



Modelos no-lineales de enmascaramiento en JPEG2000 para la codificación de imágenes en color

Memoria para optar al Diploma de Estudios Avanzados
del Doctorado de Informática y Matemática Computacional de la
Universitat de València.

Yolanda Navarro Herráiz

`yolanda.navarro@uv.es`

Dirigido por
Jesús Malo López

Septiembre 2005

ESTHER DE VES CUENCA (tutora en el programa de Doctorado),
profesora del Departament de Informàtica de la Universitat de València y

JESÚS MALO LÓPEZ (director del trabajo de investigación), profesor del
Departament d'Òptica de la Universitat de València

CERTIFICAN:

Que la presente memoria titulada *Modelos no-lineales de enmascaramiento en JPEG2000 para la codificación de imágenes en color*, ha sido realizada bajo su dirección y constituye su Trabajo de Investigación para optar al Diploma de Estudios Avanzados del Doctorado de Informática y Matemática Computacional.

ESTHER DE VES CUENCA

JESÚS MALO LÓPEZ

Agradecimientos

A mis padres sin los cuales esta aventura no habría sido posible, por su ánimo y apoyo para seguir avanzando.

A Quino por estar siempre a mi lado tanto en los buenos como en los malos momentos pasados estos años.

A Jesús Malo López por lo mucho que me has enseñado así como por los buenos momentos pasados en la terraza de físicas.

Gracias también a todos los amig@s que han estado ahí.

Índice general

1. Introducción	1
2. Modelo no-lineal de respuesta en el córtex: <i>Normalización Divisiva</i>	4
3. El estándar JPEG2000	9
3.1. Transformación cromática	10
3.2. Transformación wavelet	13
3.3. Transformación al dominio de la respuesta	17
3.4. Modelos perceptuales <i>simplificados</i>	17
3.4.1. Modelo lineal	17
3.4.2. Modelo no lineal dependiente del punto	18
3.4.3. Modelo no lineal: <i>Normalización Divisiva simplificada</i>	19
3.5. Diseño del cuantizador	19
4. Beneficios estadísticos de la <i>Normalización Divisiva</i> en la Codificación de la Transformada	22
4.1. Independencia entre coeficientes wavelet	22
4.2. Probabilidad condicionada e Información mutua	24
5. Inversión de la <i>Normalización Divisiva</i>	29
5.1. El problema de la inversión	29
5.2. Solución: el método de expansión en serie	30
5.3. Condición de invertibilidad	31
6. Resultados	33
6.1. Imágenes codificadas	33
6.2. Curvas Rate-Distorsion	34
7. Conclusiones y trabajo de futuro	38

Capítulo 1

Introducción

Las imágenes naturales tienen unas características propias que las hacen especiales, las comentaremos muy brevemente.

En las imágenes naturales existen dependencias estadísticas y perceptuales entre los píxeles de la imagen. Es decir, existe una gran relación entre los valores de luminancia de píxeles vecinos, pudiendo predecir el valor de un pixel a partir del de un vecino. Lo mismo ocurre con las componentes de color, donde la sensación de un pixel de color depende de los colores de su entorno.

Como la correlación entre píxeles y la varianza (anchura de la distribución), son tan altas. Serían necesarios muchos representantes para cuantizar escalarmente la imagen en el dominio espacial. Resulta evidente la necesidad de trabajar en un dominio donde los datos no estén correlacionados, para de este modo aprovechar más eficientemente el número de representantes empleados por el cuantizador.

El presente trabajo se ha centrado en la búsqueda de una representación eficiente de las imágenes en color, donde se hayan eliminado las dependencias estadísticas y perceptuales existentes entre los píxeles de la imagen, [1], [2]. Para su posterior compresión, empleando un cuantizador escalar uniforme dentro de un modelo como el de la *Codificación de la Transformada* (o Transform Coding) [3].

El concepto de la *Codificación de la transformada* consiste en un esquema de compresión de imágenes que ha demostrado ser efectivo, y es la base de todos los estándares de compresión con pérdidas que se emplean en la actualidad. En la figura 1.1 se ilustra la estructura en bloques de este esquema de compresión, en las configuraciones de Codificador y Decodificador.

El funcionamiento básico del *Codificador* consiste en la aplicación sobre la imagen de una transformación general \mathbf{T} , es decir puede ser tanto lineal como no-lineal, cuyo objetivo es la eliminación de las relaciones existentes

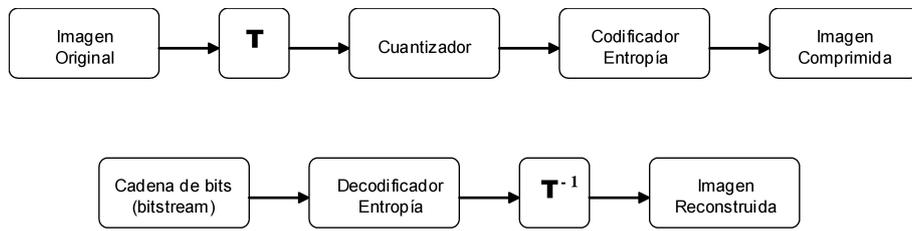


Figura 1.1: *Diagrama de bloques de un codificador (arriba) y decodificador (abajo).*

entre los píxeles junto con una acumulación de la energía en un número reducido de los coeficientes transformados. Para su posterior cuantización y codificación, empleando modelos perceptuales. De este modo, se asigna un mayor número de bits a los coeficientes con mayor relevancia, ya que la transformada \mathbf{T} los pesa de manera diferente.

El *decodificador* realiza el proceso inverso al codificador, llevando a cabo los mismos procesos pero en orden inverso. Convirtiendo el flujo de bits comprimido en una nueva imagen reconstruida. La imagen original ya no podrá recuperarse, se ha eliminado información de manera irreversible con el proceso de cuantización, por tanto a la salida se obtendrá una versión de la misma donde se ha introducido una cierta distorsión.

Organización de la memoria

El presente trabajo de investigación se ha organizado como sigue:

En el capítulo 2 se describen las características del modelo perceptual más potente de que se dispone en la actualidad, la *Normalización Divisiva* [4], [5], [6]. En él se incluye mecanismos de control de ganancia adaptativos, que están inspirados en la propia normalización realizada por las neuronas corticales de V1. Por tanto, esta transformación proporciona una buena caracterización de las propiedades de la respuesta (no-lineal) del sistema visual.

En el capítulo 3 se describe la metodología aplicada por el estándar JPEG2000 [7], [8], para la compresión de imágenes en color, siguiendo un esquema similar al de la codificación de la transformada. A fin de conseguir la mejor representación de las imágenes, es decir eliminar las dependencias estadísticas y perceptuales existentes entre los píxeles, el estándar incluye diferentes modelos perceptuales de visión humana: lineal, no-lineal dependiente del punto y *Normalización Divisiva* simplificada.

En el capítulo 4 se presentan los beneficios estadísticos conseguidos, en el aumento en la independencia estadística de los coeficientes transformados,

tras la aplicación de la *Normalización Divisiva* [2]. Lo que justifica la inclusión de esta transformada dentro del esquema de compresión de imágenes del estándar JPEG2000.

En el capítulo 5 se describen los problemas de inversión que presenta la versión más general de la *Normalización Divisiva*. Razón por la que el estándar JPEG2000, incluye una versión simplificada de la misma [8].

En el trabajo se ha resuelto este problema, recurriendo a un método de inversión ágil y robusto basado en la expansión en serie. Que ya había sido aplicado con éxito, en el contexto de la transformada DCT por bloques [2], ahora lo hemos generalizado para la transformada Wavelet.

En el capítulo 6 se presentan los resultados obtenidos en las curvas de Rate-Distorsion (compresión-distorsión), tras la inclusión en el esquema de compresión de imágenes de la *Normalización Divisiva* en su versión más general, superando a los logrados por la versión más refinada del nuevo estándar JPEG2000 [8].

Para ilustrarlo se presentan unas imágenes codificadas empleando esta transformación, donde se aprecian las mejoras conseguidas en la tasa de compresión (35-40%) para la misma distorsión. Esto repercute en un aumento de la calidad final de la imagen, a la misma compresión.

Para finalizar en el capítulo 7 se presentan las conclusiones del presente trabajo de investigación, así como las posibilidades de futuro de la línea de investigación que se ha seguido.

La contribución del presente trabajo de investigación, ha sido la inclusión en el nuevo estándar JPEG2000, del modelo perceptual de la *Normalización Divisiva* en su versión más general. Gracias a que hemos resuelto el problema de su inversión, recurriendo al método de expansión en serie que ya se había aplicado con éxito en el contexto de la transformada DCT por bloques [2]. En este caso se ha generalizado la solución de las series de expansión, para la transformada Wavelet que es la empleada por el estándar JPEG2000.

El presente trabajo ha dado como resultado la publicación de los siguientes artículos:

- 'Gain control for the chromatic channels in JPEG2000' presentado en el 10º congreso de la Asociación Internacional de Color, Mayo 2005, [9].
- 'V1 non-linearities in JPEG2000' enviado a IEEE Signal Processing Letters, [10].

Capítulo 2

Modelo no-lineal de respuesta en el córtex: *Normalización Divisiva*

El sistema visual humano ha evolucionado hasta conseguir una completa adaptación a las características de las imágenes naturales [11], [12]. Para ello ha desarrollado complejos mecanismos de procesamiento de la información que le permiten un eficaz aprovechamiento de las señales a las que está expuesto.

En la literatura de ciencias de la visión [5], vemos como estos mecanismos de procesamiento de la información, plantean una estructura de etapas en cascada:

$$A \xrightarrow{T} a \xrightarrow{R} r \quad (2.1)$$

Sobre la imagen A definida en el dominio espacial, se aplica una primera transformación lineal T , que consiste en la actuación de un banco de filtros lineales pasa-banda. Éstos caracterizan a los campos receptivos del córtex visual primario (o area V1) y permiten extraer en a el contenido espacio-frecuencial de la imagen, puede tratarse de transformaciones del tipo DCT por bloques o Wavelet, [13], [14], [15], [5].

A continuación, se aplica una transformación no-lineal R , que normaliza la respuesta lineal del sistema por una combinación de la respuesta de otras neuronas, (similar a un mecanismo de control de ganancia).

La transformación conocida como *Normalización Divisiva*, incluye mecanismos de control de ganancia adaptativos, que están inspirados en la propia normalización que realizan las neuronas corticales de V1, [4], [5], [16], [17], [18]. Por tanto, esta transformación proporcionará una buena caracterización de las propiedades de la respuesta (no-lineal) del sistema visual.

El dominio de la respuesta es Euclídeo por definición, [5]. Y en los modelos del sistema visual humano *HSV*, se supone además uniformidad perceptual, es decir no existen interacciones perceptuales entre los coeficientes y por tanto las distorsiones de cada coeficiente contribuyen de igual forma en la distorsión global.

Como veremos en el capítulo 4, además de las ventajas perceptuales, la *Normalización Divisiva* tiene unas propiedades estadísticas que la hacen interesante en diferentes aplicaciones de procesado de imágenes. Con un interés especial en la codificación eficiente de imágenes: permite reducir las dependencias estadísticas (más complejas), que todavía quedan entre los coeficientes, después de la aplicación sobre la imagen de una transformación lineal (del tipo DCT por bloques o Wavelet).

Esto permite proponer un nuevo dominio de representación para la imagen, el dominio de frecuencia-local normalizado, útil para la compresión de imágenes dentro de un esquema como el de la codificación de la transformada. Ahora sí se puede emplear un cuantizador escalar uniforme [3], ya que el dominio de salida es perceptualmente Euclídeo y por tanto el diseño del cuantizador se simplifica enormemente.

El modelo de *Normalización Divisiva*, R , usado es básicamente un modelo de control de ganancia donde la energía de cada coeficiente transformado, a_{ik} (expresado en contraste [19]), se normaliza mediante una combinación de la energía de sus vecinos en escala, orientación y espacio, [5], [4], [17], [18]:

$$r_{ik} = \frac{\text{sgn}(a_{ik}) \cdot |\alpha_{ik} \cdot a_{ik}|^\gamma}{\beta_{ik} + \sum_{j=1}^{N_k^2} h_{ij} \cdot |\alpha_{jk} \cdot a_{jk}|^\gamma} \quad (2.2)$$

donde a_{ik} se refiere al elemento i del vector de coeficientes transformados a y $k = 1, 2, 3$, a los tres canales cromáticos que constituyen la imagen en color.

La matriz, h , define las interacciones o *enmascaramiento* entre los coeficientes de la transformada, a_{ik} , se asume una vecindad gaussiana [5], [17]:

$$h_{ij} = B_i \cdot e \left(-\frac{|s_i - s_j|^2}{\sigma_s^2} - \frac{|o_i - o_j|^2}{\sigma_o^2} - \frac{|x_i - x_j|^2}{\sigma_x^2} \right) \quad (2.3)$$

donde s_i , o_i y x_i son respectivamente la escala, orientación y posición espacial del coeficiente transformado i , y B_i es una constante de normalización que hace que $\sum_{j=1}^{N_k^2} h_{ij} = 1$.

Dada una imagen, A , de tamaño $N \times N$, el tamaño de los vectores α , β y r , es N^2 . Mientras que el tamaño de la matriz, h_{ij} , es $N^2 \times N^2$. Considerando estos tamaños, un patrón de interacción arbitrario en la matriz h ,

supondría unos cálculos de $h_{ij} \cdot |\alpha_{ik} \cdot a_{ik}|^\gamma$ muy costosos. Afortunadamente la naturaleza de las interacciones entre los coeficientes de la transformada wavelet a_{ik} , es local y por tanto la matriz h es poco densa o *sparse*. En la figura 2.1, se ilustra la vecindad gaussiana asumida por el modelo, así como la naturaleza *sparse* de la matriz h .

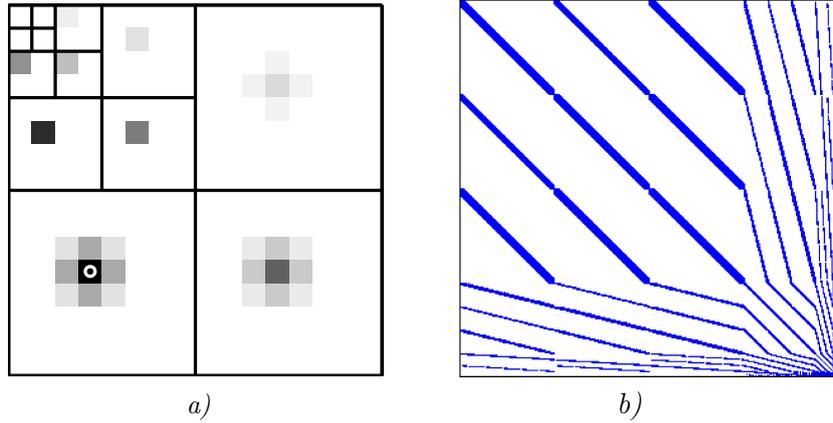


Figura 2.1: a) *Vecindad gaussiana asumida entre los coeficientes de diferentes escalas, orientaciones y posiciones espaciales, para un coeficiente particular de la transformada wavelet (marcado en círculo), que se incluye en la matriz h_{ij} . (Para el ejemplo se ha tomado $\sigma_{s_i} = 1$, $\sigma_{o_i} = 1$ y $\sigma_{x_i} = 0,4$, expresadas en grados). La transformada Wavelet se explicará con más detalle en el capítulo 3, ya que es la transformación escogida por el estándar JPEG2000 [8]. b) Representación de la matriz h . Se obtiene de la ordenación por filas, de la interacción de cada coeficiente consigo mismo y sus vecinos. Las frecuencias a su vez, se organizan de la más fina a la más baja. Debido a que las interacciones entre los coeficientes son de carácter local, el kernel de interacción es poco denso o *sparse*, los valores distintos de cero se han representado en azul.*

La *Normalización Divisiva* también incluye estos otros parámetros, α , β y γ , donde:

1. α_{ik} , es una función pasa banda similar a la CSF, ver figura 2.2. Aplica un peso selectivo a los coeficientes de cada subbanda y canal cromático o acromático.
2. β_{ik} , es un parámetro que ajusta los umbrales incrementales de contraste de las funciones base; es decir determina el contraste para el que se tiene máxima pendiente (ie. sensibilidad) de la curva de respuesta en cada frecuencia [5], [20].
3. γ es el exponente de excitación o inhibición, que controla la convexidad de la curva de respuesta.

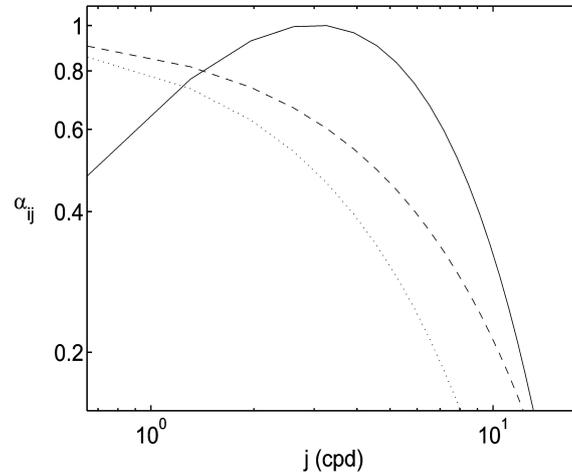


Figura 2.2: Pesos aplicados sobre los coeficientes, simulando una CSF acromática (línea continua) y cromática rojo-verde (línea discontinua) y azul-amarillo (línea punteada).

Estos parámetros pueden determinarse de diferentes maneras: *psicofísicas* [4], [17], *neurológicas* [5] y finalmente *estadísticas* [16].

En la figura 2.3 se muestra a modo ilustrativo, el comportamiento no-lineal de la respuesta. Cada respuesta r_{ik} , depende principalmente del coeficiente a_{ik} , aunque existen también contribuciones debidas a los vecinos, a través del kernel h_{ij} .

Las curvas se han calculado para el caso particular de fondo cero, es decir sin contar con el enmascaramiento producido sobre el coeficiente por los vecinos [5], [2]. Cabe destacar que, la forma sigmoide de la función de respuesta da lugar a unas propiedades de discriminación (pendiente de la respuesta) mayores para baja frecuencia y amplitud, justamente donde se concentran los valores de los coeficientes de las transformadas [3], [6]. Esto condicionará la posterior asignación de un mayor número de bits en la codificación, para estas zonas.

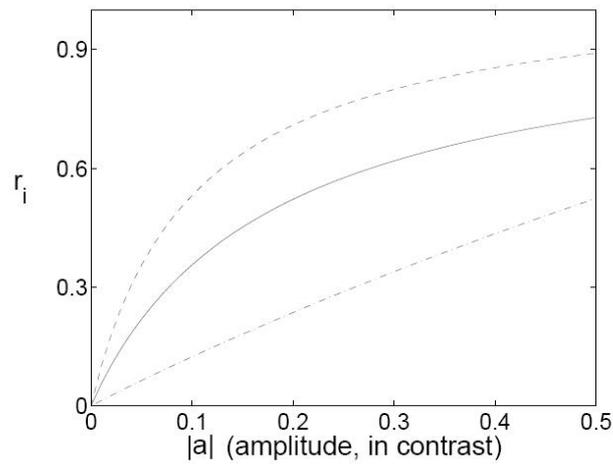


Figura 2.3: Ejemplos de la respuesta normalizada en función de la amplitud de los coeficientes (expresados en contraste [19]), para diferentes frecuencias (cpd, ciclos por grado): 4 cpd (línea continua), 8 cpd (línea discontinua) y 16 cpd (línea punteada). Los cálculos de este ejemplo ilustrativo, se han realizado en el contexto de la transformada DCT por bloques.

Capítulo 3

El estándar JPEG2000

Con el fin de proporcionar un estándar universal para la compresión de imágenes, *JPEG o Joint Photographic Expert Group* desarrolló un formato de almacenamiento para las imágenes digitales basado en estudios realizados sobre la percepción humana. El estándar JPEG [21], describe toda una familia de técnicas de compresión para imágenes naturales, tanto en escala de grises como en color (24 bits).

Para eliminar las redundancias estadísticas y perceptuales presentes en las imágenes naturales, JPEG incluye un conjunto de transformaciones que tienen como objetivo representar la imagen en un dominio de respuesta perceptualmente Euclídeo [5], [3] y así emplear un cuantizador escalar uniforme.

El estándar JPEG sigue un esquema de compresión similar al de la *codificación de la transformada*, que hemos descrito anteriormente. En particular [21], para la primera etapa T usa la transformada DCT por bloques y una aproximación lineal sencilla para la segunda etapa R . La asignación de bits se basa en las Funciones de Sensibilidad al Contraste (CSF): acromática y cromática. Empleando aproximaciones más precisas para la segunda etapa, [22], [23], [24], [25], es decir añadiendo no-linealidades, se mejoran los resultados obtenidos con el estándar JPEG.

A fin de conseguir la representación más adecuada para las imágenes, en el nuevo JPEG2000, [7], [8] se utilizan modelos actuales de procesamiento visual humano, que involucran dos etapas [4, 5]:

$$A_{ik} \xrightarrow{T} a_{ik} \xrightarrow{R} r_{ik} \quad (3.1)$$

Tomaremos como ejemplo una imagen en color, A_{ik} , de tamaño $N \times N$ donde $i = 1, \dots, N^2$ son sus componentes espaciales y $k = 1, 2, 3$, los canales del espacio cromático RGB. La imagen está definida en el dominio espacial, donde resulta evidente la existencia de interacciones estadísticas y perceptuales entre los píxeles de la imagen.

En la primera etapa, sobre cada componente i y para cada canal cromático k de la imagen A , se aplica la transformación *Wavelet* T , [7], que representará la imagen en un dominio frecuencial local tras la aplicación de un banco de filtros lineales pasa-banda.

En la segunda etapa, sobre los coeficientes de la transformada wavelet a_{ik} , se aplican diferentes versiones para esta transformación \mathbf{R} : lineal, no-lineal dependiente del punto y Normalización Divisiva simplificada [8]. Pese a que JPEG2000 incluye transformaciones más complejas que las de su predecesor JPEG, todavía necesita emplear la versión simplificada de la *Normalización Divisiva*, debido a problemas en su inversión.

La finalidad de esta segunda transformación es eliminar las dependencias perceptuales existentes entre los coeficientes y representarlos en un dominio de la respuesta a_{ik} , perceptualmente Euclídeo. A fin de poder aplicar el cuantizador escalar uniforme [5], [3].

A continuación procederemos a describir detalladamente, todas las operaciones que se realizan sobre las imágenes dentro del estándar JPEG2000, basado en el modelo de procesado visual humano que acabamos de describir.

3.1. Transformación cromática

Sobre la imagen en color que se quiere comprimir se debe aplicar antes que nada una transformación para cambiar el espacio de representación del color, (ec. 3.2). Pasando del espacio RGB , propio del registrado por los píxeles de la cámara, a un espacio de representación cromática oponente donde se separan las contribuciones luminosa y cromática. Este nuevo espacio de color YUV , incluye una oponencia del tipo azul-amarillo en U y rojo-verde en V .

$$\begin{pmatrix} Y \\ U \\ V \end{pmatrix} = \begin{pmatrix} 0,222 & 0,707 & 0,071 \\ -0,109 & -0,348 & 0,457 \\ 0,682 & -0,620 & -0,063 \end{pmatrix} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (3.2)$$

El resultado de la transformación cromática al espacio oponente YUV , se ilustra en la figura 3.1.

La aplicación de este tipo de transformación cromática oponente sobre las imágenes en color, va a permitir:

1. Reducir la anchura frecuencial global de los canales del espacio transformado YUV , respecto de su anchura en el original RGB . Lo que permitirá codificar los coeficientes transformados con un menor número

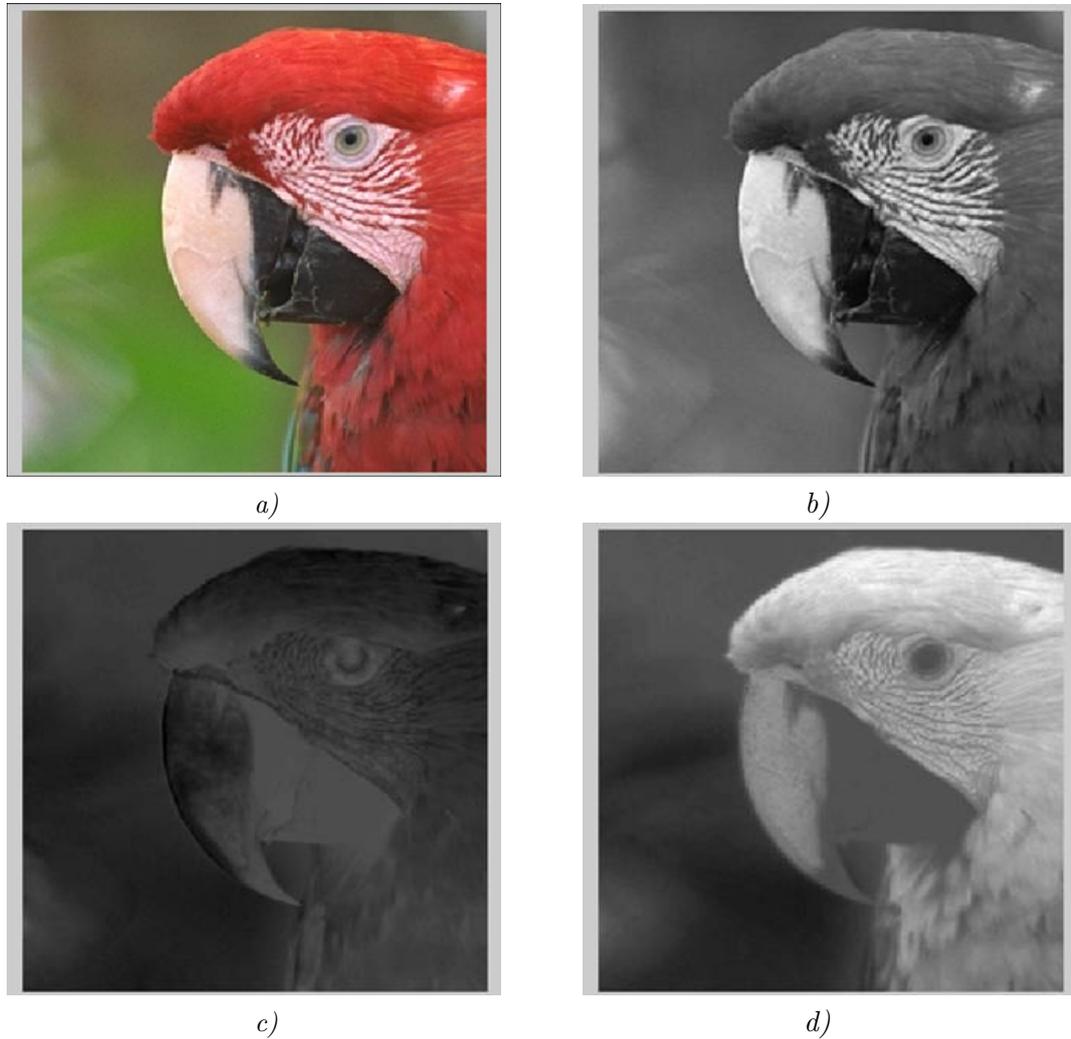


Figura 3.1: *Descomposición en el espacio oponente YUV, de la imagen del loro: a) Imagen original. b) Canal Y, contiene toda la información luminosa de la imagen. c) Canal cromático U, con oponencia amarillo-azul y d) Canal cromático V, con oponencia rojo-verde. Estos canales contienen la información cromática de la imagen.*

de niveles de cuantización.

2. Así como concentrar la máxima cantidad de información en un sólo canal, Y. En los espacios oponentes, la anchura de banda de los canales está jerarquizada, de modo que $Y > V > U$.

Estas mejoras conseguidas tras la aplicación de una transformación oponente, en la representación de la información contenida en las imágenes en

color, se ilustran mejor en la figura 3.2.

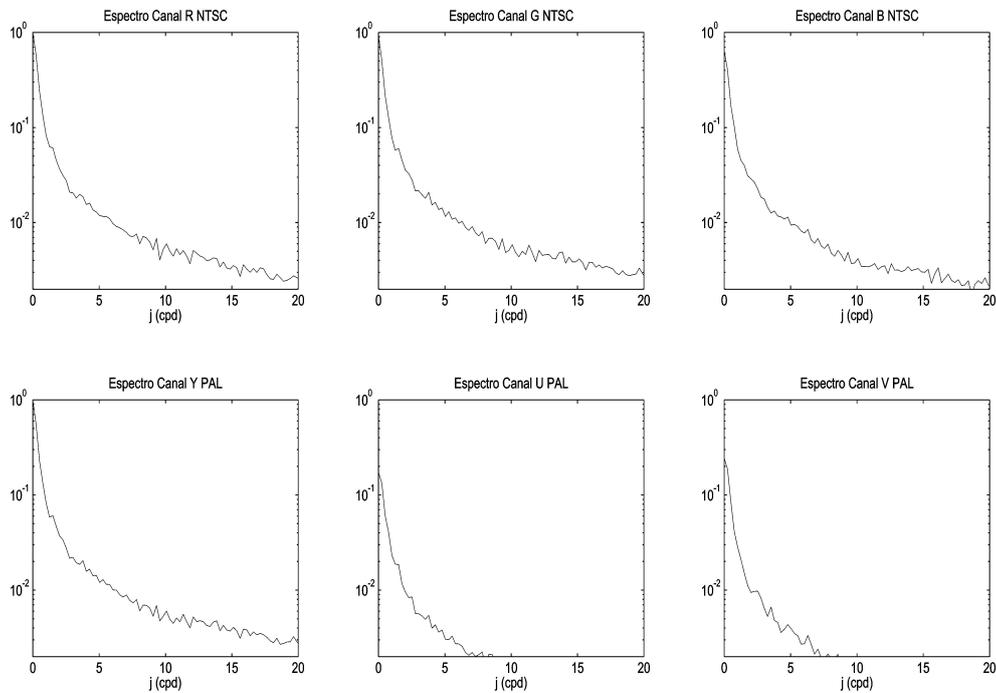


Figura 3.2: *Espectros promedio de 10 imágenes naturales en: a) el espacio RGB y b) en el espacio de primarios oponentes YUV. Vemos como se ha reducido la anchura de banda en la nueva representación YUV, respecto de la original RGB. Las siglas NTSC y PAL hacen referencia a los estándares de TV utilizados en EE.UU. y Japón (NTSC) y (PAL) en Europa. Junto con especificaciones técnicas, en estos estándares se fijan también los espacios cromáticos empleados para la adquisición/reproducción, RGB y para el almacenamiento/transmisión, YUV.*

La sensibilidad frecuencial del sistema visual humano es consistente con la composición frecuencial de las imágenes naturales, (figura 3.2). Por tanto, la mayor anchura de la banda acromática en las imágenes naturales, coincide con la mayor anchura de la CSF acromática, (figura 2.2). Con la anchura de las bandas cromáticas paso lo mismo, el canal V (rojo-verde) tiene una mayor anchura frecuencial que el U (azul-amarillo), lo cuál explicaría la relación de anchuras entre las CSF cromáticas.

Como el sistema visual humano es relativamente insensible al contenido de altas frecuencias de los canales cromáticos, éstos pueden submuestrearse en las dos direcciones del espacio a la mitad respecto al canal acromático, sin que por ello aumente la distorsión. De esta forma se elimina parte de la información cromática, en el sentido de Gabor [26], aprovechando el hecho de que en las imágenes naturales los canales cromáticos no aportan tanta

información perceptual, [27].

No hay que olvidar que en la búsqueda de una representación adecuada de las imágenes, para su posterior compresión, no se aplican las transformaciones directamente sobre los valores digitales de la imagen (números enteros en el rango [0-255]). Siendo necesario el paso a sus valores triestímulo, mediante una serie de transformaciones previas, para las que se ha empleado *Colorlab* [28]. Se trata de una herramienta matemática, usada bajo el entorno de Matlab como si fuera una toolbox más.

3.2. Transformación wavelet

La transformación más apropiada para la eliminación de las redundancias estadísticas contenidas en las imágenes naturales, estará basada en un modelo capaz describir sus características más comunes.

Debido a que el espectro de las imágenes naturales sigue una ley del tipo $1/f^2$, son las bajas frecuencias las que concentran la mayor parte de la energía de la imagen, (figura 3.2). El uso de la transformación *Wavelet* [29], [30], permite aprovechar esta peculiaridad de las imágenes, favoreciendo la concentración de la energía en unos pocos coeficientes y de este modo realizar una cuantización más eficiente al reducir el número de niveles necesarios.

La transformada Wavelet consiste en la actuación sobre la imagen de un banco de filtros lineales pasa-banda, que se asemejarían a los campos receptivos del córtex visual primario (o area V1) [13], [31], [32], [30] y van a permitir conocer su contenido espacio-frecuencial. Algunos autores proponen obtener estadísticamente las funciones base de la transformada, aplicando Análisis en Componentes Independientes más conocida como ICA [30], sobre las imágenes naturales. Todas las aproximaciones lineales de esta familia de transformaciones ICA, eliminan relaciones de segundo orden (correlación), y además, dependiendo del algoritmo concreto, minimizan información mutua, maximizan no-gaussianidad, o eliminan momentos concretos de orden superior a dos, por ejemplo el momento de orden 4, [30].

Por eso, el uso de la transformada wavelet (es el resultado de aplicar ICA sobre imágenes naturales [32]), mejora los resultados de independencia, respecto a transformadas de tipo PCA (Análisis en Componentes Principales) basada exclusivamente en la diagonalización de la matriz de covarianza o su homóloga DCT por bloques, cuyas funciones base son exponenciales extendidas por todo el dominio. En ambas transformadas sólo eliminan relaciones de segundo orden.

Por tanto, la transformada *wavelet* no sólo permite conocer las frecuencias que aparecen en la imagen, sino que da información adicional de su localización espacial dentro de la imagen. No obtendremos toda la informa-

ción de la frecuencia, pero a cambio si podremos localizarla en el momento de la reconstrucción de la imagen.

En la figura 3.4, podemos ver algunas funciones base de una transformación Wavelet ortogonal (de 3 escalas), con filtros QMF (Quadrature Mirror Filter). Para obtenerlas se han colocado δ 's en varias posiciones espacio-frecuencia del dominio wavelet (fig. 3.3) y se invierte la transformada.

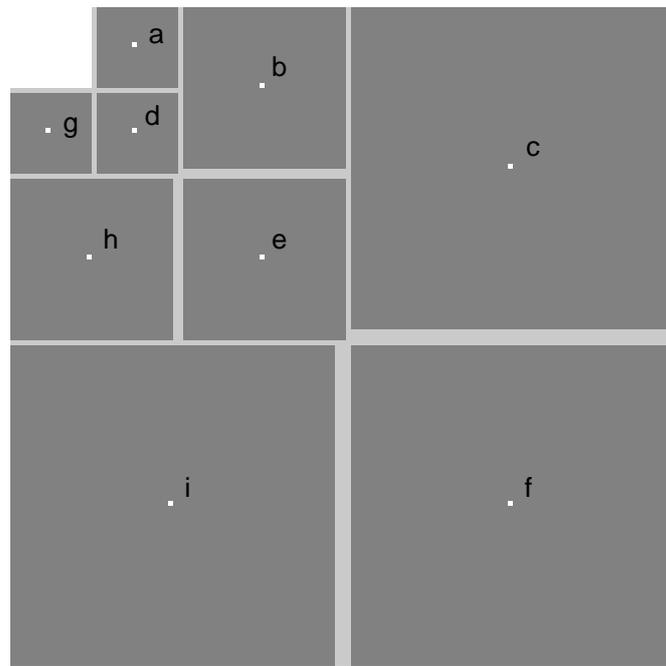


Figura 3.3: Posición espacio-frecuencial (en una transformada wavelet ortogonal de 3 escalas y 3 orientaciones) de las funciones filtro QMF mostradas en la figura 3.4.

La transformada wavelet de una imagen tiene estructura piramidal [29] y se va organizando por escalas, ver figuras 3.3 y 3.5. Para la implementación en el trabajo de la transformada wavelet, se ha utilizado la toolbox de QMF de 4 escalas y tres orientaciones espaciales [33].

En el recuadro superior izquierdo se encuentra la componente de continua, que contiene información de la luminancia media local de la imagen. Los coeficientes de la transformada se expresan en unidades de contraste previa realización de una normalización por la luminancia local media, usando la definición de contraste de Peli [19].

El tamaño de las subimágenes que la componen va creciendo con cada escala, siendo más pequeña para la primera escala y más grande para la última. En el ejemplo de la figura 3.5, la transformada wavelet de 4 escalas es

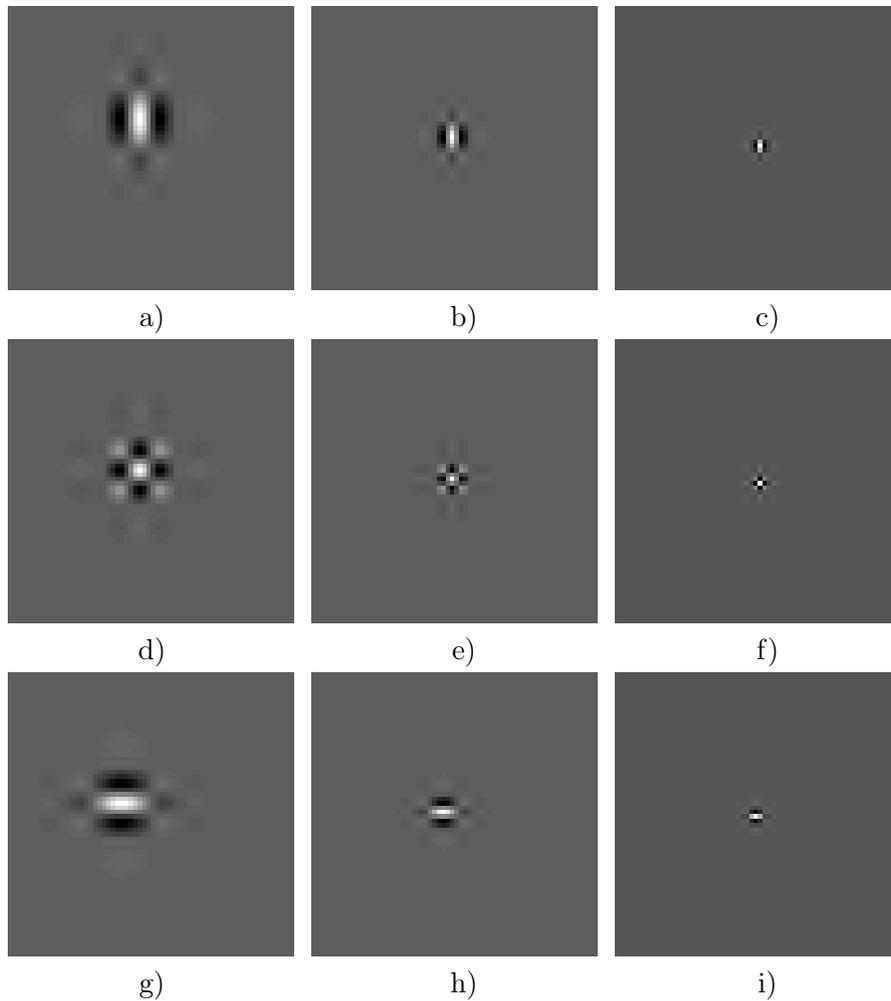


Figura 3.4: Ejemplos de las funciones base de la transformada wavelet ortogonal empleada (con gran parecido a los campos receptivos de V1 [14], [32]) para diferentes escalas y orientaciones con posición espacial central. Vemos como no se extienden por todo el espacio sino que están localizadas en una región. Lo que aporta información adicional para la localización de las frecuencias dentro de la imagen.

aplicada sobre una imagen de 256×256 , el tamaño de las subimágenes de la primera escala es de 16×16 mientras que para la cuarta es de 128×128 . Cada escala contiene a su vez tres subimágenes que se corresponden con las orientaciones del espacio: horizontal, diagonal y vertical, si lo vemos desde la izquierda.

Un simple análisis del comportamiento de los coeficientes de la transformada wavelet, a través de su varianza (figura 3.5.b). Permite destacar

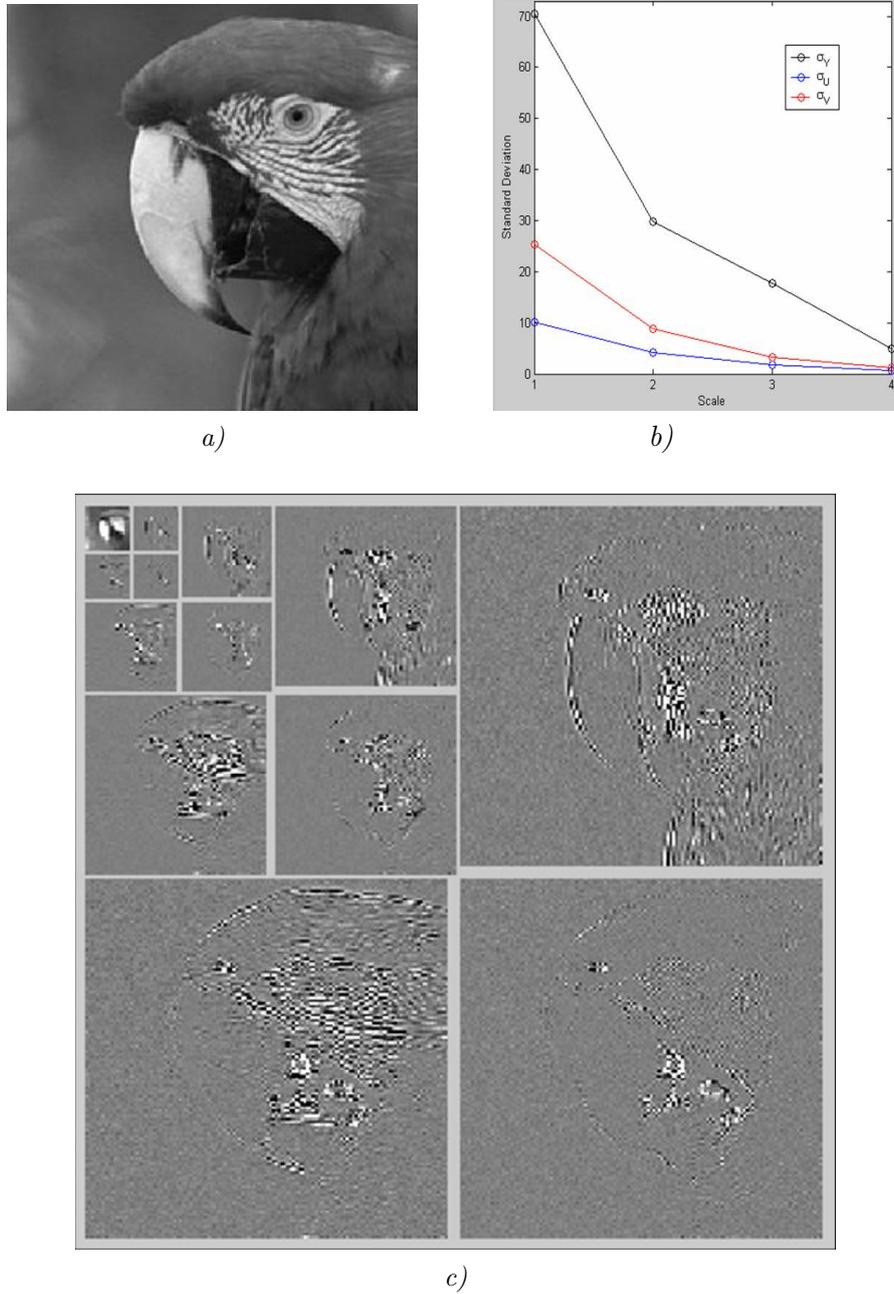


Figura 3.5: a) Imagen original. b) Varianza de los coeficientes, con las 4 escalas de la transformada wavelet. Para el canal acromático Y (línea negra), canales cromáticos U (línea azul) y V (línea roja). c) Transformada wavelet de 4 escalas de la imagen de luminancia.

la analogía existente entre el diferente decaimiento de la varianza en los canales cromáticos y la forma del espectro de las imágenes en color en el espacio oponente YUV, (figura 3.2). Esto evidencia una buena adaptación de la transformación wavelet, frente a las características de las imágenes naturales.

Tras aplicarla se han eliminado las dependencias de segundo orden y se consigue un código *sparse*, pero las imágenes contienen dependencias más complejas, que hacen que la energía de los coeficientes siga fuertemente correlacionada con la de sus vecinos. Esto es debido a que las imágenes naturales no proceden de una combinación lineal de señales de fuentes independientes, que es la suposición principal de las transformaciones lineales del tipo ICA. Se hace evidente la necesidad de aplicar un procesado adicional, una transformación no-lineal.

3.3. Transformación al dominio de la respuesta

Se trata de la última transformación, R , a la que se someten los coeficientes de la imagen. Permite representar los coeficientes de la transformada wavelet en el dominio de la respuesta, y en ella se incluyen diferentes modelos perceptuales. Es la transformación más crítica e importante debido a que determina el diseño del cuantizador, por tanto fijará la cantidad de bits asignados a cada subbanda e incluso la cantidad de coeficientes escogidos.

3.4. Modelos perceptuales *simplificados*

En el estándar JPEG2000 se incluyen tres diferentes versiones para esta transformación, asentadas en varios modelos perceptuales que irán aumentando en grado de dificultad [8].

A continuación procederemos a una descripción más detallada de cada uno de ellos:

1. Modelo lineal.
2. Modelo no lineal dependiente del punto (Point-wise).
3. Modelo no lineal: *Normalización divisiva simplificada*.

3.4.1. Modelo lineal

La asignación de bits en cada subbanda, se basa en la aplicación de la CSF. Es decir, para considerar la diferente visibilidad de las funciones base, dependiendo de su escala y canal cromático [34], se aplicará un peso diferente sobre cada escala, de acuerdo con la CSF cromática y acromática.

$$r_{ik} = \alpha_{s_i k} \cdot a_{ik} \quad (3.3)$$

donde s_i es la escala de la transformada para el coeficiente i , y k se refiere a cada uno de los canales cromáticos. En la figura 3.6 vemos un ejemplo de los pesos aplicados sobre cada escala y canal, para una transformada wavelet de 4 escalas [8].

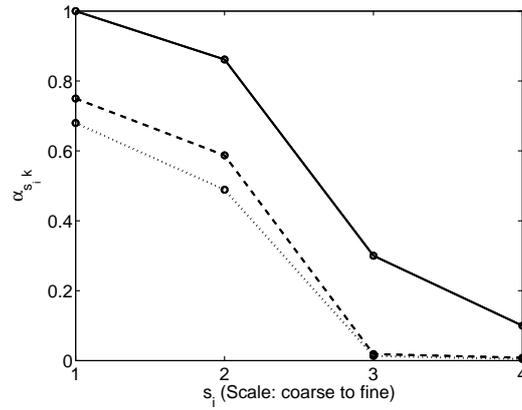


Figura 3.6: Pesos para cada escala de la transformada wavelet. Hace distinción entre los canales: acromático (línea continua) y cromáticos rojo-verde (línea discontinua) y azul-amarillo (línea punteada).

Este pesado puede obtenerse también de la varianza de los propios coeficientes de la transformada wavelet (figura 3.5). Asumiendo una óptima transmisión de la información en el sistema visual humano o una asignación óptima de la información para la codificación de las imágenes naturales con la mínima distorsión, [3], [35].

3.4.2. Modelo no lineal dependiente del punto

Debido a que los umbrales incrementales aumentan con el contraste [36], [20], una no linealidad fija, el exponente γ , es aplicada después del pesado en frecuencia.

$$r_{ik} = \text{sgn}(a_{ik}) \cdot |\alpha_{s_i k} \cdot a_{ik}|^\gamma \quad (3.4)$$

El valor recomendado del exponente, se mueve dentro del intervalo $[0,2 - 0,7]$, [8]. Esta no linealidad implica que sobre los coeficientes de mayor contraste se aplicarán los pasos de cuantización más grandes.

3.4.3. Modelo no lineal: *Normalización Divisiva simplificada*

La sensibilidad de los mecanismos de frecuencia (local) no sólo se reduce con el contraste de los estímulos de esa misma frecuencia y posición espacial, sino que está también influenciada por la presencia de estímulos vecinos en escala, orientación y espacio [5], [4], [17]. Los modelos actuales de enmascaramiento, también conocidos como modelos de control de ganancia, tienen en cuenta estas influencias y por tanto dividen cada coeficiente por una combinación lineal de sus vecinos.

A este modelo se le conoce formalmente como *Normalización Divisiva*, se pueden ver más detalles de la misma en el capítulo 2.

$$r_{ik} = \frac{\text{sgn}(a_{ik}) \cdot |\alpha_{s_{ik}} \cdot a_{ik}|^\gamma}{\beta_{ik} + \sum_{j=1}^{N_k^2} h_{ij} \cdot |\alpha_{s_{jk}} \cdot a_{jk}|^\gamma} \quad (3.5)$$

La función signo $\text{sgn}()$, el valor absoluto y el exponente γ , son aplicados sobre cada elemento i del vector a . $k = 1, 2, 3$, hace referencia a los tres canales cromáticos que constituyen la imagen en color.

El parámetro crítico de esta transformación es la matriz h_{ij} , donde se recogen las interacciones existentes entre los coeficientes a_{ik} , que posteriormente influirán en cada respuesta r_{ik} . Siguiendo recomendaciones del estándar [8], el valor de los parámetros β y γ , se ha fijado en $\beta_{ik} = 1$ y $\gamma = 0,5$.

En la literatura relacionada [5], [17], para el modelo más general de esta normalización, se asume una vecindad gaussiana en escala, orientación y espacio. Sin embargo, como se verá en el capítulo 5, este esquema de interacción general tiene un proceso de inversión (obtención de los coeficientes a_{ik} a partir de la respuesta r_{ik}) bastante costoso.

Esta es la razón de por qué el estándar JPEG2000 restringe las interacciones recogidas en la matriz h_{ij} , a sólo los vecinos que caen dentro de una misma subbanda [8], ver figura 3.7. Esto supone despreciar las interacciones en escala y orientación, de ahí viene el nombre de *Normalización Divisiva simplificada*.

3.5. Diseño del cuantizador

Las transformaciones aplicadas sobre la imagen, han tenido como único objetivo la eliminación de las dependencias existentes entre los píxeles de la imagen. Debido a que las transformaciones son invertibles, en sí mismas no incurren en compresión alguna. Por tanto para reducir la cantidad de bits empleados en codificar la imagen, lo que habrá que hacer es cuantizarla.

El cuantizador reduce la complejidad de la imagen representándola con el menor número de representantes, lo que reduce a su vez la entropía, pero

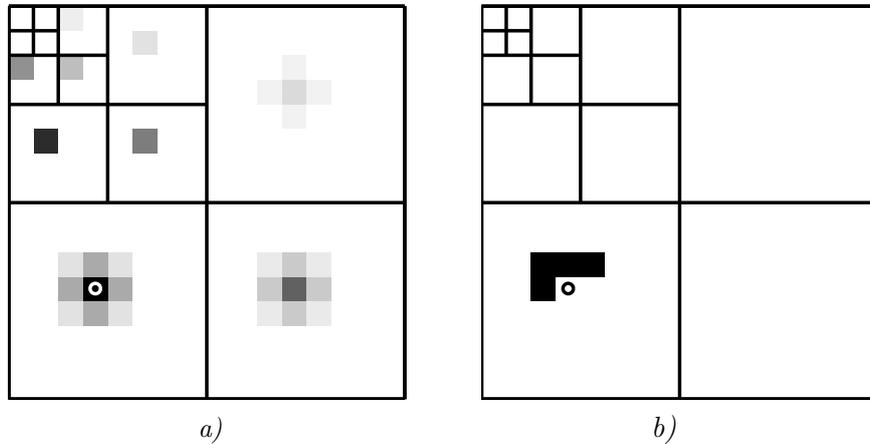


Figura 3.7: a) Ejemplos de diferentes vecindades descritas por la matriz de interacción h_{ij} , en su versión general y simplificada respectivamente, para un coeficiente particular de la transformada wavelet (marcado en círculo). a) Kernel gaussiano, incluye interacciones en escala, orientación y espacio. b) Kernel simplificado usado por el estándar JPEG2000, incluye interacciones espaciales dentro de una misma subbanda. El tamaño de la vecindad considerada, es típicamente una ventana de 5×5 [8].

introduciendo una cierta distorsión en la imagen debido a que se ha eliminado información de manera irreversible.

Un buen cuantizador debe tratar de minimizar simultáneamente, la complejidad de la imagen y la distorsión introducida. Pero al tratarse de requisitos contrapuestos, bastará con que se alcance un buen compromiso entre complejidad y distorsión.

En el diseño del cuantizador se han de considerar las propiedades de discriminación del sistema visual humano, para así concentrar los errores de cuantización en las zonas perceptualmente menos relevantes, sin peligro de introducir errores graves.

De igual modo, el cuantizador debe tener en cuenta las propiedades estadísticas de las imágenes naturales a codificar y de este modo adaptarse mejor a los datos. Asignando más niveles de cuantización a los elementos de mayor varianza, para así evitar errores graves.

Se usan varios criterios a la hora de diseñar el cuantizador:

1. Basados en la minimización del error cuadrático medio, MSE, con soluciones no uniformes para el diseño del cuantizador, se fundamentan en el cálculo de las PDF marginales [3] o en alguna modificación de estas donde se incluye la métrica perceptual, [35].
2. O basados en la restricción del máximo error perceptual, MPE, es decir reducir al mínimo la distorsión provocada por los coeficientes cuyo valor

se sale de lo normal (outliers). Esto conseguirá un mejor comportamiento en la distorsión final, que la reducción del error medio, [1], [2], [24], [3].

En un dominio perceptualmente uniforme, la restricción del MPE es equivalente a la aplicación de un cuantizador escalar uniforme con un codificador de bits también uniforme. De acuerdo con esto, el diseño del cuantizador pasaría a ser no uniforme en el dominio intermedio de las wavelet.

Por tanto, el objetivo de todas las transformaciones que se aplican sobre la imagen, será el de poder representarla en un dominio perceptualmente Euclídeo y de este modo aplicar un sencillo cuantizador uniforme junto con un codificador también uniforme.

La diferencia fundamental entre las aproximaciones que siguen esta idea de minimizar el MPE, [1], [24], [21], radica en el grado de precisión del modelo perceptual empleado para proponer que el dominio sea perceptualmente Euclídeo.

El esquema de compresión recomendado por el estándar JPEG2000 [8], [7], estará basado en la restricción del MPE y en la suposición de dominio perceptualmente Euclídeo con cada uno de los modelos perceptuales que incluye: lineal, no-lineal dependiente del punto y Normalización Divisiva simplificada.

Por último decir que en las imágenes en color, el cuantizador se aplica secuencialmente sobre cada banda (acromática o cromática), produciendo a la salida una imagen codificada en dichas bandas.

Capítulo 4

Beneficios estadísticos de la *Normalización Divisiva* en la Codificación de la Transformada

En este capítulo ilustraremos ejemplos del comportamiento de las imágenes naturales, que sugieren la necesidad de normalizar la energía de cada coeficiente de la transformada wavelet por una combinación de la de sus vecinos en escala, orientación y espacio, [5], [4], [17], [18], para de este modo aumentar la independencia estadística entre los coeficientes. Es decir, la aplicación de la *Normalización Divisiva* en su versión más general.

Además se presentan unas medidas cuantitativas que permitirán cuantificar, la independencia estadística conseguida entre los coeficientes transformados, tras la aplicación de ésta normalización.

4.1. Independencia entre coeficientes wavelet

En una imagen de tamaño $N \times N$ tenemos N^2 coeficientes, cada uno con su distribución de probabilidad, varianza, etc. Entre ellos, existen a su vez correlaciones estadísticas [37], que se intentan eliminar con la aplicación de transformadas del tipo DCT por bloques o wavelet.

Sin embargo estas transformaciones lineales, sólo consiguen eliminar las relaciones de segundo orden y algunas de cuarto, permaneciendo las de órdenes superiores.

Esto puede verse mejor en la figura 4.1, donde la energía de un coeficiente (remarcado en amarillo) está relacionada con la de sus vecinos en escala, orientación y espacio. Esta vecindad se adapta bastante bien a la descrita por el Kernel gaussiano, [5], [17], que se incluye en la *Normalización Divisiva*.

Es decir, si un coeficiente tiene alta energía, es muy probable que sus vecinos también la tengan.

Analizando el resultado de la actuación de la transformada wavelet, desde un punto de vista cualitativo, destaca la presencia de relaciones claras entre los coeficientes, que la transformada no ha conseguido eliminar.

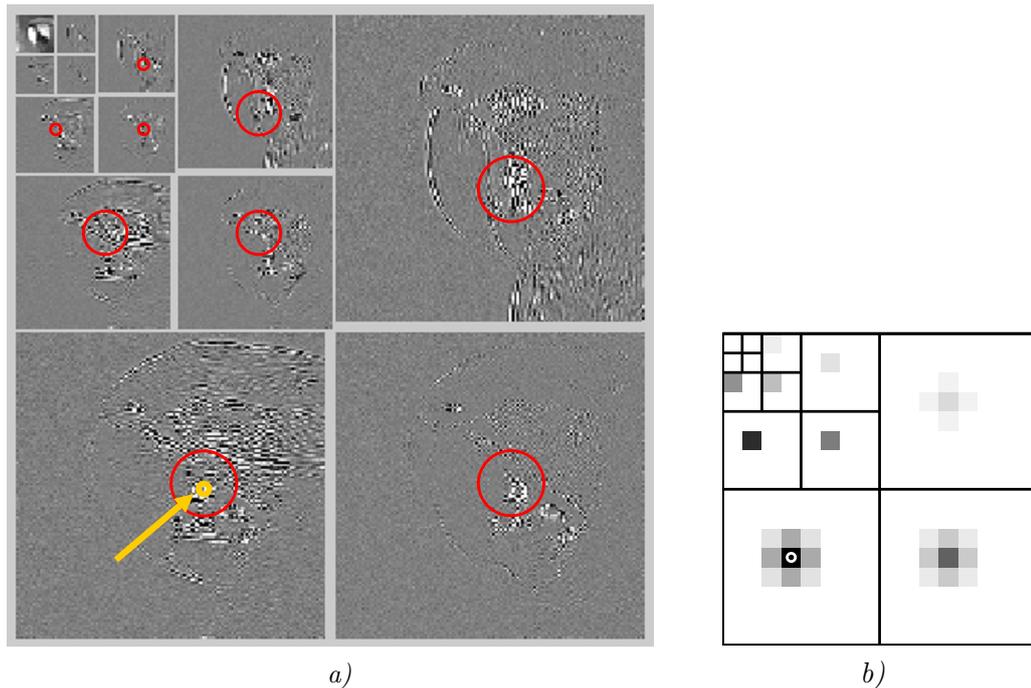


Figura 4.1: a) Transformada wavelet de 4 escalas, donde puede verse la relación entre la energía del coeficiente (en amarillo) y la de sus vecinos (en rojo). Cuya vecindad viene descrita, en b), por el Kernel de interacción gaussiano, que incluye dependencias con vecinos de diferentes escalas, orientaciones y posiciones espaciales.

Parece lógico pensar que normalizando la energía de cada coeficiente wavelet, por una combinación de la de sus vecinos, se puedan reducir las correlaciones existentes y por tanto la independencia aumente.

Haciendo uso de la *Normalización Divisiva* en su versión más general, la energía de cada coeficiente wavelet se normaliza por una combinación de la energía de sus vecinos en escala, orientación y espacio, [5], [4], [17].

El resultado de esta normalización se ilustra en la figura 4.2. Puede verse como donde antes había una cierta estructura entre los coeficientes, tras la normalización aparece ruido en la imagen.

Por tanto, la aplicación de esta normalización parece aumentar la inde-

pendencia entre los coeficientes de la transformada wavelet.

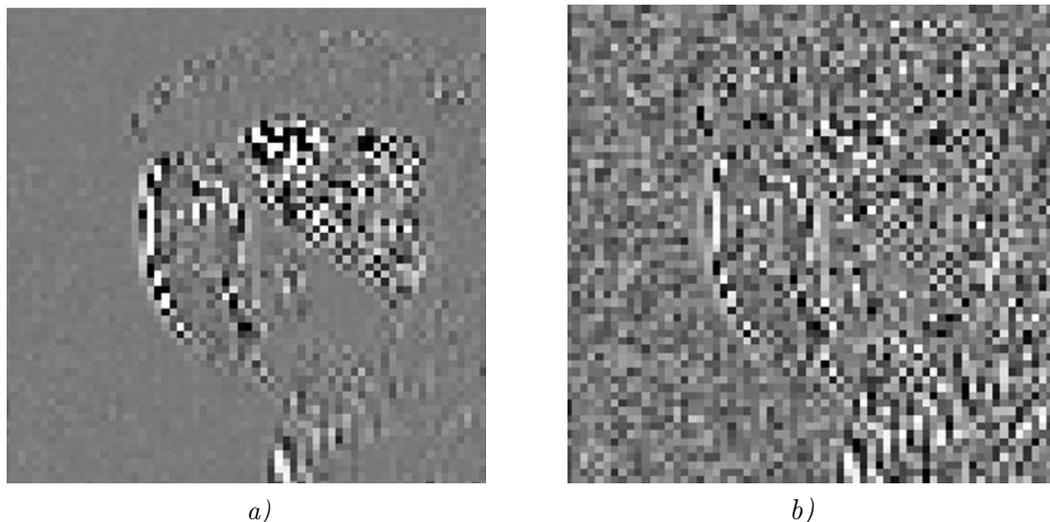


Figura 4.2: Ambas imágenes se obtienen de extraer de la estructura de la transformada wavelet (fig. 4.1), la subimagen correspondiente a la escala 3 y orientación vertical (ie. 3ª imagen tomando como origen la imagen de continua). a) Imagen original. b) Imagen obtenida tras la aplicación de la Normalización Divisiva.

Lo dicho hasta ahora, procede del análisis *cualitativo* de los datos de que se dispone. A continuación se presentarán unas medidas *cuantitativas*, que nos va a permitir cuantificar la cantidad de independencia lograda entre los coeficientes de la transformada wavelet, tras la aplicación de la *Normalización Divisiva*.

4.2. Probabilidad condicionada e Información mutua

Como medida de la existencia de correlaciones estadísticas entre los coeficientes de la transformada wavelet, calcularemos:

- La *Probabilidad condicionada* de dos coeficientes cualesquiera, de los que constituyen la imagen. Representando gráficamente el resultado de la medida, destaca la estructura característica en los datos (pajaritas), señal evidente de la existencia de correlaciones.
- La *Información mutua*, permite conocer cuánta información pueden darnos de un coeficiente sus vecinos, es decir como de relacionados están los coeficientes de la transformada wavelet.

Probabilidad condicionada

Para averiguar si entre los coeficientes a_1 y a_2 existen relaciones, procedemos a calcular la probabilidad condicionada de a_1 dado a_2 , $P(a_1/a_2)$. Es decir, para cada valor que tome el coeficiente a_2 , cuál es la distribución de probabilidad de a_1 .

$$P(a_1/a_2) = \frac{P(a_1, a_2)}{P(a_2)} \quad (4.1)$$

donde $P(a_1, a_2)$ es la probabilidad conjunta, es decir la probabilidad de que un coeficiente valga a_1 y el otro a_2 , obtenida a partir de las PDF (Función Densidad de Probabilidad) individuales de cada coeficiente.

Haciendo los cálculos de la probabilidad condicionada para dos coeficientes cualesquiera de la transformada wavelet. Si representamos la probabilidad condicionada $P(a_1/a_2)$ para todos los valores de a_2 , con a_2 en el eje X y a_1 en el eje Y, figura 4.3.

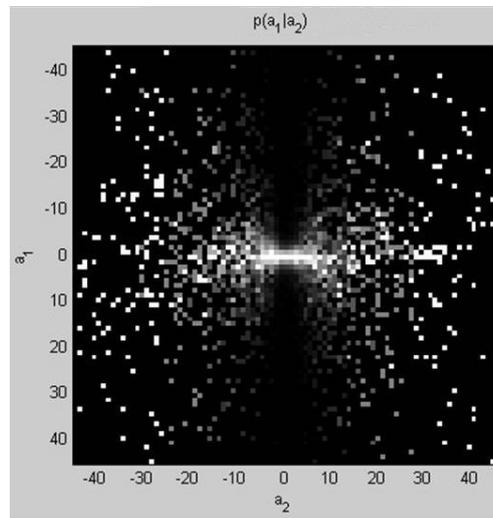


Figura 4.3: *Relación existente entre dos coeficientes, vecinos espaciales, de la transformada wavelet. Para valores pequeños de a_2 , la varianza de a_1 es pequeña. A medida que a_2 aumenta su valor, la varianza de a_1 también crece. Por tanto la varianza de a_1 , depende del valor que tome a_2 , indicando la presencia de relaciones residuales de orden superior.*

Vemos claramente la existencia de relaciones entre los coeficientes de la transformada Wavelet. Se trataría de las relaciones de orden superior de las que hemos hablado en el apartado anterior. Y que advertimos tras la

observación de los resultados, después de la actuación de la transformada Wavelet.

Aunque las transformaciones espacio-frecuenciales, en nuestro caso las Wavelet, eliminan relaciones de segundo orden e incluso algunas de cuarto orden entre los coeficientes de la imagen, su energía (varianza) continúa estando altamente correlacionada con la de sus vecinos, como acabamos de ver. Por tanto, está más que justificada la aplicación sobre los coeficientes a_1 y a_2 de la transformada Wavelet, de la *Normalización Divisiva* en su versión general y de este modo normalizar su energía por una combinación de la de sus vecinos.

El resultado de esta normalización, se ilustra en la figura 4.4.

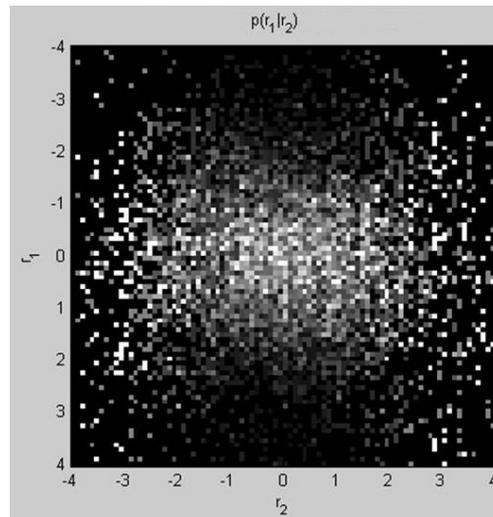


Figura 4.4: En el dominio normalizado las relaciones entre los coeficientes a_1 y a_2 de la transformada wavelet desaparecen y por tanto la estructura de 'pajaritas' de los datos también.

Puede verse como en el dominio normalizado las relaciones entre los coeficientes de la transformada wavelet han desaparecido. De este modo ya no se podrá predecir el valor de un coeficiente a partir de otro. Es decir su varianza ya no dependerá del valor que tome otro coeficiente.

Información mutua

Es una medida de la independencia estadística de las amplitudes de los coeficientes. Resulta muy útil para conocer cuánta información, en bits, pueden darnos los vecinos a cerca de un coeficiente en particular y por tanto como de fuertes son las interacciones estadísticas.

La información mutua de un conjunto de variables (a_1, \dots, a_n) , está definida como la distancia Kullback-Leibler entre las PDF conjuntas y el producto de sus marginales. Puede ser calculada a partir de las entropías marginales, $H(a_i)$, y la entropía conjunta, $H(a_1, \dots, a_n)$, de las variables [38].

$$I(a_1, \dots, a_n) = \sum_{i=1}^n H(a_i) - H(a_1, \dots, a_n) \quad (4.2)$$

Donde $I(a_1, \dots, a_n)$ puede interpretarse como el número medio de bits que se pierden al codificar las variables, asumiendo que son independientes.

Sin embargo, para un mismo valor de la información mutua pueden haber diferentes interpretaciones. Debido a que la información mutua depende de la entropía marginal de las variables consideradas, $H(a_i)$. Y ésta puede ser bastante diferente de un dominio a otro (espacial, transformado y normalizado), por la forma diferente de la PDF marginal.

Por tanto la medida de información mutua en los diferentes dominios, estará condicionada en favor de las variables con una mayor PDF marginal no-uniforme, es decir aquellas con una menor entropía.

Dicho esto definiremos la información mutua relativa, I_r [2].

$$I_r(a_1, \dots, a_n) = \frac{\frac{1}{n-1} \cdot I(a_1, \dots, a_n)}{\frac{1}{n} \cdot \sum_{i=1}^n H(a_i)} \quad (4.3)$$

Donde $I_r(a_1, \dots, a_n)$ puede interpretarse como la proporción de bits que son perdidos al asumir en la codificación, que las variables son independientes. Si $I_r = 1$, las variables a_i son completamente redundantes, (idénticas) y si $I_r = 0$, las variables son independientes.

En el cuadro 4.1 se muestran los resultados del cálculo de la información mutua relativa I_r , de dos coeficientes de la imagen (los mismos para los que antes calculamos su probabilidad condicionada), tanto en el dominio de la transformada Wavelet como en el dominio de la *Normalización Divisiva*.

Destaca la significativa reducción que se obtiene en la medida de la información mutua relativa, I_r , tras la aplicación de la *Normalización Divisiva* sobre los coeficientes de la transformada Wavelet.

Justificando de un modo cuantitativo, tanto la presencia de interacciones

	<i>Dominio Wavelet</i>	<i>Dominio Normalizado</i>
\bar{I}_r	0.1044	0.0007

Cuadro 4.1: *Resultados de la medida de información mutua relativa, entre dos coeficientes, en el dominio de la transformada Wavelet y en el normalizado, (eq. 4.3). Indica el grado de independencia estadística entre ambos coeficientes.*

entre los coeficientes, como el aumento conseguido en la independencia estadística entre los coeficientes wavelet tras la utilización de la *Normalización Divisiva*.

Capítulo 5

Inversión de la *Normalización Divisiva*

De la observación de los beneficios estadísticos introducidos, tras el uso de la *Normalización Divisiva* en su versión más general, en la mejora de la independencia de los coeficientes de la transformada wavelet, (capítulo 4). Parece más que justificada la inclusión de la versión general de ésta transformada, dentro del estándar JPEG2000.

A continuación veremos como problemas en la inversión analítica de la *Normalización Divisiva*, son los responsables de que el estándar no trabaje con la versión general y tenga que ceñirse al uso de la versión simplificada.

En el trabajo propondremos un método de inversión, basado en el uso de la expansión en serie. Propuesto originalmente en el contexto de la transformada DCT por bloques [2] y que hemos generalizado su uso para la transformada Wavelet.

5.1. El problema de la inversión

Si a una imagen de tamaño $N \times N$ se le aplica el modelo perceptual descrito anteriormente. El tamaño de los vectores a , r y β , implicados en la transformación *Normalización Divisiva*, es de N^2 y para la matriz h , $N^2 \times N^2$.

Considerando estos tamaños y el patrón de interacción entre coeficientes que describe la matriz h , la solución analítica al problema de la inversión de la *Normalización Divisiva* no es nada práctica, debido al alto coste computacional que conlleva el cálculo de la inversa de esta matriz, $(I - D_{rk} \cdot h)^{-1}$.

$$a_k = \text{sgn}(r_k)((I - D_{rk} \cdot h)^{-1} \cdot D_{\beta k} \cdot |r_k|)^{\frac{1}{\gamma}} \quad (5.1)$$

donde D_{rk} y $D_{\beta k}$ son matrices diagonales con el valor absoluto de los vectores r_k y β_k en la diagonal. Al igual que ocurría con la ecuación (3.5) de la

transformación, la función signo $sgn()$, el valor absoluto y el exponente γ , son aplicados sobre cada elemento del vector r_k .

Cabe destacar que mientras el cálculo de la transformación es eficiente, debido a la naturaleza local de la interacción entre los coeficientes a_k y por tanto la matriz h es poco densa o *sparse*. En el cálculo de la inversa, la interacción existente entre los coeficientes r_k es global, esto hace que la matriz $(I - D_{rk} \cdot h)^{-1}$ sea muy densa. Y por tanto que el cálculo de a_k a través de la solución analítica sea muy costoso, incluso para imágenes de un tamaño moderado.

El estándar JPEG2000 resuelve este problema considerando un kernel de interacción h simplificado [8], con ello la inversión de la transformada requiere unos cálculos más sencillos. Pero esta simplificación supondrá un empeoramiento de los resultados obtenidos en la codificación de las imágenes, como se verá en el capítulo 6 de resultados.

5.2. Solución: el método de expansión en serie

La solución alternativa al método de inversión analítico, se propuso originalmente para imágenes acromáticas dentro del contexto de la transformada DCT por bloques [2]. Ampliarlo para uso con imágenes en color y en un dominio wavelet, no supondrá ningún problema y es por esto que lo emplearemos como método de inversión de la *Normalización Divisiva*.

Se fundamenta en expresar la matriz inversa como una expansión en serie y de este modo no será necesario invertir ninguna matriz densa:

$$(I - D_r \cdot h)^{-1} = \sum_{n=0}^{\infty} (D_{rk} \cdot h)^n \quad (5.2)$$

De esta forma la inversa se puede calcular hasta un cierto grado de aproximación, n , considerando un número finito de pasos de la serie:

$$\begin{aligned} |a_k|_{(1)}^\gamma &= D_{\beta k} \cdot r_k + (D_{rk} \cdot h) \cdot D_{\beta k} \cdot r_k \\ |a_k|_{(2)}^\gamma &= D_{\beta k} \cdot r_k + (D_{rk} \cdot h) \cdot D_{\beta k} \cdot r_k + (D_{rk} \cdot h)^2 \cdot D_{\beta k} \cdot r_k \\ |a_k|_{(3)}^\gamma &= D_{\beta k} \cdot r_k + (D_{rk} \cdot h) \cdot D_{\beta k} \cdot r_k + (D_{rk} \cdot h)^2 \cdot D_{\beta k} \cdot r_k + \\ &\quad + (D_{rk} \cdot h)^3 \cdot D_{\beta k} \cdot r_k \\ &\vdots \end{aligned}$$

Estas ecuaciones pueden ser reescritas de la siguiente manera:

$$\begin{aligned} |a_k|_{(1)}^\gamma &= D_{\beta k} \cdot r_k + (D_{rk} \cdot h) \cdot D_{\beta k} \cdot r_k \\ |a_k|_{(2)}^\gamma &= D_{\beta k} \cdot r_k + (D_{rk} \cdot h) \cdot ((D_{rk} \cdot h) \cdot D_{\beta k} \cdot r_k + D_{\beta k} \cdot r_k) \end{aligned}$$

$$\begin{aligned}
 |a_k|_{(3)}^\gamma &= D_{\beta k} \cdot r_k + (D_{rk} \cdot h) \cdot ((D_{rk} \cdot h) \cdot ((D_{rk} \cdot h) \cdot D_{\beta k} \cdot r_k + \\
 &\quad + D_{\beta k} \cdot r_k) + D_{\beta k} \cdot r_k) \\
 &\vdots
 \end{aligned}$$

Una implementación simplista de estas ecuaciones nos puede llevar a trabajar con potencias del tipo $D_{rk} \cdot h$, lo cual también sería un problema. Sin embargo es posible escribir la expansión en series de un modo recursivo, donde sólo aparecen sumas entre vectores y multiplicaciones de matriz-vector:

$$\begin{aligned}
 |a_k|_{(0)}^\gamma &= D_{\beta k} \cdot r_k \\
 |a_k|_{(n)}^\gamma &= D_{\beta k} \cdot r_k + D_{rk} \cdot h \cdot |a_k|_{(n-1)}^\gamma
 \end{aligned} \tag{5.3}$$

En la ecuación 5.3, vemos como todas las matrices que aparecen son *sparse* y por tanto la inversión de la transformación *Normalización Divisiva* puede realizarse de forma rápida y eficiente.

5.3. Condición de invertibilidad

Los procedimientos de inversión de la *Normalización Divisiva* que se han propuesto, tanto el *analítico* como el basado en la *expansión en serie*, comparten la misma condición de invertibilidad. Es decir, la condición que garantiza la existencia de solución al problema de la inversión en ambos procedimientos es la misma.

Planteando el problema como una descomposición en valores propios de la matriz $D_{rk} \cdot h$:

$$D_{rk} \cdot h = V \cdot \lambda \cdot V^{-1}$$

donde λ y V son el autovalor y autovector respectivamente.

Por tanto la condición de invertibilidad, sería:

$$\lambda_{max} = \max(\lambda_i) < 1 \tag{5.4}$$

Analizaremos esta condición para ambas soluciones:

1. En la solución *analítica*, para que la matriz $(I - D_{rk} \cdot h)$ sea invertible su determinante, $\det(\lambda_i I - D_{rk} \cdot h) \neq 0$. En el caso de que algún autovalor, λ_i , fuera igual a uno, el determinante se anularía y no habría solución. Por tanto se debe cumplir la condición (5.4).

2. En la solución basada en la *expansión en serie*, la convergencia de las series debe estar garantizada. Empleando en las series de expansión la descomposición en valores propios de la matriz $D_{rk} \cdot h$, tendremos:

$$\sum_{m=0}^{\infty} (D_{rk} \cdot h)^m = V \cdot \left(\sum_{m=0}^{\infty} \lambda^m \right) \cdot V^{-1} \quad (5.5)$$

Esto claramente converge sólo si el máximo autovalor, λ , es menor que uno, como indica la condición 5.4.

En el presente trabajo no hemos comprobado exhaustivamente si se cumple la condición de invertibilidad, para el caso de la transformada wavelet: tan solo hemos aplicado el método sin explorar los valores propios λ_{max} obteniendo hasta el momento resultados positivos en la convergencia para todos los casos analizados (aprox. 20 imágenes). Resulta, por tanto, pendiente la realización de una comprobación sobre una base de imágenes suficientemente extensa.

Capítulo 6

Resultados

Una vez superado el problema de la inversión de la *Normalización Divisiva*, estamos en condiciones de aplicar esta versión general sobre los coeficientes de la transformada wavelet. Y de este modo representar las imágenes naturales en un dominio de la respuesta realmente *Euclídeo*, ya que se han conseguido eliminar las interacciones existentes entre los píxeles de la imagen. Por tanto en este dominio ya no habrá que preocuparse por el diseño del cuantizador, empleando el sencillo cuantizador escalar uniforme.

A diferencia del estándar JPEG2000, donde los modelos perceptuales incluidos son versiones simplificadas, que no consiguen eliminar las interacciones entre los coeficientes y de este modo suponer un dominio perceptualmente euclídeo. Para así emplear un cuantizador escalar uniforme [8], [7], basado en la restricción del máximo error perceptual (MPE), [25], [2], junto con un codificador de bits también uniforme.

6.1. Imágenes codificadas

En las figuras 6.1 y 6.2 se muestran los resultados obtenidos tras la aplicación de los tres modelos perceptuales simplificados que incluye JPEG2000 (lineal, no lineal dependiente del punto y Normalización Divisiva simplificada), frente al uso de la versión general de la *Normalización Divisiva* que se propone en el trabajo, para una tasa de compresión de la imagen de 0.2bits/pix.

Del análisis de las imágenes obtenemos las consecuencias que conlleva el uso del estándar JPEG2000:

1. En las imágenes codificadas con los modelos que no incluyen enmascaramiento (*lineal y no lineal dependiente del punto*), se aprecia una sobresaturación de los colores, así como la aparición de falsos colores en la dirección cromática oponente. Es decir, aparecen falsos azules en

los bordes de la imagen rojo-verde (6.1) y falsos detalles en rojo en la imagen azul-amarillo (6.2).

2. Tras emplear el modelo con enmascaramiento, la *Normalización Divisiva aunque en su versión simplificada*, se aprecia una mejora. Las imágenes ya no presentan estos falsos colores, pero se ha reducido el contraste de los detalles de alta frecuencia y por tanto la agudeza de los contornos.

Los defectos descritos, sobre las imágenes codificadas con el estándar JPEG2000, se deben al ruido cromático añadido en las imágenes, como consecuencia de emplear modelos simplificados de enmascaramiento en los canales cromáticos.

Por contra, en las imágenes codificadas con el modelo propuesto de *Normalización Divisiva en su versión general*, se reproduce correctamente el color, además de conservar los detalles espaciales contenidos en la imagen.

6.2. Curvas Rate-Distorsion

El uso de estos modelos simplificados *repercute no sólo* en la pérdida de calidad visual de la imagen comprimida, como acabamos de ver.

También se obtienen peores resultados para las curvas de Rate-Distorsion, ver figura 6.3, que han sido calculadas para diferentes valores de la entropía dentro del rango $[0,1 - 0,7]$ bits/pix (equivalente a un ratio de compresión entre 240:1 y 34:1 respectivamente).

Como medida de distorsión, es decir diferencia entre la amplitud del coeficiente transformado y su versión cuantizada, se suele emplear el error cuadrático medio, MSE. Sin embargo el MSE es una medida objetiva y no refleja desde un punto de vista subjetivo, el comportamiento de los algoritmos que se han descrito [39, 40]. Por tanto obtendremos nuevas medidas de la distorsión empleando las métricas subjetivas S-CIELab [41] y SSIM [42].

Los beneficios del modelo propuesto, se confirman con las curvas de Rate-Distorsion obtenidas. Mejorando para todos los valores de entropía, la distorsión introducida en la compresión de la imagen. Reduciendo el volumen de la señal hasta en un 35 %, respecto al resultado obtenido con el estándar JPEG2000, para un valor de MSE constante (de igual modo para SSIM). La reducción en volumen es incluso mayor, empleando el S-CIELAB. Pero más importante que los datos numéricos, es comprobar con una inspección directa, la mejora de la calidad visual conseguida sobre la imagen.

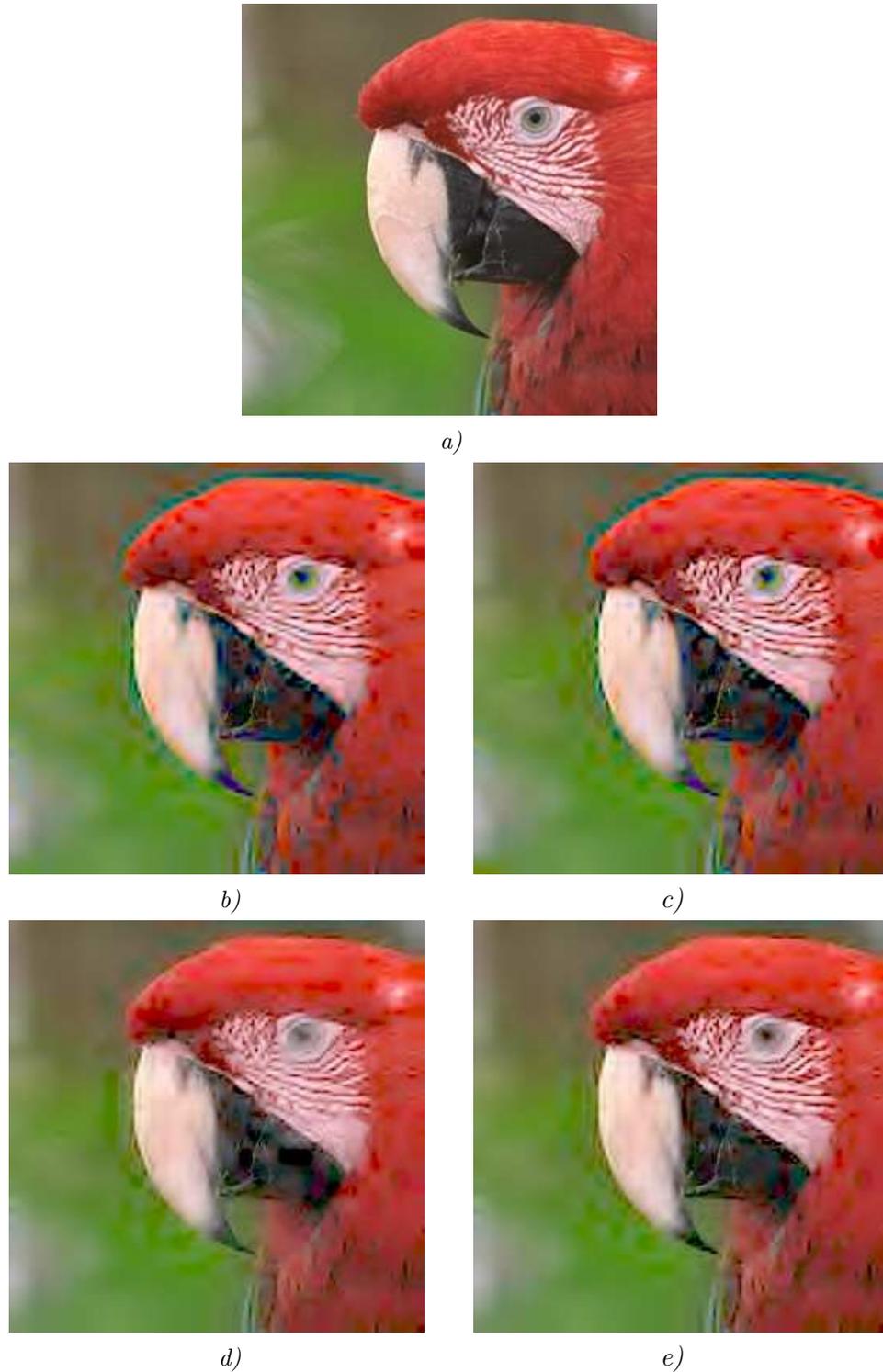


Figura 6.1: Imágenes codificadas empleando diferentes modelos perceptuales: a) Imagen original del loro rojo-verde. b) Lineal JPEG2000. c) No lineal dependiente del punto JPEG2000. d) Normalización Divisiva simplificada JPEG2000. e) Normalización Divisiva general (modelo propuesto), para una tasa de compresión de 0.2 bits/pix.



a)



b)



c)



d)



e)

Figura 6.2: Imágenes codificadas empleando diferentes modelos perceptuales: a) Imagen original del loro azul-amarillo. b) Lineal JPEG2000. c) No lineal dependiente del punto JPEG2000. d) Normalización Divisiva simplificada JPEG2000. e) Normalización Divisiva general (modelo propuesto), para una tasa de compresión de 0.2 bits/pix.

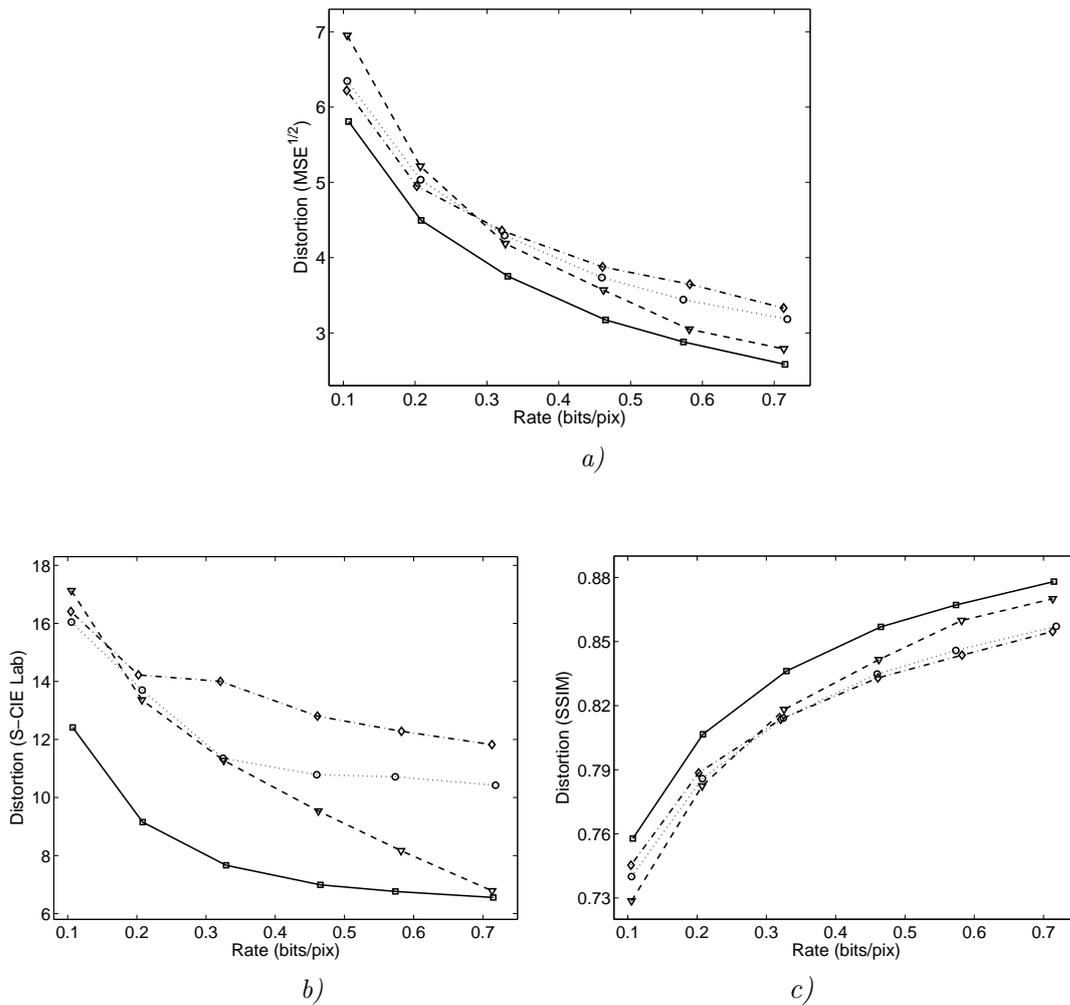


Figura 6.3: Curvas R-D, reflejan la relación existente entre la tasa de compresión alcanzada y la distorsión introducida en la imagen. Se han empleado varias medidas de distorsión: Objetiva en a) MSE y Subjetivas en b) S-CIELAB y c) SSIM, obtenidas sobre una imagen media resultado de promediar distintas imágenes. Cada gráfica incluye los resultados obtenidos con los 4 modelos perceptuales: Lineal (línea discontinua), No-lineal dependiente del punto (línea punteada), Normalización Divisiva simplificada (línea punto y raya) y Normalización Divisiva general (línea continua).

Capítulo 7

Conclusiones y trabajo de futuro

En este trabajo hemos incluido en el esquema general de JPEG2000 el modelo más actual de no-linealidades en V1 (*Normalización Divisiva* general), que hasta el momento no había sido tenido rigurosamente en cuenta en el estándar.

Hemos mostrado que, más allá de las mejoras obvias que podrían esperarse por la introducción de un modelo perceptual más refinado, existen argumentos estadísticos que indican las ventajas de la representación no-lineal utilizada frente a la representación wavelet lineal. Específicamente, hemos mostrado que una representación de este tipo obtiene unos coeficientes con menores relaciones estadísticas que los coeficientes wavelet.

Desde el punto de vista técnico hemos mostrado que dicha representación no-lineal es aplicable porque es fácilmente invertible utilizando la técnica de la expansión en serie que había sido utilizada previamente en el dominio de la DCT.

Las ventajas de la mejora propuesta se reflejan en los buenos resultados obtenidos en las imágenes codificadas, (Cap. 6), en las que se ha empleado la versión general de la *Normalización Divisiva*. Donde se consigue reducir la tasa de compresión, lo que repercute en un aumento de la calidad de la imagen, como puede verse en las curvas Rate-Distorsion.

Llegados a este punto, cabe destacar que estos buenos resultados se han obtenido empleando para la representación del color un espacio oponente tan sencillo como el YUV (*lineal*). Por tanto, resulta obvia la futura aplicación de espacios cromáticos no lineales mucho más precisos [43], que mejorarían significativamente los resultados aquí presentados.

Otra idea a desarrollar en el futuro, sería la inclusión en el Kernel de

interacción h , de los efectos de enmascaramiento entre los coeficientes de los diferentes canales cromáticos, lo que reduciría aún más la distorsión introducida en la compresión y por tanto se mejorarían aún más las curvas de R-D obtenidas.

Bibliografía

- [1] I. Epifanio, J. Gutiérrez, and J. Malo. Linear transform for simultaneous diagonalization of covariance and perceptual metric in image coding. *Pattern Recognition*, 36:1799–1811, 2003.
- [2] J. Malo, I. Epifanio, R. Navarro, and E. Simoncelli. Non-linear image representation for efficient perceptual coding. *IEEE Transactions on Image Processing*, 2005. Accepted for publication (in press). Available at <http://www.uv.es/vista/vistavalencia/>.
- [3] A. Gersho and R.M. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic Press, Boston, 1992.
- [4] D. J. Heeger. Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9:181–198, 1992.
- [5] A.B. Watson and J.A. Solomon. A model of visual contrast gain control and pattern masking. *JOSA A*, 14:2379–2391, 1997.
- [6] E.P. Simoncelli and O. Schwartz. Modeling surround suppression in V1 neurons with a statistically-derived normalization model. In M.S. Kearns, editor, *Adv. in Neural Inf. Proc. Syst.*, volume 11. MIT Press, 1999.
- [7] D.S. Taubman and M.W. Marcellin. *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Kluwer Academic Publishers, Boston, 2001.
- [8] W. Zeng, S. Daly, and S. Lei. An overview of the visual optimization tools in JPEG2000. *Sig.Proc.Im.Comm.*, 17(1):85–104, 2002.
- [9] Y. Navarro, J. Rovira, J. Gutiérrez, and J. Malo. Gain control for the chromatic channels in JPEG2000. *Proc. of the 10th Intl. Conf. AIC*, 1:539–542, 2005.
- [10] Y. Navarro, J. Gutiérrez, and J. Malo. V1 non-linearities in JPEG2000. *Enviado a: IEEE Signal Processing Letters*, 2005.

-
- [11] H.B. Barlow. Possible principles underlying the transformation of sensory messages. In WA Rosenblith, editor, *Sensory Communication*, pages 217–234. MIT Press, Cambridge, MA, 1961.
- [12] H B Barlow. Redundancy reduction revisited. *Network: Computation in Neural Systems*, 12:241–253, 2001.
- [13] D. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4(12):2379–2394, 1987.
- [14] B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.
- [15] A.B. Watson. *Perceptual Aspects of Image Coding*. In A.B. Watson, editor, *Digital Images and Human Vision*, pages 61–138, Massachusetts, 1993. MIT Press.
- [16] O. Schwartz and E.P. Simoncelli. Natural signal statistics and sensory gain control. *Nat. Neurosci.*, 4(8):819–825, 2001.
- [17] J R Cavanaugh, W Bair, and J A Movshon. Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons. *J Neurophysiology*, 88(5):2547–2556, November 2002.
- [18] W S Geisler and D G Albrecht. Cortical neurons: Isolation of contrast gain control. *Vision Research*, 8:1409–1410, 1992.
- [19] E. Peli. Contrast in complex images. *JOSA A*, 7:2032–2040, 1990.
- [20] E. Martinez-Uriegas. Color detection and color contrast discrimination thresholds. In *Proceedings of the OSA Annual Meeting ILS-XIII*, page 81, Los Angeles, 1997.
- [21] G.K. Wallace. The JPEG still picture compression standard. *Communications of the ACM*, 34(4):31–43, 1991.
- [22] S. Daly. Application of a noise-adaptive CSF to image data compression. *Opt. Eng.*, 29(8):977–987, 1990.
- [23] A.B. Watson. DCT quantization matrices visually optimized for individual images. In B.E. Rogowitz, editor, *Human Vision, Visual Processing and Digital Display IV*, volume 1913, 1993.
- [24] J. Malo, F. Ferri, J. Albert, and J. Soret. Comparison of perceptually uniform quantization with average error minimization in image transform coding. *Electronics Letters*, 35(13):1067–1068, 1999.

-
- [25] J.Malo, J.Gutierrez, I.Epifanio, F.Ferri, and J.M.Artigas. Perceptual feed-back in multigrid motion estimation using an improved DCT quantization. *IEEE Transactions on Image Processing*, 10(10):1411–1427, 2001.
- [26] D. Gabor. Theory of communication. *J. Inst. Elect. Eng.*, 93:429–457, 1946.
- [27] J. Malo. *Tecnología del Color*, chapter 4: *Almacenamiento y transmisión de imágenes en color*. Servei de Publicacions de la Universitat de València, Universitat de València, 2002.
- [28] J. Malo and M.J. Luque. *Colorlab: Matlab toolbox for color image processing*. Universitat de València, 2002. Available at <http://www.uv.es/vista/vistavalencia/software>.
- [29] E.P. Simoncelli and E.H. Adelson. *Subband Image Coding*, chapter Subband Transforms, pages 143–192. Kluwer Academic Publishers, Norwell, MA, 1990.
- [30] A. Hyvarinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, New York, 2001.
- [31] B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision Research*, 37:3311–25, 1997.
- [32] A.J. Bell and T.J. Sejnowski. The ‘independent components’ of natural scenes are edge filters. *Vision Research*, 37(23):3327–3338, 1997.
- [33] E.P. Simoncelli. Matlabpyrtools: Wavelet toolbox. 1996. Available at <http://www.cns.nyu.edu/lcv/software.html>.
- [34] D.H. Kelly. Spatiotemporal variation of chromatic and achromatic contrast thresholds. *JOSA A*, 73(6):742–749, 1983.
- [35] J. Malo, F. Ferri, J. Albert, J.Soret, and J.M. Artigas. The role of perceptual contrast non-linearities in image transform coding. *Image & Vision Computing*, 18(3):233–246, 2000.
- [36] A.B. Watson. Efficiency of a model human image code. *Journal of Optical Society of America A*, 4(12):2401–2417, 1987.
- [37] E.P. Simoncelli and B.O. Olshausen. Natural image statistics and neural representation. *Annu. Rev. Neurosci.*, 24:1193–1216, 2001.
- [38] T.M. Cover and J.A. Tomas. *Elements of Information Theory*. 1991.

-
- [39] J. Malo, A.M. Pons, and J.M. Artigas. Subjective image fidelity metric based on bit allocation of the human visual system in the DCT domain. *Image & Vision Computing*, 15(7):535–548, 1997.
- [40] B. Girod. What’s wrong with mean-squared error. In A. B. Watson, editor, *Digital Images and Human Vision*, pages 207–220. the MIT press, 1993.
- [41] X. Zhang and B.A. Wandell. Color image fidelity metrics evaluated using image distortion maps. *Signal Processing*, 70:201–214, 1998.
- [42] Z Wang, A C Bovik, H R Sheikh, and E P Simoncelli. Perceptual image quality assessment: From error visibility to structural similarity. *IEEE Trans Image Processing*, 13(4):600–612, 2004.
- [43] M.D Fairchild. *Color Appearance Models*. Addison-Wesley, New York, 1997.