➤ $\text{csf_mean} = \text{csf_pesos} (\text{csf}, \text{tipo_filtro}, \text{zona})$.

Calcula el valor de la CSF en cada banda como una media. *Csf_mean* es una matriz de $(N_f+1) \times N_o$ o un vector de N_f+1 con el valor de la CSF medio para cada canal. Recibe como entrada la función *csf* sobre la que se calcula, el tipo de filtro que se va a utilizar (*tipo_filtro* = 1 si filtros Cortex, 2 si filtros Gabor y 3 si filtros anillo) y *zona* serán los distintos filtros según el caso.

➤ $[\text{coefRose}, \text{roseta}] = \text{gabores} (N_f, N_o, \text{sizefft})$.

Función para el cálculo de los filtros de Gabor. Toma como parámetros de entrada el número de escalas N_f (el paso bajo se corresponde con N_f+1), el número de orientaciones N_o y el tamaño de la FFT de la imagen. *CoefRose* es una matriz 4D donde se guarda la familia de filtros Gabor generada. *Rosetta* es una matriz 2D que contiene la roseta o suma de los filtros Gabor en el dominio de frecuencia al cuadrado. La función gabores se limita a llamar a la función *rossette*, implementada por Sylvain Fischer en el instituto de Óptica del CSIC en la elaboración de su PFC para la generación del banco de filtros.

➤ $[\text{cortexband}, \text{dom}, \text{fan}] = \text{cortex} (\text{sizefft}, \text{tipo})$.

Función para el cálculo de la transformada Cortex. Utiliza como parámetro de entrada el tamaño de la FFT de la imagen (*sizefft*) y proporciona, además de 31 filtros Cortex (*cortexband*), 6 filtros de frecuencia (anillos *dom*) y 6 selectivos en orientación (filtros de abanico *fan*). Los valores de salida de los filtros que se proporcionan están en el dominio de frecuencia. La variable *tipo* con valor 1 (por defecto) calcula todo y 2 calcula solo los filtros *dom*.

- $[banda, imgrec] = \text{filtragabor}(imag, csf, coefRose)$.

Filtrado de la imagen (*imag*) en canales mediante el banco de filtros de Gabor definidos en *coefRose*. Si se introduce en *csf* el valor -1 no se utiliza CSF como filtro, sino, se debe introducir la función CSF deseada. Obtiene como resultado la matriz 4D *banda* que contiene las bandas de la imagen en el dominio espacial y la imagen reconstruida, *imgrec*.

- $[banda, bandafft, imgrec] = \text{filtradocortex}(imag, csf, cortexband)$.

Función que descompone la imagen (*imag*) en 30 bandas más el paso bajo. Se tendrá, por tanto, una matriz de 4D con las distintas bandas de la imagen o bien en frecuencia (*bandafft*) o en el dominio espacial (*banda*). El valor *imgrec* es la imagen reconstruida. Si se introduce en *csf* la matriz 2D con los valores de la función de sensibilidad al contraste se realizará un filtrado previo a la descomposición en canales, en caso contrario el valor de *csf* introducido es -1. *Cortexband* son los filtros de la transformada Cortex que son devueltos por la función *cortex*.

- $[banda, bandafft, imgrec] = \text{filtradom}(imag, csf, dom)$.

Función que descompone la imagen en 6 bandas mediante los filtros anillo. Es igual que la función anterior pero con una descomposición en canales que no es selectiva en orientación. El resultado, tanto en el dominio espacial (*banda*), como en frecuencia (*bandafft*), se almacenará en una matriz 3D.

- $im_contrastes = \text{cal_contraste}(bandas, tipo)$.

Función de cálculo de contrastes para una imagen descompuesta en *bandas* de frecuencia y orientación (matriz 4D) o solo en frecuencia (matriz 3D). Se obtendrá un conjunto de imágenes (*im_contrastes*) que representan el contraste en las distintas bandas en el dominio espacial y con las distintas definiciones de contraste propuestas según el valor elegido para *tipo*: 0 (global), 1 (local), 2 (Peli) y 3 (Lubin).

- $umbral = \text{thresh_elevation}(im_contrastes)$.

Función para cálculo de enmascaramiento como una serie de umbrales de elevación (*umbral*), según la ecuación proporcionada por Daly (ver 4.6.1). Toma como entrada *im_contrastes* que son los valores de contraste de la imagen en bandas normalizados previamente por la CSF.

- *umbral_p = thresh_elevation_p (csf_mean, img, im_contrastes).*

Igual que el caso anterior pero con la utilización de la CSF como una serie de pesos para el cálculo de los mapas de elevación de umbral (ver 4.6.2). Los parámetros de entrada son *csf_mean* (valor de la CSF medio para cada canal), *img* (imagen de entrada sobre la que se calculan umbrales) y *im_contrastes* (imágenes de contraste para la imagen *img* en cada canal).

- *mutualmasking = mutual_thres (umbral1, umbral2).*

Obtiene el umbral de elevación mutuo considerando la imagen original e imagen ruidosa. El resultado (*mutualmasking*) será el mínimo de los umbrales de elevación (ver *thresh_elevation* y *thresh_elevation_p*) calculado para cada una de las entradas, que son las matrices 3D o 4D con los mapas de enmascaramiento para la imagen de referencia (*umbral1*) y de la imagen distorsionada (*umbral2*).

- *[VMs, PMs, PMs_sig, T_VM, T_PM, T_PMs_sig, ICP, ICP_1a5] = cal_error (im_contrastes1, im_contrastes2, TH).*

Función para el cálculo de los distintos mapas de visibilidad del error, de probabilidad e índices de calidad perceptible (ver 4.7). Recibe como entrada la matriz con los valores de la imagen original y la distorsionada en bandas (*im_contrastes1* y *im_contrastes2*) y la matriz con los mapas de enmascaramiento mutuo (*TH*). A la salida se obtienen los mapas de visibilidad del error (*VMs*), mapas de probabilidad del error (*PMs*), mapas de probabilidad del error sin signo (*PMs_sig*), mapa de visibilidad total del error (*T_VM*), mapa de probabilidad total del error (*T_PM*), mapa de probabilidad total del error sin signo (*T_PMs_sig*), índice de calidad perceptible (*ICP*), ICP con rango de 1 a 5 (*ICP_1a5*).

El esquema para la formación de un sistema de medida de calidad utilizando las funciones descritas se muestra en las figuras C1 y C2. La primera de ellas se corresponde con un sistema que utiliza la CSF como filtro, la segunda con un modelo de medida de calidad que usa la CSF como una serie de pesos. Estos esquemas sirven para ver como se relacionan las distintas funciones entre sí y el orden de las mismas dentro de un sistema de medida de calidad.

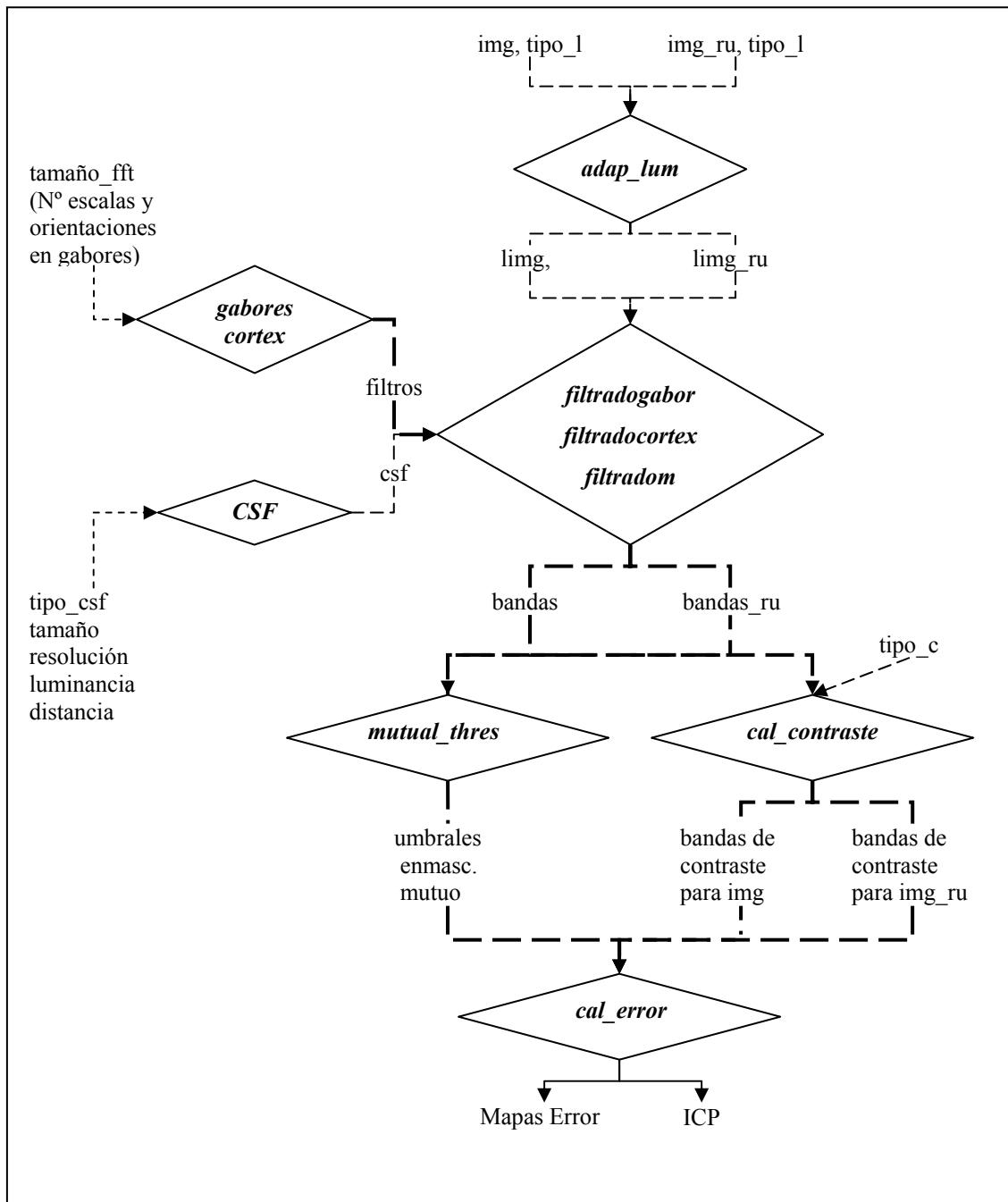


Figura C.1: Esquema de funciones para modelo con CSF como filtro.

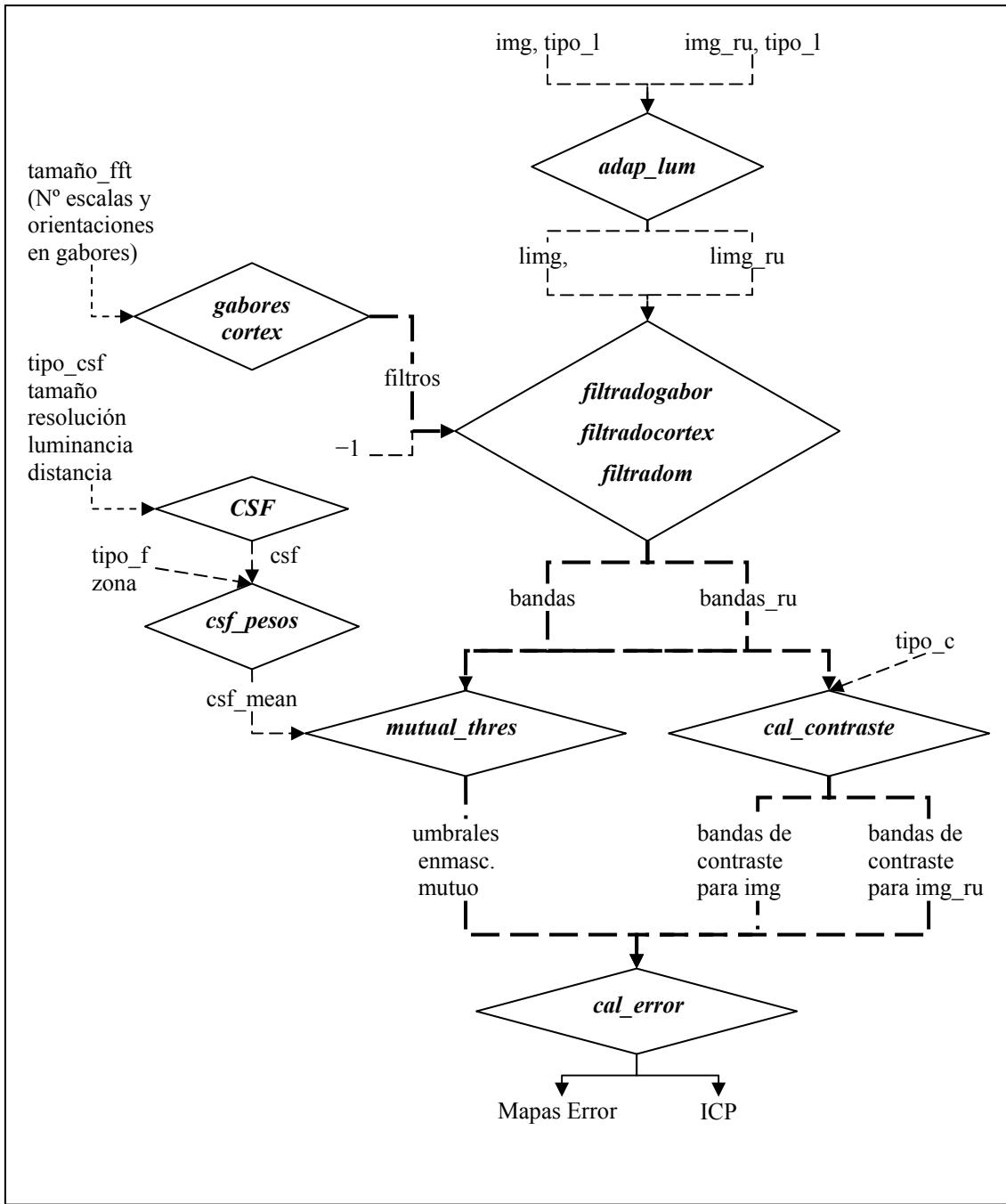


Figura C.2: Esquema de funciones para modelo con CSF como pesos.

Finalmente, en la figura C.3 se muestran las funciones más importantes para el manejo del entorno gráfico, de forma que resulte fácil saber cómo y dónde modificar para añadir nuevas funcionalidades a la interfaz. La función `interfaz_usuario.m` es donde se definen todos los elementos del entorno. El resto de funciones que aparecen son las que van a ejecutarse al pulsar cada botón, desplegable, barra, etc.

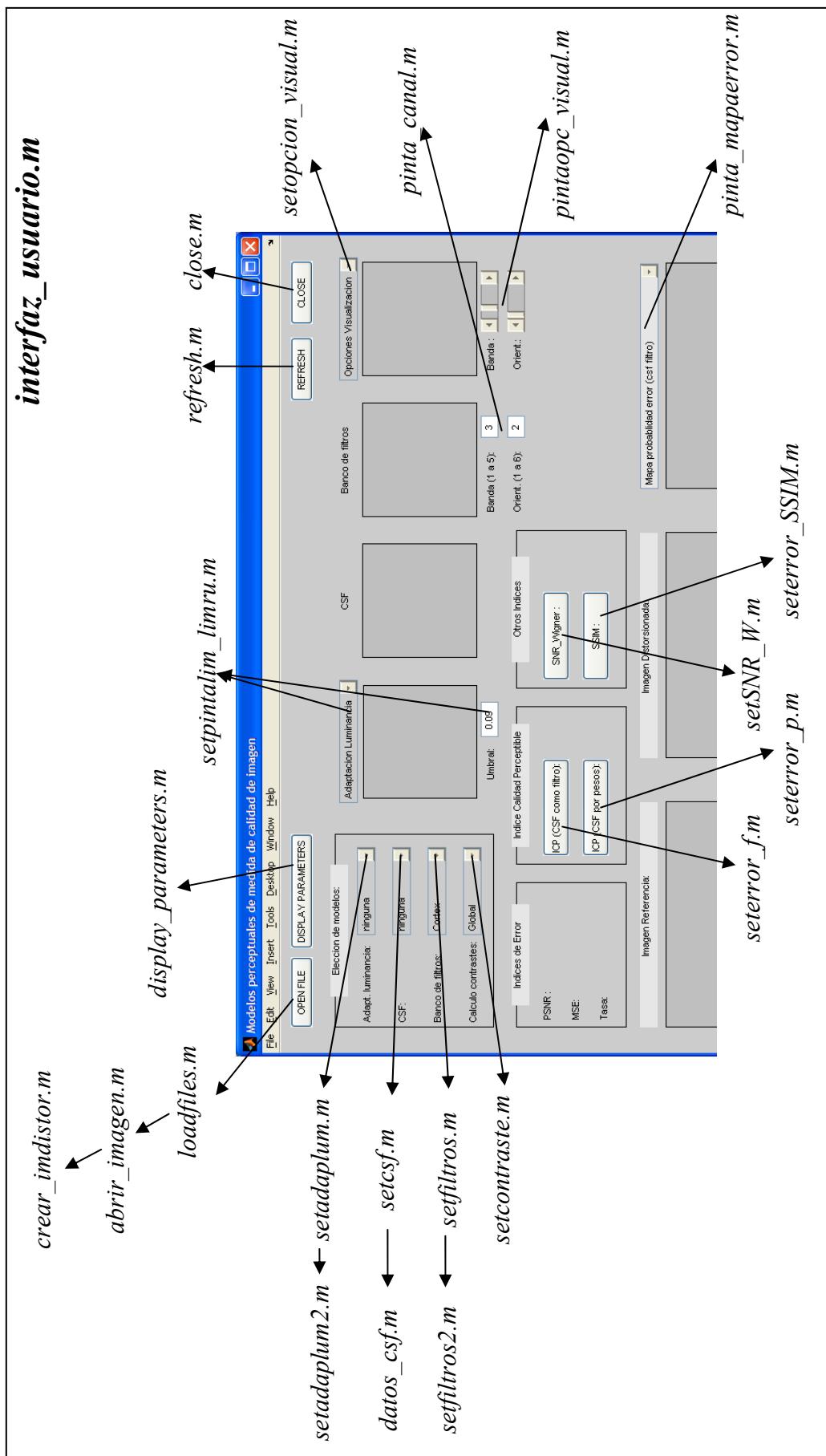


Figura C.3: Esquema de funciones ejecutadas mediante los distintos controles del entorno gráfico.

BIBLIOGRAFÍA

- [1] ARTIGAS, J.M., et al. “Óptica Fisiológica. Psicofísica de la Visión”. Interamericana McGraw-Hill (ed.), 1995.
- [2] AZNAR, J.A. “Análisis multiescala y multiorientación de imágenes mediante un banco de filtros de Gabor-2D”. Dpto. de Psicología Básica. Facultad de Psicología. Universidad de Barcelona, 2000.
- [3] BEGHDADI, A.; PESQUET, B. “A new image distortion measure based on wavelet decomposition”. Proc. IEEE ISSPA2003, 1-4 Julio 2003, París.
- [4] BEGHDADI, A.; IORDACHE, R. “Image Quality Assessment Using the Joint Space/Spatial-Frequency Representation”, EURASIP J. on Applied Signal Processing (en revisión).
- [5] BRAINARD, D.H.; PELLI, D.G.; ROBSON, T. “Display Characterization”. Encyclopedia of Imaging Science and Technology. J. Hornak (ed.), Wiley. pp 172-188.
- [6] CADIK, M.; SLAVIK, P. “Evaluation of two Principal Approaches to Objective Image Quality Assessment”. Eight International Conference on Information Visualisation (IV’04), pp. 513-518, 2004.
- [7] DALY, S. “The visible difference predictor: An algorithm for the assessment of image fidelity”. En A. B. Watson (ed.), Digital Images and Human Vision, pp. 179-206. MIT Press, Cambridge, MA, 1993.
- [8] DAUGMAN, J.G. “Two dimensional spectral analysis of cortical receptive fields profiles”, Vision Research, 20: 847-856, 1980.

- [9] DAUPHIN, G.; BEGHDADI, A.; LESEGNO, P. V. “A local directional band limited contrast”. Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium, vol.2, pp 197- 200, 2003.
- [10] ECKERT, M.P.; BRADLEY, A.P. “Perceptual quality metrics applied to still image compression”, Signal Proc., 70:177-200, 1998.
- [11] ERIKSSON, R.; ANDREN, B.; BRUNNSTRÖM, K. “Modelling the perception of digital images: A performance study.” Proc. SPIE Conf. Human Vision and Electronic Imaging III. San José, Cal., pp. 88-97, 1998.
- [12] FISCHER, S. “Modelos perceptuales en la cuantificación y codificación de imágenes”. Proyecto Fin de Carrera. Dpto. de Ingeniería Electrónica. ETSIT. Junio 2000.
- [13] GABOR, D. “Theory of communication”, J.IEE (London), 93(III), pp. 429-457, 1946.
- [14] GIROD, B. “What’s wrong with mean-square error.” En A. B. Watson (ed.), Digital Images and Human Vision, pp. 207-220. MIT Press, Cambridge, MA, 1993.
- [15] HORMIGO, J.; CRISTOBAL, G. “Image segmentation using the Wigner-Ville distribution”, Advances in imaging and electron physics. 2003.
- [16] HUBEL, D. H. “Ojo, cerebro y visión.” Servicio de Publicaciones, Universidad de Murcia, 2000.
- [17] ITU. “Methodology for the subjective assessment of the quality of television pictures”. Recommendation ITU-R BT.500-10, 2000.
- [18] LI, B.; MEYER, G.W.; KLASSEN,R.V. “A comparison of two image quality Models”, Computer Science Department, University of Oregon, Eugene, OR 97403 Color and Digital Imaging Systems, Xerox Corporation, Webster NY 14580

- [19] LUBIN, J. “The use of psychophysical data and models in the analysis of display performance.” En A. B. Watson (Ed.), Digital Images and Human Vision, pp. 163-178. MIT Press, Cambridge, MA, 1993.
- [20] MARCELA, S. “Mathematical description of the responses of simple cortical cells”, Journal of the Optical Society of America A, 70(11):1296-1300, 1980.
- [21] MIYAHARA, M.; KOTANI, K.; ALGAZI, V.R. “Objective picture quality scale (PQS) for image coding”, IEEE Transactions on Communications, vol. 46, no. 9, September 1998.
- [22] NADENAU, M.J. “Integration of human color vision models into high quality image compression”. PhD thesis, Signal Processing, Laboratory, Swiss Federal Institute of Technology, Lausanne, Nov 2000.
- [23] PELI, E. “Contrast in complex images”. Journal of the Optical Society of America A, 7(10):2032-2040, Octubre 1990.
- [24] PELI, E. “In search of a contrast metric: Matching the perceived contrast of gabor patches at different phases and bandwidths.” Vision Research, 37(23): 3217-3224, 1997.
- [25] OH, J. et al. “A Multistage Perceptual Quality Assessment for Compressed Digital Angiogram Images”. IEEE Transactions on medical imaging, vol. 20, no 12, Diciembre 2001.
- [26] OSBERGER, W. “Perceptual Vision Models for Picture Quality Assessment and Compression Applications”. Ph.D. dissertation, Sch. Elect. Electron. Syst. Eng., Queensland Univ. Technol., Queensland, Australia, 1999.
- [27] REED, T.; WECHSLER, H. “Spatial/spatial-frequency representations for image segmentation and grouping”, Image and vision computing, vol. 9, no. 3, 1991.

- [28] RUST, B.W. "A new representation of the contrast sensitivity function for human vision." In Proc. Int. Conf. on Imaging Science, Systems and Technology, pp. 1-15, 1997.
- [29] TAYLOR, C. "Image Quality Assessment via a Human Visual System Model", Ph.D. Dissertation, Purdue University, Agosto 1998.
- [30] WANG, Z.; BOVIK, A.C. "A universal image quality index", IEEE Signal Processing Letters, vol. 9, pp. 81-84, Marzo 2002.
- [31] WANG, Z., et al., "Image quality assessment: from error measurement to structural similarity", IEEE Transactions on image processing, vol. 13, no. 4, pp. 600-612, 2004.
- [32] WATSON, A.B. "The Cortex Transform: Rapid Computation of Simulated Neural Images", Computer Vision, Graphics, and Image Processing, 39, 311-327, 1987.
- [33] WATSON, A.B.; BORTHWICK, R.; TAYLOR, M. "Image quality and entropy masking", SPIE Proceedings, vol. 3016, paper 1 (1997), 358-371
- [34] WESTEN, S.J.P.; LANGENDIJK, R.L.; BIEMOND, J. "Perceptual image quality based on multiple channel HSV model", Proc. ICASSP, 1995, pp.2351-2354.
- [35] WINKLER, S. "Vision models and quality metrics for image processing applications". PhD thesis, Ecole Polytechnique Fédérale de Lausanne. Dic. 2000.
- [36] <http://www.cns.nyu.edu/~lcv/ssim/>