



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE  
ESCUELA DE INGENIERIA

# **CUANTIFICACIÓN DE LA CALIDAD PERCEPTUAL DE IMÁGENES ESTÁTICAS**

**FRANCISCO PRIETO CIFUENTES**

Tesis presentada a la Dirección de Investigación y Postgrado  
como parte de los requisitos para optar al grado de  
Magíster en Ciencias de la Ingeniería

Profesor Supervisor:  
**MARCELO W. GUARINI H.**

Santiago de Chile, Diciembre 2008

© MMVIII, FRANCISCO PRIETO CIFUENTES



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE  
ESCUELA DE INGENIERIA

# **CUANTIFICACIÓN DE LA CALIDAD PERCEPTUAL DE IMÁGENES ESTÁTICAS**

**FRANCISCO PRIETO CIFUENTES**

Miembros del Comité:

MARCELO W. GUARINI H.

PABLO IRARRÁZAVAL M.

MARIO N. FAVA P.

MARCOS E. SEPÚLVEDA F.

Tesis presentada a la Dirección de Investigación y Postgrado  
como parte de los requisitos para optar al grado de  
Magíster en Ciencias de la Ingeniería

Santiago de Chile, Diciembre 2008

© MMVIII, FRANCISCO PRIETO CIFUENTES

*A mis Padres, que son una  
verdadera muestra del amor  
incondicional*

## **AGRADECIMIENTOS**

Esta tesis no habría sido posible sin el apoyo de los otros miembros del Centro de Imágenes Biomédicas de la Pontificia Universidad Católica de Chile, en especial los doctores Marcelo Andía y Luis Meneses , y los ingenieros Claudia Prieto, Leonardo Ramírez y Carlos Sing Long. Mención especial merecen el Director del Centro, Pablo Irrázaval Mena, el profesor supervisor, Marcelo Guarini Hermann y el Administrador del Centro, Cristián Tejos Núñez, que siempre apoyaron muy de cerca la investigación y al investigador.

## ÍNDICE GENERAL

AGRADECIMIENTOS . . . . .	iv
ÍNDICE DE FIGURAS . . . . .	vii
ÍNDICE DE TABLAS . . . . .	x
RESUMEN . . . . .	xi
ABSTRACT . . . . .	xii
1. INTRODUCCIÓN . . . . .	1
1.1. Descripción del problema . . . . .	1
1.2. Características de la visión humana . . . . .	2
1.2.1. Sensibilidad a la luminancia . . . . .	3
1.2.2. Sensibilidad a la frecuencia espacial . . . . .	4
1.2.3. Enmascaramiento por contraste . . . . .	6
1.3. Métricas de calidad de imágenes . . . . .	9
1.3.1. Modelos basados en propiedades matemáticas . . . . .	10
1.3.2. Modelos perceptuales . . . . .	16
1.4. Necesidad de una métrica representativa de la visión humana . . . . .	24
1.5. Organización de la tesis . . . . .	25
2. MÉTODOS PROPUESTOS . . . . .	26
2.1. Métrica basada en espacio de energías ponderadas normalizadas . . . . .	26
2.2. Métricas basadas en JND . . . . .	28
2.2.1. JND - Variación de contraste . . . . .	29
2.2.2. JND - Compresión . . . . .	35
2.3. Métrica basada en transformaciones lineales . . . . .	36
2.3.1. Transformación que maximice el espacio nulo . . . . .	39

2.3.2.	Cambio de dominio . . . . .	40
2.4.	Experimentos . . . . .	41
2.4.1.	Experimento 1: comparación con RMSE . . . . .	44
2.4.2.	Experimento 2: aumento gradual de las distorsiones . . . . .	44
2.4.3.	Experimento 3: diferencia entre ruido aleatorio y ruido estructurado . . . . .	44
2.4.4.	Experimento 4: percepción subjetiva . . . . .	45
3.	RESULTADOS EXPERIMENTALES Y VALIDACIÓN . . . . .	47
3.1.	Experimento 1: comparación con RMSE . . . . .	47
3.2.	Experimento 2: aumento gradual de las distorsiones . . . . .	49
3.3.	Experimento 3: diferencia entre ruido aleatorio y ruido estructurado . . . . .	51
3.4.	Experimento 4: percepción subjetiva . . . . .	53
4.	CONCLUSIONES Y TRABAJO FUTURO . . . . .	61
4.1.	Revisión general de los resultados y discusión . . . . .	61
4.2.	Revisión y discusión de cada métrica en particular . . . . .	61
4.3.	Trabajo futuro . . . . .	63
	BIBLIOGRAFÍA . . . . .	65

## ÍNDICE DE FIGURAS

1.1 Sistema Visual Humano ( <a href="http://members.fortunecity.es/alcones/imagen/ojo.gif">http://members.fortunecity.es/alcones/imagen/ojo.gif</a> ). . . . .	3
1.2 Rango de sensación subjetiva del brillo mostrando un nivel particular de adaptación (Gonzalez et al., 2001). . . . .	4
1.3 Función de Sensibilidad al Contraste. . . . .	5
1.4 Modelo de experimento visual para determinar empíricamente la CSF. . . . .	6
1.5 Ilusión de E. H. Adelson . . . . .	7
1.6 La percepción de la región central depende de la intensidad del fondo. . . . .	8
1.7 La capacidad de detección de la región central depende del fondo en que se encuentra. . . . .	8
1.8 El ruido gaussiano es menos apreciable en los bordes que en las regiones homogéneas de la imagen. . . . .	9
1.9 Ejemplo de imágenes calificadas igualmente según RMSE, pero perceptualmente distintas. . . . .	11
1.10 Imágenes perceptualmente iguales calificadas de modo distinto según el $RMSE_{rel}$ . . . . .	12
1.11 Máscara para detectar el fondo local. . . . .	18
1.12 Gráfico de la función de enmascaramiento $f_{fondo}$ . . . . .	18
1.13 Filtros orientados para detectar textura. . . . .	19
2.1 Energías Ponderadas Normalizadas. . . . .	28
2.2 Umbral bajo el cual la distorsión es invisible, según el paradigma JND. . . . .	30
2.3 Para el mismo píxel, cada nivel de contraste generará un nuevo umbral. . . . .	31
2.4 Curva JNDA: suma de todas las curvas de los umbrales. . . . .	31

2.5	Distorsión del píxel y su umbral de detección en las variaciones del contraste para los píxeles que originalmente estaban por sobre el umbral. . . . .	32
2.6	Distorsión del píxel y su umbral de detección en las variaciones del contraste para los píxeles que originalmente estaban por bajo el umbral. . . . .	32
2.7	Imagen original y una distorsionada, proveniente de un algoritmo de reconstrucción de imágenes submuestreadas de MRI. . . . .	34
2.8	Imágenes distorsionadas por efecto del truncamiento del espacio de Fourier de la imagen (anillos de Gibbs) y los respectivos mapas de distorsiones de las imágenes.	35
2.9	Histogramas de los mapas presentados en la figura 2.8. . . . .	35
2.10	Se puede descomponer $\mathbf{A}$ en la parte que no se perderá al hacer la transformación $(\mathbf{A}^I)$ y en la parte que se perderá al hacerla $(\mathbf{A}^K)$ . . . . .	38
2.11	Imágenes originales de los experimentos. . . . .	42
2.12	Imagen original y ejemplos de imágenes distorsionadas con cada tipo de degradación: borrosidad, ruido gaussiano, ruido sal y pimienta, efecto de Gibbs, inhomogeneidades de campo, traslación y rotación. . . . .	43
2.13	El experimento consistió en que se presentan 15 imágenes y el observador debe calificarlas entre 1 y 5. . . . .	46
3.1	Correlación entre el RMSE y la métrica SSIM. . . . .	47
3.2	Correlación entre el RMSE y la métrica EPN. . . . .	48
3.3	Correlación entre el RMSE y la métrica $I_{JND\_C}$ . . . . .	49
3.4	Correlación entre el RMSE y la métrica $I_{JND\_W}$ . . . . .	50
3.5	Correlación entre el RMSE y la métrica TL. . . . .	51
3.6	Correlación entre el RMSE y la métrica TLD. . . . .	52
3.7	Desempeño de la métrica basada en el Espacio de Energías Ponderadas Normalizadas para cada tipo de distorsión. . . . .	52

3.8	Desempeño de la métrica basada en IJND con variación de contraste, para cada tipo de distorsión. . . . .	53
3.9	Desempeño de la métrica basada en IJND con variación de grado de compresión, para cada tipo de distorsión. . . . .	53
3.10	Desempeño de la métrica basada en transformaciones lineales, en que se busca la transformación que maximice el espacio nulo, para cada tipo de distorsión. . . .	54
3.11	Desempeño de la métrica basada en transformaciones lineales, en que se busca la transformación que cambie el dominio a uno que tenga información perceptual, para cada tipo de distorsión. . . . .	54
3.12	Experimento que mide la sensibilidad de las métricas frente al ruido aleatorio y frente al ruido estructurado. . . . .	55
3.13	Relación de la calificación subjetiva con el RMSE. . . . .	56
3.14	Relación de la calificación subjetiva con el SSIM. . . . .	56
3.15	Relación de la calificación subjetiva con la métrica basada en EPN. . . . .	57
3.16	Relación de la calificación subjetiva con la métrica basada en JND con variación de contraste. . . . .	57
3.17	Relación de la calificación subjetiva con la métrica basada en JND con variación de compresión. . . . .	58
3.18	Relación de la calificación subjetiva con la métrica basada en transformaciones lineales en que se busca maximizar el espacio nulo. . . . .	59
3.19	Relación de la calificación subjetiva con la métrica basada en transformaciones lineales en que se cambia de dominio. . . . .	60

## ÍNDICE DE TABLAS

2.1 Para distintos grados de Gibbs se presentan los valores porcentuales de la métrica IJND_C. . . . .	34
3.1 Valores $R^2$ de los ajustes lineales entre el RMSE y las métricas. El valor más alto del ajuste lo tiene la métrica SSIM. . . . .	49
3.2 Valores de las distintas métricas para el caso de la figura 3.12. . . . .	53
3.3 Valores $R^2$ de los ajustes lineales entre la calificación subjetiva de las imágenes y cada métrica. . . . .	55

## RESUMEN

Los nuevos métodos para reconstruir imágenes médicas representan un avance importante para la capacidad de diagnóstico clínico, sin embargo es necesario validarlos midiendo la calidad de las imágenes por ellos producidos. Algo similar ocurre con las técnicas de compresión de imágenes. Ambos métodos, reconstrucción y compresión, requieren de métricas para evaluar las distorsiones que se introducen en las imágenes y el grado de esas distorsiones producto de los algoritmos. La métrica utilizada en forma estándar ha sido el error cuadrático medio (RMSE del inglés *Root Mean Squared Error*), pero esta métrica no refleja necesariamente la forma en que el sistema visual humano (SVH) percibe las imágenes. En este trabajo se han desarrollado tres nuevos enfoques alternativos al RMSE, de los que se obtuvieron cinco nuevas métricas que logran una correlación mayor con la percepción subjetiva de observadores. Estas métricas incorporan algunas características del SVH. Para validar estas métricas se realizaron experimentos que reflejan la correlación entre las métricas y la calificación subjetiva. Las métricas presentan buenos desempeños, y por su facilidad de implementación y adecuación al SVH se recomienda su uso en áreas de reconstrucción y compresión de imágenes. Además, pueden ser el punto de partida de nuevos desarrollos en esta línea de investigación.

**Palabras Claves:** calidad de imágenes, root mean squared error (RMSE), just noticeable differences (JND).

## ABSTRACT

New reconstruction methods for medical images are an important advance for clinical diagnosis ability, however they have to be validated by assessing the quality of reconstructed images. For new image compression techniques a metric for image quality assessment is also required. Both methods, reconstruction and compression, need to quantify the degree of distortion in images due reconstruction or compression algorithms. This is a difficult task because it involves both, objective measures and subjective appreciation. The Root Mean Squared Error (RMSE) has been commonly employed for this task in spite of its inability to capture truly relevant differences and its tendency to accentuate perceptually irrelevant nuances. In this work, three new approaches for image quality assessment are presented, with five new image quality metrics that incorporate some HVS characteristics. Experiments show that they are closer to the subjective opinion than the RMSE value. We recommend its use in reconstruction and compression areas due to its simple implementation and good correlation with HVS. Moreover, they can be the starting point for new developments in this research area.

**Keywords:** image quality, root mean squared error (RMSE), just noticeable differences (JND).

# 1. INTRODUCCIÓN

## 1.1. Descripción del problema

La calidad de una imagen es un concepto difícil de cuantificar, porque depende esencialmente de opiniones subjetivas de observadores. En algunas aplicaciones de procesamiento de imágenes es necesario establecer una medida objetiva de la calidad de una imagen, para lo que se suele evaluar la calidad relativa de una imagen respecto a otra de referencia, y así plantear el problema de la calidad como una comparación entre dos imágenes.

En el campo de las imágenes médicas, base fundamental en el diagnóstico de patologías y evaluación de tratamientos médicos, la comparación de imágenes permite avanzar en la investigación y evaluación de nuevos métodos de reconstrucción de imágenes para ir superando las limitaciones y problemas inherentes a las técnicas de resonancia magnética (MRI, del inglés *magnetic resonance imaging*), tomografía computada (CT, del inglés *computed tomography*), ultrasonido, etc.

Por ejemplo, el proceso de adquisición de MRI es relativamente lento y muchas veces no permite observar con precisión fenómenos dinámicos (como funciones cardíacas, flujo sanguíneo, etc.). En consecuencia, se está buscando permanentemente nuevas formas de reconstruir imágenes a partir de datos submuestreados del espacio de Fourier; de trayectorias de muestreo no cartesianas u otras alternativas que permitan optimizar y acelerar la adquisición de los datos. Las imágenes reconstruidas con estos métodos no son perfectas, por lo que se busca un balance entre rapidez de la adquisición y la cantidad de distorsiones de la imagen respecto de la imagen original sin distorsiones.

En los ámbitos de compresión de imágenes y en aplicaciones con imágenes biométricas, también existe la necesidad de comparar imágenes. En el caso de compresión de imágenes y videos, como JPEG y MPEG, respectivamente, la idea es reducir al máximo el tamaño de la imagen o película perdiendo cierta información, pero manteniendo lo más posible la calidad de la imagen. La gran mayoría de las técnicas que se han desarrollado para evaluar

la calidad de las imágenes tienen como objetivo medir la calidad de imágenes comprimidas con diversos métodos (Cadik & Slavik, 2004). Para el caso de imágenes biométricas, el objetivo es obtener una imagen que sea suficientemente buena como para obtener un determinado número de parámetros para así reconocer una huella digital, por ejemplo, pero que sea obtenida lo más fácilmente y en el menor tiempo posible (Fronthaler et al., 2006).

En todos estos casos, la forma tradicional de comparar las imágenes reconstruidas con las originales es usando el error cuadrático medio (RMSE por el inglés *root mean squared error*), que mide la distorsión píxel a píxel en el espacio de la imagen. El RMSE es muy utilizado para estos propósitos porque es sencillo de implementar y fácil de entender, pero, como veremos más adelante, no es necesariamente un fiel reflejo de lo que el hombre puede percibir de una imagen (Girod, 1993). Es por esto que nace la motivación de crear una métrica de comparación de imágenes, como el RMSE, pero tomando en cuenta cómo el Sistema de Visión Humano (SVH) califica la calidad de las imágenes (Heeger & Teo, 1995), y acercarse a una cuantificación de la calidad de la imagen basada en lo que la visión humana dice que es bueno (Manning et al., 2005).

## **1.2. Características de la visión humana**

El sistema de visión humano posee características que no son fáciles de entender. La complejidad de sus componentes, desde los ojos hasta la corteza cerebral, hace que sea prácticamente imposible construir un modelo fisiológico completo del SVH, dejándonos como herramienta de análisis los resultados de experimentos en los que se pretende caracterizar su comportamiento bajo distintas circunstancias. Los modelos del SVH que se utilizan en procesamiento de imágenes usualmente están basados en estudios psicofísicos (Kai et al., 2005).

El mismo hecho que el SVH tenga muchos componentes implica que la percepción que se tiene de una imagen no dependa exclusivamente de la recepción de luz por parte de los conos y bastones en la retina, o de la capacidad de difracción de la luz del cristalino

(figura 1.1), sino que hay un proceso, y la visión se da finalmente en la corteza cerebral. Si la visión consistiera simplemente en la recepción de los fotones, como una cámara CCD, el sólo hecho de comparar la diferencia de intensidades de la imagen sería un buen modelo del SVH, pero no lo es, como quedará claro en las secciones siguientes.

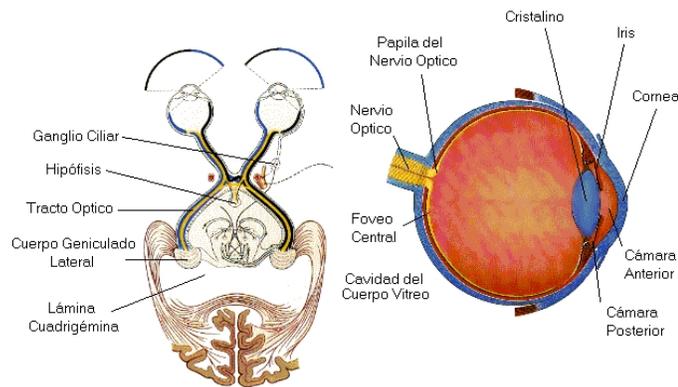


FIGURA 1.1. Sistema Visual Humano (<http://members.fortunecity.es/alcones/imagen/ojo.gif>).

Por lo general, la información contenida en imágenes médicas contemplan sólo un canal (intensidad), por lo que se suelen desplegar en escala de grises. Es por esto que el estudio de los efectos del color lo omitiremos y nos centraremos en imágenes desplegadas en escala de grises. A continuación revisaremos las características más relevantes del SVH que influyen en la percepción de imágenes.

### 1.2.1. Sensibilidad a la luminancia

La sensibilidad del SVH a la luminancia, es decir, la intensidad percibida por el hombre, es una función logarítmica (figura 1.2) de la intensidad incidente en el ojo (Gonzalez & Woods, 2001). Para compensar este efecto en el despliegue de imágenes en un monitor, lo que usualmente se hace es calibrarlo con lo que se conoce como corrección  $\gamma$ . Esto consiste en modificar la relación entre el valor numérico de un píxel y la intensidad del brillo

en que se despliega. La relación está dada por

$$I_{ij} = x_{ij}^{\gamma} \quad (1.1)$$

donde  $I_{ij}$  es la intensidad desplegada en el monitor para el píxel  $(i, j)$ ,  $x_{ij}$  es la intensidad real del píxel y  $\gamma$  es una constante que suele ir entre 1,3 y 2,7. Aplicar la corrección  $\gamma$  consiste en hacer que las intensidades desplegadas por el monitor sigan la relación inversa a la de la ecuación (1.1), es decir

$$y_{ij} = x_{ij}^{\frac{1}{\gamma}} \quad (1.2)$$

en que  $y_{ij}$  es el nuevo valor del píxel.

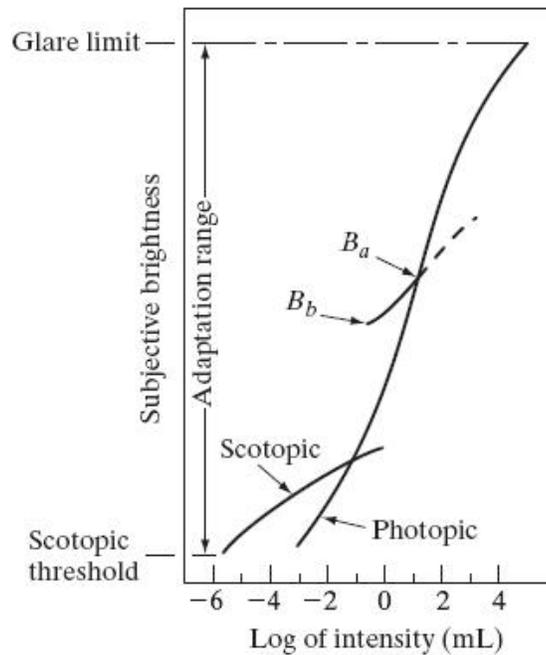


FIGURA 1.2. Rango de sensación subjetiva del brillo mostrando un nivel particular de adaptación (Gonzalez et al., 2001).

### 1.2.2. Sensibilidad a la frecuencia espacial

El SVH posee receptores de luz que tienen distintas respuestas para distintas bandas de frecuencias espaciales (Blakemore & Campbell, 1969), es decir, los receptores tienen

distintas sensibilidades para distintos cambios espaciales de luminancia. La mayor sensibilidad se encuentra para las frecuencias medias, en una banda en torno a los 10 cpd (ciclos por grado) y se atenúa a medida que las frecuencias se alejan de esta banda. Este efecto se ha simulado con la función de transferencia de la modulación (MTF del inglés *modulation transfer function*), también conocida como la función de sensibilidad al contraste (CSF del inglés *contrast sensitivity function*), que indica que el SVH tiene una característica pasabanda (figura 1.3), (Mannos & Sakrison, 1974)) (Peli, 2001). La CSF es la inversa de la función de umbral de contraste (CTF del inglés *contrast threshold function*), que es obtenida normalmente a partir de experimentos de detección de umbrales, esto es, midiendo el contraste requerido por un observador para detectar patrones sinusoidales de distintas frecuencias (figura 1.4).

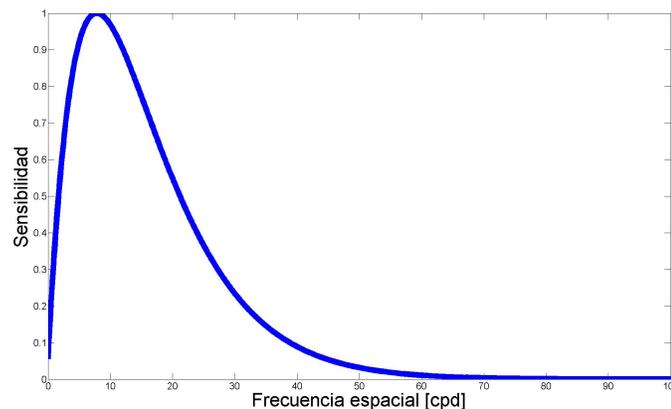


FIGURA 1.3. Función de Sensibilidad al Contraste.

La CSF en su forma completa es una función multivariada de la frecuencia espacial, frecuencia temporal (para el caso de video), la orientación, la distancia al monitor y el color. En el caso de la orientación hay más sensibilidad para las direcciones vertical y horizontal que para las oblicuas (Bradley, 1999). La CSF que se muestra en la figura 1.3 es la que se obtiene para imágenes estáticas desplegadas en escala de grises dejando fija la distancia al monitor y la orientación.

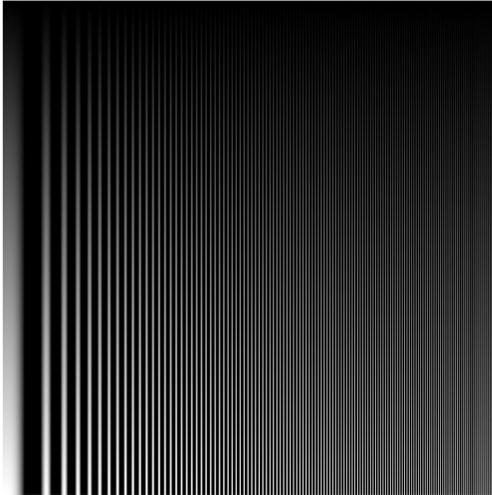


FIGURA 1.4. Modelo de experimento visual para determinar empíricamente la CSF.

### 1.2.3. Enmascaramiento por contraste

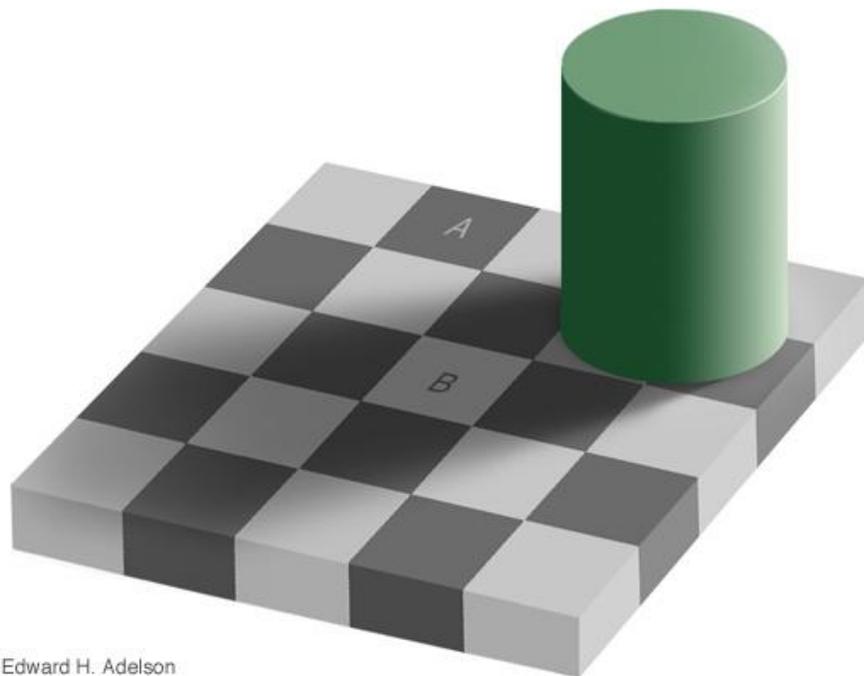
Además de los efectos nombrados anteriormente, la relación que tiene un píxel con su entorno también produce efectos en el SVH, llamados *enmascaramientos por contraste*, esto es, que bajo determinadas circunstancias hay cosas que no se ven o que se perciben de otro modo a como se despliega.

El primero de esos efectos es el producido por el valor de un píxel que está inmerso en un determinado fondo: la percepción del píxel dependerá de la intensidad del fondo. Este efecto puede llevar a percibir de modo distinto un mismo valor de gris (figuras 1.5 y 1.6) o no ser capaz de distinguirlo del fondo (figura 1.7). En el caso de la detección de un píxel respecto del fondo se produce el efecto descrito por la Ley de Weber (Gonzalez & Woods, 2001), que dice

$$\frac{\Delta I}{I} \geq k \quad (1.3)$$

donde  $I$  es la intensidad desplegada, y  $k$  es una constante, es decir, que un estímulo para ser percibido debe ser a lo menos un  $k \cdot 100\%$  más intenso que la base (el fondo para el caso de las imágenes). Esto es una regla general, para valores de  $I$  dentro de un rango acotado de intensidades. Si vemos lo que pasa para un rango más grande de intensidades, se

puede apreciar que en los fondos más oscuros se necesita un mayor contraste para detectar un píxel, mientras que para fondos muy claros se necesita menos contraste para detectarlo. Para los fondos intermedios un contraste bajo es suficiente para detectar el píxel. Un simple experimento se muestra en la figura 1.7. Para un fondo de intensidad 0 (asumiendo que estamos desplegando intensidades con 8 bits por píxel) se necesita que la región de prueba tenga una diferencia de alrededor de 20 intensidades de gris para comenzar a ser detectada. Por otro lado, si se tiene un fondo de intensidad 127 se requiere una diferencia de 1 nivel de gris para que la región sea detectada. Por último, en un fondo de intensidad 200, se requiere que una región tenga una diferencia de 6 valores de intensidad de gris para ser distinguida. Este efecto lo llamaremos *efecto de luminancia de fondo*.



Edward H. Adelson

FIGURA 1.5. En esta ilusión de E. H. Adelson, los cuadrados marcados con las letras A y B tienen la misma intensidad de gris. Se perciben distintos por el entorno de cada uno de ellos ([http://web.mit.edu/persci/people/adelson/checkershadow\\_illusion.html](http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html)).

El segundo efecto del enmascaramiento es el producido por los bordes o altas frecuencias espaciales. En una zona de bordes de una imagen es más difícil percibir distorsiones

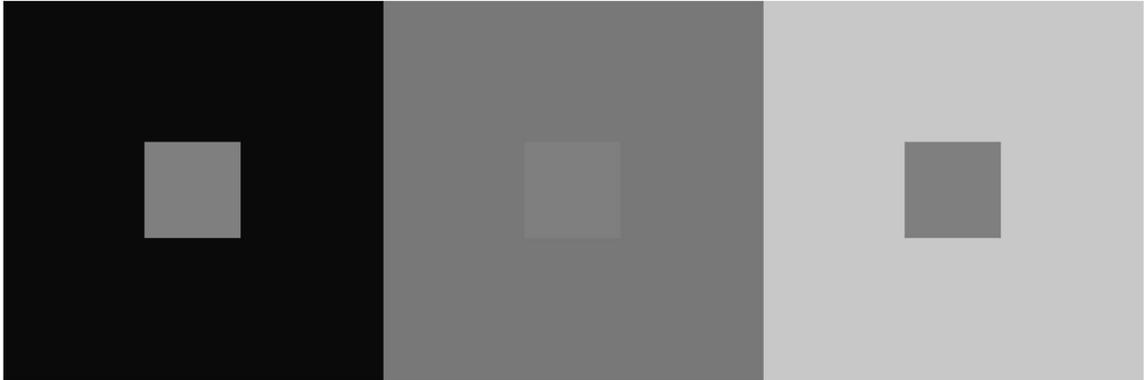


FIGURA 1.6. La percepción de la región central depende de la intensidad del fondo. La región central de las tres imágenes tiene la misma intensidad.

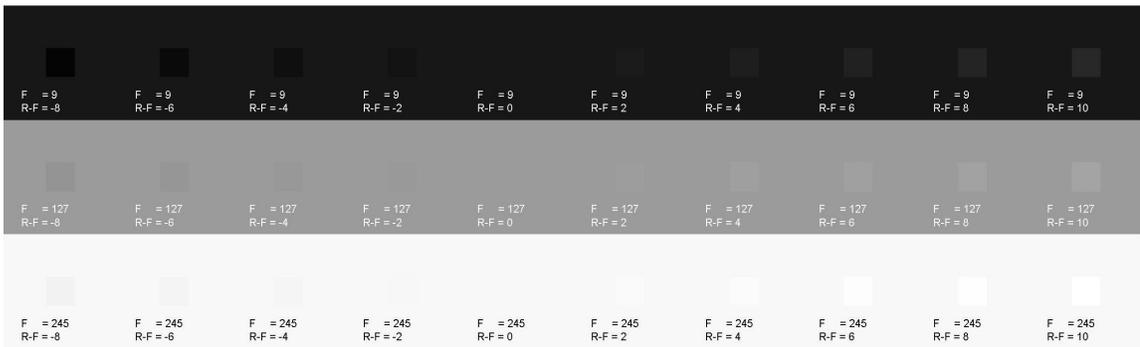


FIGURA 1.7. La capacidad de detección de la región central depende del fondo en que se encuentra. En cada cuadro,  $F$  es la intensidad del fondo, y  $R - F$  es la diferencia de intensidad entre la región central y el fondo. Para fondos oscuros se necesita mayor contraste para ser distinguido que en el caso de fondos claros. Para fondos intermedios basta un pequeño contraste para ser detectada la región. Para efectos de impresión de la figura, se aplicó una corrección con  $\gamma = 1,4$

que en una zona homogénea de la imagen. Por ejemplo, si se tiene una imagen a la que se le agrega ruido gaussiano (figura 1.8), es decir, que la nueva intensidad de cada píxel sea

$$y_{ij} = x_{ij} + r \quad (1.4)$$

donde  $y_{ij}$  es el nuevo valor del píxel,  $x_{ij}$  es la intensidad original y  $r$  es una variable aleatoria que distribuye  $N(0, \sigma^2)$  (en que la varianza  $\sigma^2$  es la misma para toda la imagen), en las zonas homogéneas el ruido se aprecia inmediatamente, mientras que en los bordes

el ruido se oculta un poco más, siendo que el ruido es el mismo en toda la imagen. Este efecto lo llamaremos *efecto de textura*.

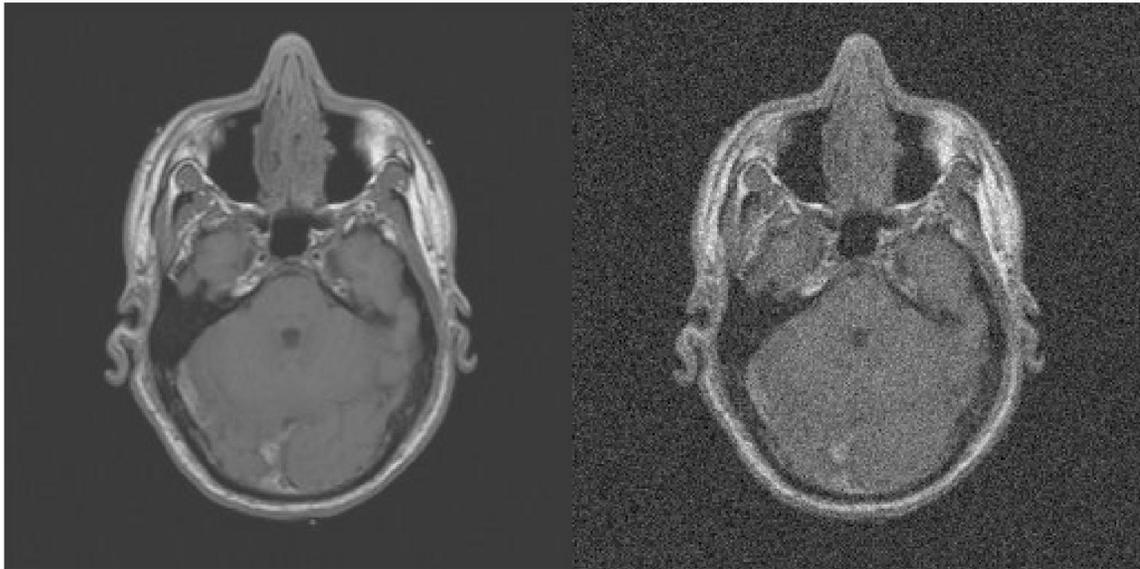


FIGURA 1.8. El ruido gaussiano es menos apreciable en los bordes que en las regiones homogéneas de la imagen.

### 1.3. Métricas de calidad de imágenes

Ante la necesidad de tener una métrica de comparación de imágenes han aparecido diversos métodos matemáticos para evaluar la calidad de las imágenes, enfocados en distintas características. En esta sección haremos un breve repaso de algunos de ellos. En su gran mayoría, la motivación de estos métodos es medir la calidad en aplicaciones de compresión de imágenes digitales, lo que es sustancialmente equivalente a comparar calidad de imágenes para otro tipo de aplicaciones.

Podríamos distinguir en esta área dos grandes líneas de desarrollo. En primer lugar tenemos los modelos que intentan evaluar la calidad mediante propiedades matemáticas de las imágenes. El más conocido de éstos es el RMSE, pero hay otros también, como el PSNR (del inglés *peak signal-to-noise ratio*) o el SSIM (*structural similarity*), presentado en (Wang & Bovik, 2002). Además, dentro de esta misma línea pero introduciéndose en

las probabilidades, se han elaborado algunos métodos estadísticos que tienen por fin la evaluación de la calidad de la imagen en función del desempeño de la ejecución de una tarea (Barrett et al., 1993).

La otra gran línea de desarrollo son los modelos perceptuales, que buscan incorporar las propiedades del SVH al modelo y así establecer una métrica de calidad de la imagen que vaya en directa consonancia con la percepción subjetiva. Los más utilizados son los que incorporan el concepto de las diferencias apenas apreciables (JND del inglés *just noticeable differences*) y el predictor de diferencias (Daly, 1992).

### 1.3.1. Modelos basados en propiedades matemáticas

#### 1.3.1.1. Métricas basadas en estadísticas simples

El RMSE es por lejos el método de comparación de imágenes más utilizado. Se puede definir de varias formas, dependiendo si se quiere tener una medida absoluta o relativa. La definición clásica es

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_i (x_i - y_i)^2} \quad (1.5)$$

en donde  $x$  e  $y$  son las dos imágenes a comparar y  $N$  es el número de píxeles de las imágenes. La ecuación (1.5) representa la distancia euclidiana entre las dos imágenes, o norma  $L_2$ , normalizada por el número de píxeles de las imágenes.

Para tener una medida que sea útil para comparar distintos pares de imágenes, se puede definir un RMSE relativo, en donde una de las dos imágenes se toma como factor de normalización.

$$\text{RMSE}_{rel} = \sqrt{\frac{\sum_i |x_i - y_i|^2}{\sum_i |x_i|^2}} \quad (1.6)$$

Esto significa que el RMSE de la imagen estudiada y es proporcional a la energía de la imagen de referencia  $x$ .

Como el concepto es fácil de entender y es además fácil de implementar, se ha transformado en el método estándar para la comparación de imágenes. Sin embargo, al no considerar las características de visión humana, en muchas ocasiones no se comporta como un buen estimador de la percepción subjetiva (Ivkovic & Sankar, 2004).

Por ejemplo, en la figura 1.9 se muestra la imagen original a la izquierda y luego dos imágenes distorsionadas que tienen un mismo valor de RMSE (calculado según la ecuación (1.6), y que sin embargo son perceptualmente muy diferentes. El caso contrario se muestra en la figura 1.10, en que la imagen de la izquierda es la original y la de la derecha está distorsionada. A pesar de verse similares, el RMSE dice que son distintas.



FIGURA 1.9. Ejemplo de imágenes calificadas igualmente según RMSE, pero perceptualmente distintas. La imagen original es la de la izquierda, mientras que la central y la de la izquierda están distorsionadas. Ambas imágenes distorsionadas tienen un valor  $RMSE_{rel}$  de 0,11 respecto de la original.

El PSNR es la razón entre la máxima potencia de la imagen y la energía del ruido de la imagen. Se suele definir en términos del RMSE en la forma de la ecuación (1.5), y en escala logarítmica, porque el rango dinámico de las imágenes suele ser grande. Con esto, el PSNR se define

$$PSNR = 20 \log_{10} \left( \frac{x_{max}}{RMSE} \right) \quad (1.7)$$

donde  $x_{max}$  es el valor del máximo píxel de la imagen  $x$ . Esta métrica tampoco considera las propiedades del SVH, y evalúa en función del RMSE, con lo que esta métrica cae en los mismos errores que el RMSE.



FIGURA 1.10. La imagen de la izquierda es la original y la de la derecha está distorsionada. El  $RMSE_{rel}$  para la segunda respecto de la original es 0,2112 a pesar de que ambas imágenes se ven similares.

Además del RMSE y PSNR se han usado otras métricas para comparar imágenes, que son evaluadas en (Eskicioglu & Fisher, 1995) según los criterios de compresión de imágenes. Entre ellas están, además de las mencionadas anteriormente, la diferencia promedio, el error absoluto, la máxima diferencia, el PMSE (*peak mean squared error*), MSE Laplaciano, histograma, etc. En el mismo trabajo se muestra que cada una de esas métricas se comporta bien para una determinada técnica de compresión, pero no necesariamente para otra. De todos modos, la mayor ventaja de estas métricas basadas en estadísticas sencillas es precisamente su simplicidad. También se muestra que combinando un modelo del SVH se puede mejorar un poco el desempeño.

### 1.3.1.2. Structural Similarity Index (SSIM)

Wang y Bovik proponen una métrica alternativa que ha sido utilizada en diversas aplicaciones de procesamiento de imágenes, llamada SSIM (*Structural SIMilarity*) (Wang & Bovik, 2002). Se basa en la combinación de tres factores: pérdida de correlación entre los píxeles de la imagen, distorsión de luminancia y distorsión de contraste. No se utiliza

explícitamente un modelo del sistema de visión humano, pero su comportamiento es significativamente mejor que el del RMSE y del PSNR para evaluar la calidad en imágenes comprimidas (Wang et al., 2004).

En su versión más sencilla, el SSIM se define como

$$\text{SSIM} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2\bar{x}\bar{y}}{(\bar{x})^2 + (\bar{y})^2} \cdot \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \quad (1.8)$$

en que el primer factor corresponde a la correlación entre las imágenes  $x$  e  $y$ ; el segundo factor corresponde a la cercanía de las luminancias de ambas imágenes, y el tercer factor mide la similitud de los contrastes. Otras derivaciones de este método han utilizado el dominio de Wavelets, como (Wang et al., 2004) o (Wang & Simoncelli, 2005).

Respecto al modo en que esta métrica evalúa la calidad, no es claro que la multiplicación entre la correlación de las imágenes, la luminancia y el contraste logre reproducir el comportamiento de la percepción visual del SVH. Aun cuando se le pudieran agregar ponderadores, éstos deberían ser determinados experimentalmente, lo que quitaría robustez al modelo.

### 1.3.1.3. Modelo estadístico

En (Barrett et al., 1993) se propone un método para evaluar la calidad de imágenes médicas basado en una tarea, es decir, según la capacidad y exactitud de hacer alguna *deducción* por parte de un observador al realizar una tarea específica a partir de la imagen. Esa deducción puede ser cualitativa, como la clasificación de una estructura entre sana o enferma, o puede ser una deducción cuantitativa, como la medición de algún parámetro en la imagen. Para que la tarea a evaluar sea una métrica válida de la calidad de la imagen, es necesario que sea cuantitativa, objetiva, escalar y calculable.

El observador que ejecuta la tarea puede ser una persona o un modelo matemático. Si es una persona, hay que evaluar la precisión en la ejecución de la tarea para distintas imágenes, y así determinar cuál es mejor. El caso del modelo matemático, en cambio, se trata de un modelo entrenado con muchas imágenes sanas y enfermas, y realiza un test

estadístico para determinar si la imagen de prueba es sana o enferma. Como la decisión es binaria, hay que comparar el resultado del test con un umbral de decisión.

La confiabilidad del método depende del tipo de modelo computacional que se utilice como observador (bayesiano o lineal). También depende del entrenamiento que se le haya aplicado al modelo, porque necesita una gran base de datos previos para lograr buenos resultados.

Estos métodos han demostrado un buen desempeño como estimador del comportamiento humano para una variedad de tareas SKE (del inglés *signal known exactly*) en que se conoce exactamente la distribución de probabilidades de la señal, y como punto de término de procesos iterativos de reconstrucción de imágenes de SPECT (por *single photom emission computed tomography*) (Gifford et al., 2000).

La necesidad de entrenamiento de este método, la capacidad computacional que se requiere para poder llevarlos a cabo y la restricción de tener que conocer exactamente la distribución de probabilidades de la señal que se está buscando hacen que no sea un método que pueda reemplazar fácilmente el uso del RMSE en la evaluación de las imágenes médicas.

#### **1.3.1.4. Otros acercamientos basados en propiedades matemáticas**

Otras alternativas han aparecido, como la presentada en (Shnayderman et al., 2006), basada en la descomposición de una matriz en sus valores singulares (SVD, del inglés *singular value decomposition*) que, aunque no contiene en su definición las características de la visión humana, se comporta, para algunas distorsiones, perceptualmente mejor que el RMSE, pero para otros no, como se ve de los resultados que el autor presenta.

En (Chandler et al., 2003) se explica un método para determinar si las distorsiones presentes en una imagen son visibles o no. Luego, para ver qué tan visibles son las distorsiones mide la distancia entre una razón señal a ruido ideal del contraste y la real, calculadas en cada banda de la descomposición de Wavelets de la imagen. Para imágenes comprimidas

con métodos basados en Wavelets tiene una correlación bastante directa con la apreciación subjetiva, pero no hay evidencia que esta métrica funciona para otros tipos de distorsiones.

Una derivación del SSIM es el *Content-Based Metric* (CBM) (Gao et al., 2005), que utiliza la información local de la estructura de las imágenes evaluadas para luego identificar regiones en estas imágenes, clasificándolos en bordes, texturas y regiones planas, dependiendo de su contenido. Luego utiliza lógica difusa para obtener la similitud de regiones correspondientes en cada una de las imágenes, y realiza una suma ponderada para obtener un índice de calidad.

La teoría de la información también ha sido usada para estos fines, como en (Sheikh et al., 2005), en que se ve la imagen de referencia como la fuente, el proceso de distorsión como el canal que impone ciertas limitaciones, y el receptor como el observador de la imagen. De esta forma se establecen los criterios de la información necesarios para una métrica de la fidelidad de una imagen y no de sus distorsiones. Una gran ventaja es que esta métrica no tiene parámetros de ajuste, sin embargo, no considera directamente efectos visuales como el enmascaramiento por contraste. El método es profundizado en (Sheikh & Bovik, 2006).

En (Suthaharan et al., 2005) se propone un método que combina información matemática con las propiedades del SVH, clasificando las regiones de la imagen en cuatro categorías dependiendo del contraste local de la imagen. Este método está fuertemente enfocado en medir el efecto de bloques producido por la compresión de imágenes, en particular mediante JPEG.

Más recientemente, en (Pang et al., 2007) se propone un método que evalúa la calidad de la imagen midiendo la energía de la distorsión estructural que hay entre dos imágenes. Cada píxel es normalizado por la energía del bloque en que se encuentra, y se calcula el producto interno de los píxeles normalizados con los píxeles originales, obteniendo así, según los autores, un valor para la energía estructural de ese bloque de la imagen. Finalmente se calcula el valor RMS (*root mean square*) de las energías estructurales de la imagen original

y la distorsionada. Los resultados presentados muestran que la correlación entre este índice y la opinión subjetiva es mayor que la del SSIM y que la del método basado en la SVD.

También en (Yao et al., 2007) se presenta un método de evaluación de calidad de imagen. Se propone calcular la similitud entre dos imágenes usando la correlación que hay entre ellas. Para obtener un índice, suman estas correlaciones normalizándolas por la desviación estándar de los coeficientes de la descomposición de Wavelets de las imágenes. El método es evaluado con imágenes comprimidas mediante JPEG y JPEG2000, aunque los resultados no son significativamente mejores que los presentados en (Wang & Bovik, 2002).

Silva, Panetta y Agaian proponen en (Silva et al., 2007) usar las medidas de mejoramiento de contraste como métrica de similitud entre imágenes. Los resultados tampoco son significativamente mejores que los presentados por otros autores.

En (Liu & Laganière, 2007) se utiliza el método de congruencia de fase (Kovesi, 2000) para evaluación de calidad de imagen. Este método es bueno para detectar bordes y características estructurales de la imagen, lo que lo hace una poderosa herramienta para las áreas de registro y segmentación, pero en nuestro caso no es una buena opción por ser demasiado insensible a características de luminancia o contraste en la imagen.

### **1.3.2. Modelos perceptuales**

Uno de los primeros intentos de usar los conceptos de la ciencia de la visión fue hecho en (Mannos & Sakrison, 1974), donde usaban una CSF para ponderar la importancia de los distintos componentes de frecuencia de una imagen. A partir de ese trabajo han surgido una serie de líneas de investigación, de las cuales destacan las que veremos a continuación.

#### **1.3.2.1. JND**

El concepto de JND (*just noticeable differences*), presentado en (Jayant, 1992) para el caso de compresión de audio, hace referencia al límite de la distorsión que pueda tener una señal sin que sea percibida por el hombre como tal. Por eso se propone la existencia de un

rango de valores en torno al estímulo para los cuales la percepción subjetiva será idéntica a la original. También se ha definido el JND como el umbral en que hay una determinada probabilidad de detectar la distorsión  $Pr = 0,5$  en (Daly, 1992), o  $Pr = 0,75$  en (Zhang et al., 2005). Esta diferencia en los umbrales no afecta el fondo de la discusión.

La idea de los JND fue introducida a imágenes en (Chou & Li, 1995), donde se propone un método para calcular estos umbrales para cada píxel bajo los cuales la distorsión es imperceptible, es decir, que si tenemos la imagen de referencia  $\mathbf{x}$ , de la que se obtienen los JND, y la imagen a evaluar y tales que se cumpla que

$$|x_{ij} - y_{ij}| \leq \text{JND}(i, j), \forall x_{ij} \in \mathbf{x}, y_{ij} \in \mathbf{y} \quad (1.9)$$

ambas imágenes se verán exactamente iguales.

El concepto ha sido utilizado en otras métricas, como el *Visual Discrimination Model* (VDM) propuesto por Lubin (Lubin, 1997) o los propuestos en (Yang et al., 2003), (Yang et al., 2005), (Zhang et al., 2008), (Yao et al., 2008) y validado en trabajos como (Vujovic et al., 2002).

En la implementación de (Chou & Li, 1995) se toman en cuenta algunas de las características del SVH, mencionadas anteriormente y se evalúa cuál de ellas es la que tiene el efecto predominante para un píxel determinado. Los dos efectos que se toman en cuenta son los producidos por el enmascaramiento por contraste y el enmascaramiento por textura.

El efecto del enmascaramiento por el fondo se modela calculando el fondo local  $bg(i, j)$  en torno al píxel, como

$$bg(i, j) = \mathbf{x} * \mathbf{B} \quad (1.10)$$

donde  $\mathbf{x}$  es la imagen original,  $\mathbf{B}$  es un filtro pasabanda de frecuencias bajas (figura 1.11) y el símbolo  $*$  es la convolución. Entonces la función del enmascaramiento debido al fondo

$f_{fondo}(i, j)$ , para imágenes con 256 niveles de grises, está dada por

$$f_{fondo}(i, j) = \begin{cases} T_0 \left( 1 - \left( \frac{bg(i, j)}{127} \right)^{1/2} \right) + 3 & \text{para } bg(i, j) \leq 127 \\ \gamma (bg(i, j) - 127) + 3 & \text{para } bg(i, j) > 127 \end{cases} \quad (1.11)$$

	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>
	<b>1</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>1</b>
<b><math>\frac{1}{32}</math></b>	<b>1</b>	<b>2</b>	<b>0</b>	<b>2</b>	<b>1</b>
	<b>1</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>1</b>
	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>

FIGURA 1.11. Máscara para detectar el fondo local.

La función  $f_{fondo}$  (ecuación (1.11)) mostrada en la figura 1.12 representa lo mostrado en la figura 1.7. Para los fondos oscuros se necesita un contraste cercano a 20 intensidades de grises para que un observador pueda distinguir una región en ese fondo. Para los fondos medios (en torno a 127 si son imágenes de 8 bits por muestra) un pequeño contraste ya es suficiente para percibir una diferencia, y ese es el mínimo que presenta el gráfico. Para los fondos muy claros también se necesita mayor contraste, aunque no tanto como para los fondos oscuros: cerca de 6 intensidades de grises es suficiente para comenzar a detectar una región inmersa en ese fondo.

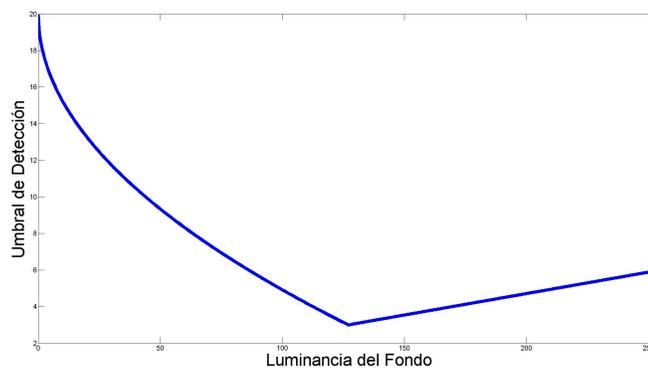


FIGURA 1.12. Gráfico de la función de enmascaramiento  $f_{fondo}$ .

Para medir el efecto de textura, es decir, lo que los bordes son capaces de enmascarar, se observan los cambios espaciales en el entorno de cada píxel, midiendo derivadas

direccionales  $\nabla_k(i, j)$ , donde  $k$  es la dirección. Cada derivada es obtenida haciendo

$$\nabla_k(i, j) = \mathbf{x} * \frac{1}{v} \mathbf{G}_k \quad (1.12)$$

donde cada  $\mathbf{G}_k$  es un filtro direccional, como los de la figura 1.13, para los que  $v = 16$ .

Con esto, se constituye una imagen  $mg$  en que el valor de cada píxel se calcula como

$$mg(i, j) = \max_k \{|\nabla_k(i, j)|\} \quad (1.13)$$

es decir, el valor del máximo gradiente orientado. De esta forma se obtienen los contrastes máximos para modelar el hecho que un mayor contraste enmascara más información.

0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	-1	0
1	3	8	3	1	0	8	3	0	0	0	0	3	8	0	0	3	0	-3	0
0	0	0	0	0	1	3	0	-3	-1	-1	-3	0	3	1	0	8	0	-8	0
-1	-3	-8	-3	-1	0	0	-3	-8	0	0	-8	-3	0	0	0	3	0	-3	0
0	0	0	0	0	0	0	-1	0	0	0	0	-1	0	0	0	1	0	-1	0
$G_1$					$G_2$					$G_3$					$G_4$				

FIGURA 1.13. Filtros orientados para detectar textura.

Experimentos visuales (Chou & Li, 1995) muestran que el umbral JND aumenta de forma lineal respecto a la derivada de la imagen, es decir, si  $mg$  es mayor, el JND será proporcionalmente mayor. Esta relación se modela con la función lineal  $f_{borde}$ , cuya pendiente  $\alpha$  es dependiente de la intensidad del fondo. De la misma forma, el umbral  $\beta$  para un contraste mínimo también varía con el fondo. Entonces, los modelos propuestos son

$$f_{borde}(mg(i, j), bg(i, j)) = mg(i, j)\alpha(bg(i, j)) + \beta(bg(i, j)) \quad (1.14)$$

donde

$$\alpha(bg(i, j)) = 0,0001 \cdot bg(i, j) + 0,115 \quad (1.15)$$

y

$$\beta(bg(i, j)) = 0,5 - 0,01 \cdot bg(i, j) \quad (1.16)$$

De este modo, el valor del JND para el píxel  $(i, j)$  es

$$\text{JND}(i, j) = \max \{f_{\text{fondo}}, f_{\text{borde}}\} \quad (1.17)$$

Así se obtiene el umbral para cada píxel bajo el cual la distorsión no se ve.

Este método es un buen estimador de lo que el SVH es capaz o no de percibir, pero no sirve para comparar la calidad de imágenes cuando ellas están por sobre el umbral (Wang et al., 2004). Es decir, si se tienen dos imágenes distorsionadas  $y_1$  e  $y_2$  tales que  $|x - y_1| > \mathbf{JND}$  y  $|x - y_2| > \mathbf{JND}$ , no se puede establecer cuál de ellas es mejor. Este problema será llamado el problema *sobre el umbral*.

Una implementación de JND basada en Wavelets se propone en (Wei et al., 1998), y una variación del método explicado se muestra en (Zhang et al., 2005). Ambos aumentan bastante la complejidad del modelo.

### 1.3.2.2. WVDP

Un método que pretende predecir las diferencias visuales que tendrán dos imágenes es el *Visible Difference Predictor* (VDP) (Daly, 1992), que fue llevado al dominio de Wavelets en (Bradley, 1999), llamándolo WVDP (*Wavelet Visible Difference Predictor*). El objetivo de este último es establecer una probabilidad de detección de la distorsión, analizando (en los coeficientes de Wavelets) si la diferencia entre ambas imágenes es apreciable o no. Para determinar esto, se asume que los coeficientes de Wavelets son una buena medida del contraste, y que por lo tanto, los coeficientes más grandes serán capaces de enmascarar más distorsión que los coeficientes pequeños.

Lo primero que hace es estimar el umbral de detección de coeficientes, es decir, el umbral bajo el cual las diferencias de contraste, representadas en Wavelets, no serán percibidas. Este umbral dependerá de la orientación y de la frecuencia espacial. Luego de esto, se busca el mínimo de los umbrales de cada banda entre las imágenes para evitar el efecto del enmascaramiento mutuo (Daly, 1992).

Con todo esto se obtiene una probabilidad de detección para cada banda, que son agrupadas para obtener una probabilidad de detección por píxel, y construir así un mapa de probabilidades.

Si consideramos un nivel de descomposición de Wavelets, el umbral de detección para los coeficientes está dado por

$$n_c(\theta, f) = \frac{y(\theta, f)}{i_\theta \cdot p_l^{2(l-1)}} \quad (1.18)$$

donde  $i_\theta$  es una constante para cada una de las orientaciones,  $p_l$  es una constante para cada banda  $l$  de descomposición, e  $y(\theta, f)$  corresponde al umbral en el espacio de la imagen, que depende de la orientación  $\theta$  y de la frecuencia espacial  $f$ , y se obtiene

$$\log(y(\theta, f)) = \log(a) + k \cdot (\log(f) - \log(g_\theta f_0))^2 \quad (1.19)$$

con  $a$ ,  $f_0$  y  $k$  constantes determinadas experimentalmente, y  $g_\theta$  un valor dependiente de la orientación.

Luego se saturan los coeficientes en el umbral, haciendo

$$\mathbf{T}_e(\theta, f, i, j) = \max \{n_c(\theta, f), b(f) \cdot |C(i, j)|\} \quad (1.20)$$

en que  $C(i, j)$  son los coeficientes de Wavelets y  $b(f)$  modela la relación que hay entre estos coeficientes y el contraste. Normalmente se puede decir que  $b(f) = 1$ .

Para evitar el efecto del enmascaramiento mutuo, se toma el mínimo de las CTF, haciendo

$$\mathbf{T}_{em} = \min \{ \mathbf{T}_e, \widehat{\mathbf{T}}_e \} \quad (1.21)$$

donde  $\mathbf{T}_e$  son los coeficientes apreciables de la imagen de referencia, y  $\widehat{\mathbf{T}}_e$  los de la imagen distorsionada.

Con esto se puede establecer la probabilidad  $P_b(i, j)$  de detectar una diferencia por cada banda, como:

$$P_b(i, j) = 1 - \exp\left(-\left|\frac{\Delta C(i, j)}{\alpha \cdot \mathbf{T}_{em}}\right|^\beta\right) \quad (1.22)$$

en que  $\Delta C(i, j)$  es la diferencia de los coeficientes de Wavelets de la imagen original y distorsionada, mientras que  $\alpha$  y  $\beta$  constantes de ajuste del modelo.

Entonces, la probabilidad de detección por cada píxel estaría dada por la combinación de las probabilidades en cada banda, es decir,

$$P_d(i, j) = 1 - \prod_b (1 - P_b(i, j)) \quad (1.23)$$

Al igual que en (Daly, 1992), el autor propone la probabilidad  $P_d = 0,5$  como el equivalente a un JND.

Este modelo tiene la ventaja que puede modelar los efectos de enmascaramiento en zonas más grandes de la imagen, dependiendo de la cantidad de descomposiciones de Wavelets que se usen, a diferencia de la implementación de (Chou & Li, 1995) que solamente es capaz de ver los efectos en una vecindad de 2 píxeles para cada lado. Sin embargo, los supuestos que hace sobre la relación que hay entre el contraste y los coeficientes de Wavelets y la forma de determinar los umbrales para cada banda dejan lugar a la discusión. Además, no soluciona el problema *sobre el umbral*.

### 1.3.2.3. Visual Image Quality Model

También se ha propuesto (Westen et al., 1995) un índice basado en un modelo multi-canal del SVH, que incluye los efectos de la sensibilidad de la luminancia, la sensibilidad a las frecuencias espaciales y enmascaramiento, trabajando con el concepto de LBC (*Local Band-Limited Contrast*) en bandas de frecuencia espacial orientadas.

Este método pretende categorizar el efecto del enmascaramiento analizando la imagen en distintas bandas de frecuencia y calculando el contraste para cada una de esas bandas según una modificación a lo propuesto en (Peli, 1990). Luego se observan las diferencias

del contraste por banda entre las dos imágenes (la de referencia y la de prueba) y se ponderan por una CTF, que elimina las diferencias de contraste que no serán apreciables por el SVH, quedándose sólo con lo que sí es perceptualmente relevante. Estas diferencias son agrupadas para obtener una medida del error perceptual (PEM, por *perceptual error measure*), que es el índice propuesto.

Matemáticamente, el PEM se obtiene evaluando:

$$\text{PEM} = \left( \sum_{x,y} \left| \sum_{k,l} |\Delta\text{MLBC}_{k,l}(x,y)|^\alpha \right|^\beta \right)^\gamma \quad (1.24)$$

en donde  $\alpha$ ,  $\beta$  y  $\gamma$  son constantes de ajuste, y  $\Delta\text{MLBC}_{k,l}(x,y)$  se define como la diferencia LBC de enmascaramiento, dada por

$$\Delta\text{MLBC}_{k,l}(x,y) = \frac{\text{LBC}_{k,l}(x,y) - \text{LBC}_{k,l}^*(x,y)}{\min(\text{TE}_{k,l}(x,y), \text{TE}_{k,l}^*(x,y))} \quad (1.25)$$

donde  $\text{LBC}_{k,l}(x,y)$  corresponde al contraste local para la banda  $k$  y la orientación  $l$  de la imagen original, y  $\text{LBC}_{k,l}^*(x,y)$  al de la imagen distorsionada.  $\text{TE}_{k,l}(x,y)$ ,  $\text{TE}_{k,l}^*(x,y)$  corresponden a las funciones de elevación del umbral (CTF) de la imagen original y distorsionada, respectivamente. Para calcular el LBC se hace

$$\text{LBC}_{k,l}(x,y) = \begin{cases} A_{k,l} \cdot \frac{B_{k,l}(x,y)}{K_{k,l} + L_{k-l}(x,y)} & \text{si } k = 2, \dots, K, l = 1, \dots, L \\ A_{k,l} \cdot \frac{B_{k,l}(x,y)}{K_{k,l} + \bar{L}_1} & \text{si } k = 1, l = 1, \dots, L \end{cases} \quad (1.26)$$

donde  $B_{k,l}(x,y)$  es la banda orientada espacialmente,  $L_k(x,y)$  es la versión pasabajos de la imagen con índice  $k$ ,  $\bar{L}_1$  es la luminancia promedio, y  $A_{k,l}$  y  $K_{k,l}$  son constantes de ajuste del modelo.

La cantidad de constantes por definir en este modelo limitan su capacidad de predecir la percepción visual humana, porque su determinación depende fuertemente de las imágenes que se estén considerando. Esto hace que en la práctica no haya sido utilizado posteriormente.

#### **1.3.2.4. Otros acercamientos basados en propiedades del SVH**

En (Teo & Heeger, 1994) se propone un método usando una descomposición piramidal de la imagen, y utiliza aproximaciones lineales para simular la respuesta de los receptores a las frecuencias espaciales y orientaciones. Este método muestra consistencia con los experimentos visuales con patrones que ellos mismos proponen, pero no se prueba con imágenes reales. Además, no soluciona el problema por sobre el umbral.

Ivkovic propone en (Ivkovic & Sankar, 2004) un método que toma en cuenta los efectos de la sensibilidad a la luminancia y de las frecuencias espaciales, pero no el del enmascaramiento por el contraste, a pesar que lo estima a través de la correlación local. Valida con diversos tipos de distorsiones, pero no compara el índice con las opiniones subjetivas de personas.

En (Fronthaler et al., 2006) se presenta un método también basado en características locales de la imagen, analizando diversos patrones de simetría que podría tener la imagen. Este método está fuertemente enfocado a las aplicaciones biométricas, como la evaluación de imágenes de huellas digitales y caras.

Estos métodos son criticados por algunos como (Shnayderman et al., 2006) o (Sutaharan et al., 2005), porque las propiedades del SVH son adaptivas, dependiendo de las condiciones de luminancia del lugar donde se muestran las imágenes, de la distancia al monitor, etc., lo que imposibilitaría a estos modelos ser una métrica objetiva.

#### **1.4. Necesidad de una métrica representativa de la visión humana**

Como se ha visto, la evaluación de la calidad de una imagen sigue siendo un problema abierto hoy en día. Hay estudios comparativos entre diversos métodos, como (Li et al., 1998), (Cadik & Slavik, 2004), (Freitas-Zampolo & Seara, 2005) y (Engelke & Zepernick, 2007), pero todavía no hay un criterio definido.

Tanto los métodos que incorporan directamente las propiedades del SVH como los que no lo hacen, tienen claro las características que un modelo debería tener, sin embargo en

la forma de representar matemáticamente esas características hay supuestos que no necesariamente se correlacionan con el SVH. Por otro lado, los métodos que se enfocan en la probabilidad de detección de distorsiones no han solucionado el problema sobre el umbral, lo que limita la evaluación de la calidad de las imágenes a las que tienen distorsiones apenas apreciables.

Además, todavía no se han aplicado masivamente los nuevos conceptos de calidad de imagen, considerando los tipos de distorsiones que pueden tener producto del método de adquisición (MRI, CT, SPECT, rayos X, ultrasonido, etc.), puesto que la gran mayoría de los métodos revisados en esta sección han sido validados para técnicas de compresión. Existen algunos intentos cercanos a la aplicación a imágenes médicas, como en (Hadhoud, 2001) o (Kai et al., 2005), pero no apuntan directamente a esta área.

### **1.5. Organización de la tesis**

En el segundo capítulo se presentan nuevos enfoques para evaluar imágenes, con aplicaciones a imágenes médicas, y la forma de evaluarlos experimentalmente. En el tercer capítulo se exponen los resultados experimentales obtenidos y en el cuarto capítulo las conclusiones de la investigación, junto a las proyecciones futuras de esta línea de trabajo.

## 2. MÉTODOS PROPUESTOS

En este capítulo se presentan tres nuevos enfoques para evaluar la calidad de imágenes, con algunas aplicaciones en imágenes fotográficas e imágenes de resonancia magnética. Dos de estos enfoques están basados en características del SVH, y el tercero en propiedades de la teoría del álgebra lineal.

A continuación se describen los tres enfoques.

### 2.1. Métrica basada en espacio de energías ponderadas normalizadas

Uno de los grandes problemas de muchas de las métricas presentadas en la sección 1.3 es que son sensibles a rotaciones y traslaciones, siendo que para el SVH una imagen que está trasladada una distancia pequeña o rotada pocos grados es prácticamente igual a la original. Por otro lado, la sensibilidad del SVH a las frecuencias espaciales es una característica que juega un rol importante en la visión. Como el espacio de Fourier es el espacio de las frecuencias y una traslación en el espacio de la imagen es un cambio de fase -y no de magnitud- en el espacio de Fourier, este espacio es un buen candidato como dominio de trabajo para buscar una métrica de calidad.

Esta métrica busca comparar regiones correspondientes del espacio de Fourier de una imagen de referencia y una distorsionada, para así establecer las diferencias entre ambas. La idea es que las diferencias que se capturen tengan correlación con las diferencias perceptuales entre las imágenes. En cada una de las regiones se calcula la energía, normalizada por el número de píxeles de la región, y se ponderan por la CSF, dando mayor importancia a las regiones que contienen componentes de frecuencias medias, que son a las que el SVH tiene mayor sensibilidad. Se comparan las diferencias de energías ponderadas y se suman las diferencias para establecer un valor numérico estimador de la calidad de la imagen distorsionada relativa a la de referencia.

Las distorsiones producen cambios en las componentes de frecuencia de la imagen, por lo que esta métrica es capaz de predecir bien la percepción subjetiva de la imagen para efectos de traslación, rotación en ángulos pequeños, entre otras. La traslación significa un cambio de fase en el espacio de Fourier y este índice toma el valor absoluto, por lo que no se ve afectada. Una rotación en la imagen implica una rotación en Fourier, y si la rotación es pequeña, la energía de cada región no cambiará significativamente.

Para introducir en el espacio de Fourier la sensibilidad a las frecuencias espaciales, dividiremos el espacio de Fourier de una imagen en  $P$  bandas de frecuencia y en  $Q$  orientaciones (figura 2.1). Si definimos  $\rho_{pq}^k$  como una región del espacio de Fourier  $\mathbf{M}^k$  de la imagen  $\mathbf{m}^k$ , para una banda  $p$  y una orientación  $q$  dados ( $p = 1, \dots, P$  y  $q = 1, \dots, Q$ ), tendremos que

$$\sum_{p,q} \rho_{pq}^k = \Psi(u, v) \mathbf{M}^k(u, v) \quad (2.1)$$

en que

$$\mathbf{M}^k = \mathcal{F} \{ \mathbf{m}^k \} \quad (2.2)$$

y  $\Psi(u, v)$  es la CSF definida en el espacio de Fourier. De esta forma podemos establecer la energía ponderada normalizada de cada imagen como

$$E_{pq}^k = \frac{1}{n_{pq}^k} \sum_{\rho_{pq}^k} |\mathbf{M}^k(u, v)|^2 \quad (2.3)$$

donde  $n_{pq}^k$  es la cantidad de píxeles que tiene la región  $\rho_{pq}^k$ , en que  $k \in \{\mathbf{A}, \mathbf{D}\}$ , siendo  $\mathbf{A}$  la imagen original o de referencia y  $\mathbf{D}$  la imagen distorsionada.

Con esto podemos comparar las regiones correspondientes  $\rho_{pq}^k$  de ambas imágenes y sumarlas para poder obtener un valor numérico de la métrica perceptual (EPN por Espacio Ponderado Normalizado de energía), haciendo

$$\text{EPN} = \sum_{p,q} \frac{E_{pq}^{\mathbf{A}} - E_{pq}^{\mathbf{D}}}{E_{pq}^{\mathbf{A}} + E_{pq}^{\mathbf{D}}} \quad (2.4)$$

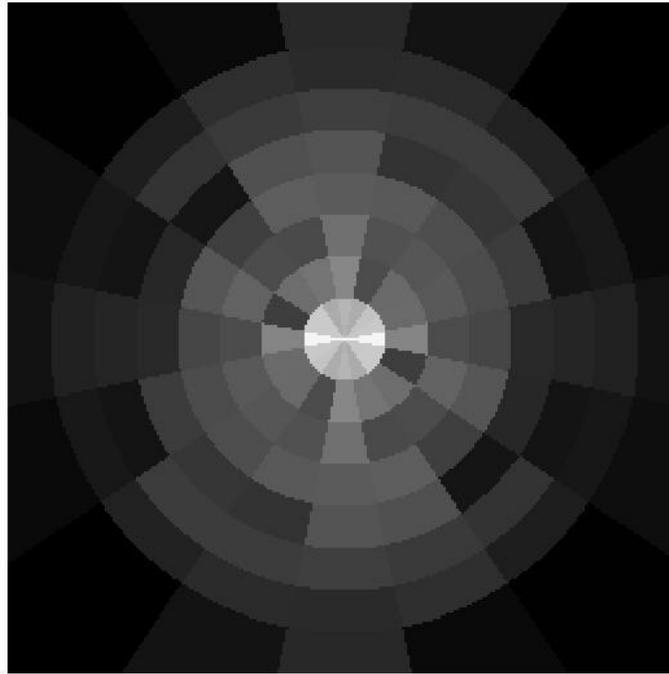


FIGURA 2.1. Energías Ponderadas Normalizadas. Las bandas de frecuencia corresponden a los anillos concéntricos y las orientaciones corresponden a las secciones de círculos. La intensidad de cada región de la imagen corresponde a la energía de esa zona dividida por la cantidad de píxeles de esa región.

Para minimizar el efecto de la luminancia, ambas imágenes tienen que estar normalizadas. En este caso se prefirió normalizarlas respecto a la media de la imagen de referencia (componente central del espacio de Fourier).

Esta métrica es fácil de implementar y contiene en su definición la sensibilidad a las frecuencias espaciales, aunque no es sensible a los efectos del enmascaramiento por el fondo ni por la textura porque no están modelados.

## 2.2. Métricas basadas en JND

Una segunda alternativa se presenta utilizando el concepto de JND. Como se describió en la sección 1.3.2.1 el principal problema de los JND es que sólo sirven para discriminar

cuándo un cambio de una imagen es visible o no. Pero si tenemos dos imágenes visiblemente distorsionadas respecto a la original, este método no puede proporcionar un estimador para discriminar cuál de ellas es mejor o está menos distorsionada. Para solucionar este problema se busca llevar progresivamente las distorsiones de una imagen hasta el umbral en que ella es apreciable, según el paradigma JND.

La idea de fondo es ir empeorando progresiva y controladamente las imágenes que se quieren comparar y analizar en cada paso si las distorsiones originales son visibles todavía o no, de tal forma que las distorsiones que se le agregan a la imagen vayan ocultando las distorsiones originales. Si una distorsión es siempre visible, aunque la imagen esté muy deteriorada, quiere decir que es una distorsión muy fuerte, por lo que la imagen será mala. Si la distorsión deja rápidamente de ser apreciable, la imagen no es tan mala.

Hay muchas formas en que se pueden ir empeorando progresivamente las imágenes, por ejemplo, agregándole ruido, haciéndolas más borrosas, etc. Lo que hay que tener presente es que el modo de degradación tenga algún sentido directo con la percepción subjetiva del SVH. En este trabajo utilizaremos dos métodos: disminución paulatina de contraste y aumento de grado de compresión por Wavelets.

La implementación de JND que se utilizó en este trabajo fue la propuesta por (Chou & Li, 1995), a la que nos referiremos en adelante como el IJND, puesto que se calcula en el espacio de la imagen.

### **2.2.1. JND - Variación de contraste**

Si tenemos una imagen con algún tipo de distorsión, esa distorsión será menos visible si se disminuye el contraste a la imagen. Si el contraste se disminuye demasiado, la imagen será plana y evidentemente no se verá la distorsión. De este modo podemos disminuir el contraste en torno a la media de la intensidad de la imagen y ver en qué momento la distorsión deja de ser percibida.

Para explicar el método, veremos lo que ocurre para un píxel en particular de la imagen. Si tenemos dos imágenes, la original y la distorsionada, a partir de la original obtenemos el umbral JND para ese píxel, es decir, la cantidad de distorsión que se le puede agregar (o quitar) a ese píxel sin que sea perceptualmente relevante (figura 2.2). Luego, si se disminuye un poco el contraste y se recalcula el valor del JND para ese píxel, este valor será distinto al original, puesto que las condiciones del fondo y la textura cambian. Si se hace este proceso repetidas veces, se tendrá un umbral para cada nivel de contraste (figura 2.3), hasta que la imagen quede completamente plana.

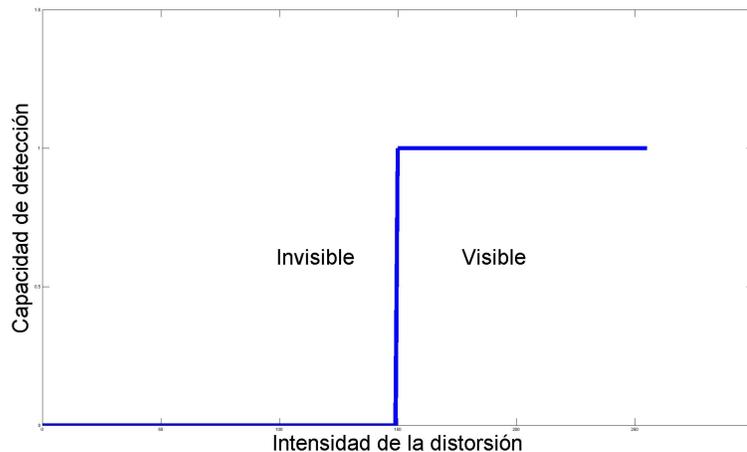


FIGURA 2.2. Umbral bajo el cual la distorsión es invisible, según el paradigma JND.

Los escalones obtenidos por los umbrales (figura 2.3) significan que una distorsión será visible si se encuentra por sobre el umbral (más a la derecha). Si se suman todos estos escalones y luego se normaliza para que quede la curva entre 0 y 1, se obtendrá una curva como la de la figura 2.4, que llamaremos *JND Acumulado* (JNDA), que indica que una distorsión de  $x$  intensidades de grises es visible en un  $r\%$  de los niveles de contraste.

Como se pudo ver en la figura 1.12, la percepción de un píxel inserto en un fondo determinado no es lineal respecto del fondo ( $f_{fondo}$ ). Al estar modificando el contraste, cada uno de estos píxeles se estará moviendo a lo largo de esa curva, por lo que su comportamiento

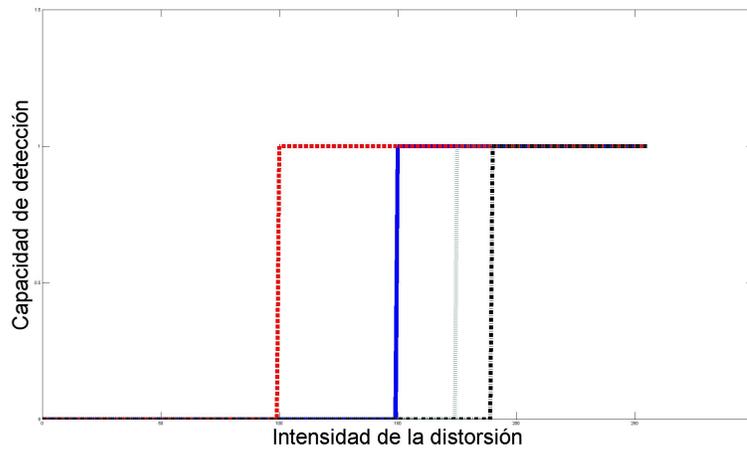


FIGURA 2.3. Para el mismo píxel, cada nivel de contraste generará un nuevo umbral. Cada curva proviene de un nivel distinto de contraste.

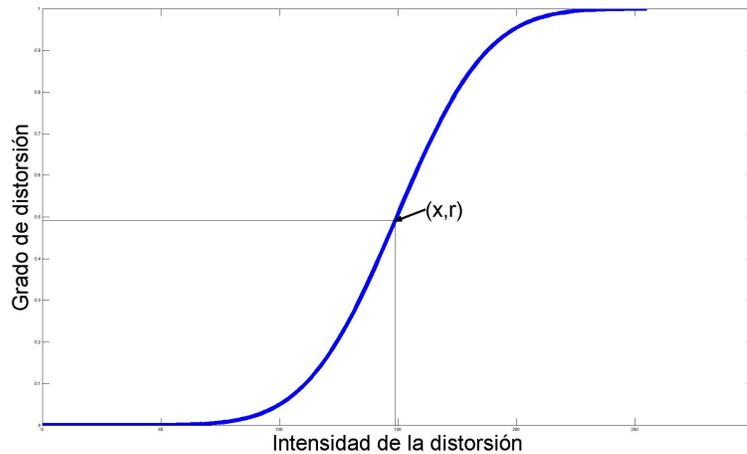


FIGURA 2.4. Curva JNDA: suma de todas las curvas de los umbrales. Una distorsión de  $x$  intensidades de grises es visible en un  $r\%$  de los niveles de contraste.

tendrá la misma forma no lineal respecto a los grados de contraste que tenga la imagen. En las figuras 2.5 y 2.6 la línea recta (sólida) representa la variación de la distorsión del píxel seleccionado a medida que disminuye el contraste. La curva punteada representa el valor JND de ese píxel determinado. Juntando ambas curvas se tiene que el espacio se divide en la región visible (bajo la curva sólida) e invisible (sobre la curva), es decir, el tramo en que

la curva punteada está por sobre la sólida quiere decir que el valor JND es mayor que la distorsión, por lo que la distorsión no será apreciada. En las zonas en que la curva punteada es menor que la curva sólida ocurre lo contrario, la distorsión es mayor al valor JND, es decir, la distorsión será visible.

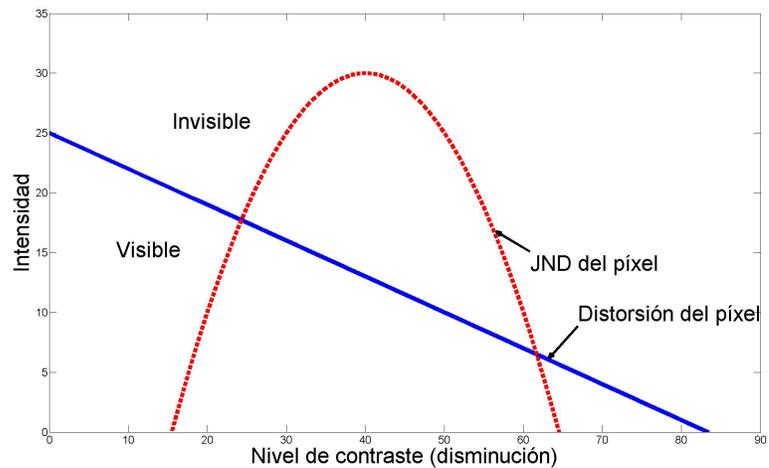


FIGURA 2.5. Distorsión del píxel y su umbral de detección en las variaciones del contraste para los píxeles que originalmente estaban por sobre el umbral.

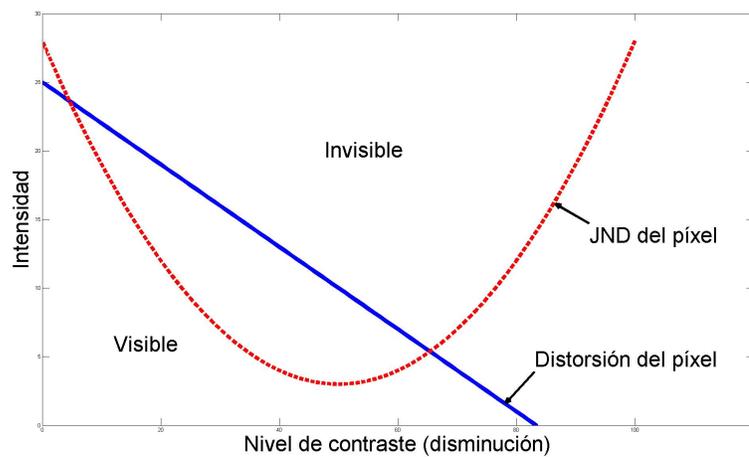


FIGURA 2.6. Distorsión del píxel y su umbral de detección en las variaciones del contraste para los píxeles que originalmente estaban por bajo el umbral.

Para estas variaciones de contraste, sólo se tomarán en cuenta los píxeles que originalmente eran visibles, es decir, que  $|x_{ij} - y_{ij}| > \mathbf{JND}_{ij}$ , con lo que nos podemos quedar sólo con las que tienen un comportamiento como el de la figura 2.5. El hecho que la curva del JND cruce dos veces la curva de la distorsión del píxel significa que para los primeros contrastes la distorsión es visible y, a medida que baja el contraste, deja de ser visible y luego comienza a serlo nuevamente. Nosotros haremos el supuesto que sólo el primer cruce es el relevante, puesto que si se comienza a ver de nuevo el píxel quiere decir que ha cambiado su relación con respecto al fondo (estamos en la otra parte de la curva de la figura 1.12), y ese efecto es producido por esta métrica y no propio del píxel en cuestión. De esta forma, podemos afirmar de la figura 2.4 no sólo que se trata de la distorsión que se vería en un  $r\%$  de los niveles de contraste, sino que es el primer  $r\%$  de los niveles de contraste.

En consecuencia, la curva de la figura 2.4 puede ser utilizada como predictor de las distorsiones de la imagen a evaluar. Se mide la diferencia píxel a píxel entre las imágenes y esa diferencia se usa como dato de entrada en la curva JNDA, con lo que se obtendrá un valor para cada píxel que diga en qué porcentaje de contrastes de la imagen esa distorsión es visible, y ese valor se utiliza como estimador de la calidad perceptual del píxel puesto que determina su grado de distorsión.

Si se hace para toda la imagen, se obtiene un mapa de calidad de la imagen (figura 2.7) que representa los grados de distorsión perceptualmente relevante que tiene la imagen.

Para obtener un valor de estimación de la calidad de la imagen a partir de la información contenida en el mapa, se puede observar su histograma, que dice cuántos píxeles tienen distorsión invisible y cuántos tienen distorsión visible en grados entre 0 y 1. Si los separamos en tres grupos, los que tienen distorsión invisible (los que están en cero), los que tienen máxima distorsión (están en uno) y los que están entre cero y uno, podemos estimar un valor para la calidad de la imagen dividiendo el número de píxeles que son muy visibles por los que son invisibles. Llamaremos a esta métrica  $IJND_C$ , por utilizar los  $IJND$  con

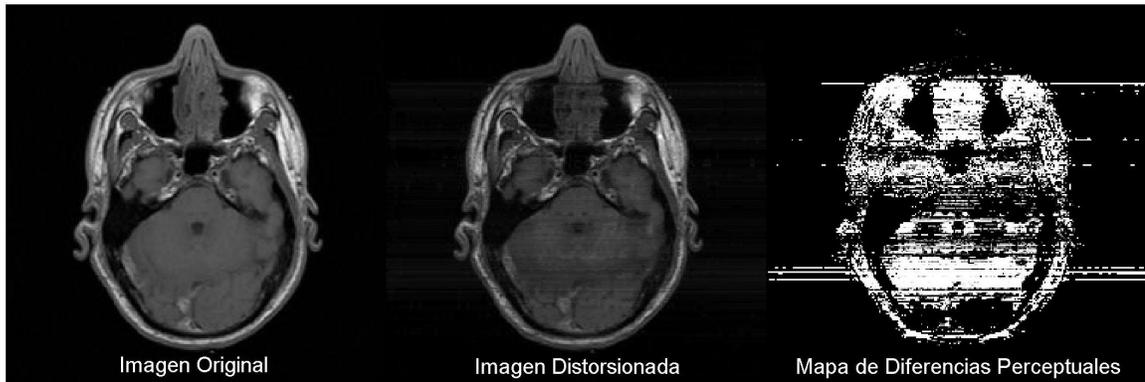


FIGURA 2.7. Imagen original y una distorsionada, proveniente de un algoritmo de reconstrucción de imágenes submuestreadas de MRI. A la derecha, el mapa en que la intensidad de cada píxel corresponde al valor obtenido de la curva JNDA.

variación de contraste.

$$I_{JND\_C} = \frac{\text{n}^\circ \text{ de píxeles con visibilidad igual a 1}}{\text{n}^\circ \text{ de píxeles con visibilidad igual a 0}} \quad (2.5)$$

De esta forma, si el valor es 0, las imágenes serán iguales, y a medida que aumenta, significará que hay más diferencias entre ellas.

Un ejemplo para analizar el desempeño de esta métrica es estudiar las distorsiones producto del efecto que tiene adquirir sólo las muestras del centro en el espacio de Fourier, como si fuera un proceso de adquisición de MRI. Los mapas de la figura 2.8 indican el grado de distorsión del píxel, información que también está contenida en sus histogramas (figura 2.9). En la tabla 2.1 se presentan los valores de donde se obtiene el valor de la métrica  $I_{JND\_C}$ , correspondiente a la ecuación (2.5).

% espacio-k	Grados de distorsión			$I_{JND\_C}$
	distorsión = 0	0 < distorsión < 1	distorsión = 1	
6, 67	0, 6893	0, 0415	0, 2692	0, 3905
13, 33	0, 7693	0, 0480	0, 1826	0, 2373
20	0, 8121	0, 0443	0, 1436	0, 1768
26, 67	0, 8448	0, 0438	0, 1114	0, 1319
33, 33	0, 8713	0, 0407	0, 0880	0, 1010

TABLA 2.1. Para distintos grados de Gibbs se presentan los valores porcentuales de la métrica  $I_{JND\_C}$  para las imágenes de la figura 2.8.

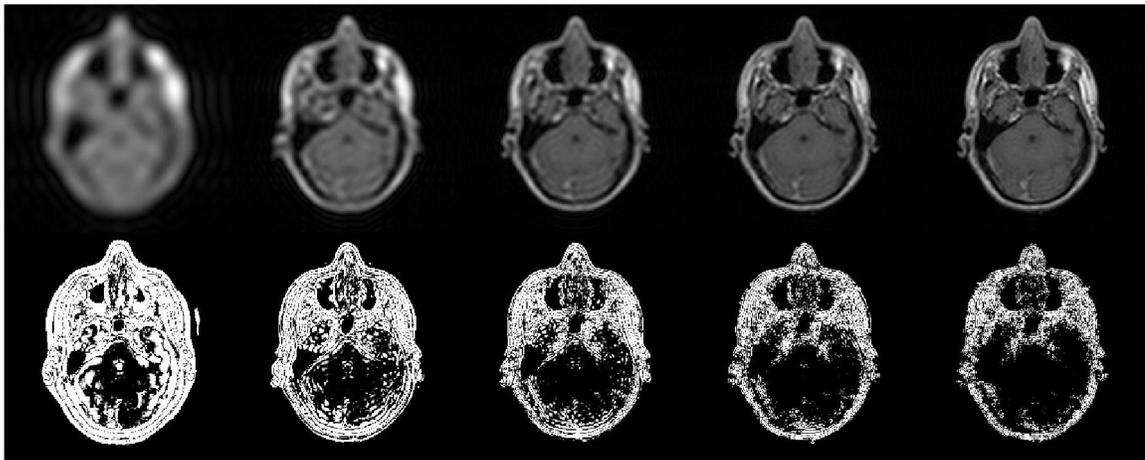


FIGURA 2.8. En la primera fila, imágenes distorsionadas por efecto del truncamiento del espacio de Fourier de la imagen (anillos de Gibbs). En la segunda fila están los respectivos mapas de distorsiones de las imágenes.

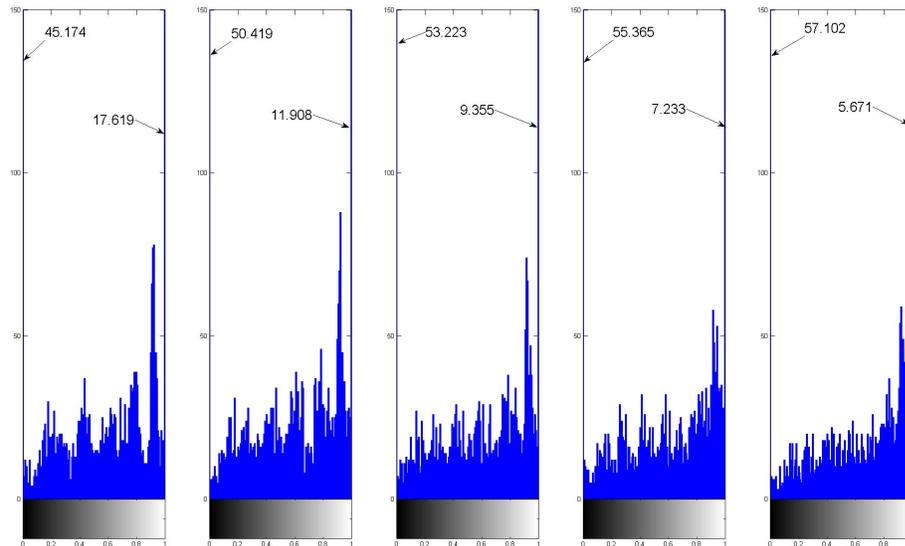


FIGURA 2.9. Histogramas de los mapas presentados en la figura 2.8.

### 2.2.2. JND - Compresión

El otro medio de degradación de las imágenes que se utilizó fue la compresión hecha en el dominio de Wavelets. Dentro de todos los tipos de compresión que existen, aquí se

utilizó uno muy sencillo, que consiste en truncar los coeficientes de Wavelets sin disminuir sus niveles de cuantización. Se hizo así porque se asumió que los coeficientes de Wavelets son una buena estimación del contraste local de la imagen y sólo se quería medir el efecto de la eliminación progresiva de los contrastes que no son apreciables. Si se modifican también los niveles de cuantización de los coeficientes, estaríamos bajando resolución a los valores de los contrastes y por lo tanto se modificarán, lo que implica por ahora un efecto no deseado. En este acercamiento no queremos modificar los contrastes que sí son apreciables por el SVH.

Con este medio de degradación se obtienen curvas equivalentes a las de las figuras 2.2 a 2.7 y se puede establecer un modo de analizar los datos análogo a la ecuación (2.5), que llamaremos  $I_{JND\_W}$  por tratarse del dominio de Wavelets:

$$I_{JND\_W} = \frac{\text{n}^\circ \text{ de píxeles con visibilidad igual a 1}}{\text{n}^\circ \text{ de píxeles con visibilidad igual a 0}} \quad (2.6)$$

### 2.3. Métrica basada en transformaciones lineales

Esta propuesta se basa en la teoría de álgebra lineal, es decir, se podría agrupar entre las medidas de calidad que buscan propiedades matemáticas de las imágenes, aunque como se verá, también incorpora algunas de las propiedades del SVH.

Si tenemos una imagen  $\mathbf{D}$  de  $N$  píxeles, desplegada como un vector de  $N \times 1$ , que queremos compararla con una imagen de referencia  $\mathbf{A}$ , también de  $N$  píxeles, desplegada como un vector de  $N \times 1$ , podemos expresar la imagen  $\mathbf{D}$  como una transformación lineal  $\mathbf{T}$  de  $\mathbf{A}$ . Así, tenemos que

$$\mathbf{D} = \mathbf{TA} + \mathbf{B} \quad (2.7)$$

en que  $\mathbf{B}$  es del mismo tamaño que  $\mathbf{A}$ . Un primer supuesto será que las distorsiones no afectan a la imagen nula (es decir que  $\mathbf{A} = \mathbf{0} \Rightarrow \mathbf{D} = \mathbf{0}$ ), por lo que podemos decir que  $\mathbf{B} = \mathbf{0}$ .

También podemos asumir que en la transformación  $\mathbf{T}$  se pierde información, aunque no toda, con lo que tendrá asociados un espacio nulo (o kernel) y un espacio no nulo (o espacio de la imagen). El kernel de la transformación  $\mathbf{T}$  ( $\text{Ker}(\mathbf{T})$ ) se define como

$$\text{Ker}(\mathbf{T}) = \left\{ \mathbf{v} \in \mathbb{R}^N : \mathbf{T}\mathbf{v} = \vec{0} \right\} \quad (2.8)$$

con los que se puede construir una base ortonormalizada de  $\text{Ker}(\mathbf{T})$ , es decir

$$\text{Ker}(\mathbf{T}) = \text{span} \{ \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k \} \quad (2.9)$$

en que

$$\langle \mathbf{v}_m, \mathbf{v}_n \rangle = \delta_{mn} \quad (2.10)$$

donde  $\langle \cdot, \cdot \rangle$  es el producto interno o producto punto entre los vectores y  $\delta_{mn}$  es el delta de Krönecker.

De la misma forma se define el espacio de la imagen de la transformación  $\mathbf{T}$  ( $\text{I}(\mathbf{T})$ )

$$\text{I}(\mathbf{T}) = \left\{ \mathbf{u} \in \mathbb{R}^N : \mathbf{T}\mathbf{u} \neq \vec{0} \right\} \quad (2.11)$$

con los que se puede construir una base ortonormalizada de  $\text{I}(\mathbf{T})$ , es decir,

$$\text{I}(\mathbf{T}) = \text{span} \{ \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n \} \quad (2.12)$$

donde

$$\langle \mathbf{u}_m, \mathbf{u}_n \rangle = \delta_{mn} \quad (2.13)$$

De esta forma tenemos los espacios  $\text{Ker}(\mathbf{T})$  e  $\text{I}(\mathbf{T})$  que son ortogonales entre sí, es decir

$$\begin{aligned} \text{I}(\mathbf{T}) \cup \text{Ker}(\mathbf{T}) &= \mathbb{R}^N \\ \text{I}(\mathbf{T}) \cap \text{Ker}(\mathbf{T}) &= \phi \end{aligned} \quad (2.14)$$

tales que

$$\langle \mathbf{v}_i, \mathbf{u}_j \rangle = 0, \forall i, j \quad (2.15)$$

Con esto podemos descomponer  $\mathbf{A}$  como una combinación lineal de vectores  $\mathbf{v}$  y  $\mathbf{u}$  haciendo

$$\mathbf{A} = \sum_i \eta_i \mathbf{v}_i + \sum_j \xi_j \mathbf{u}_j \quad (2.16)$$

en que el primer término representa la parte de la imagen que se perderá al aplicar la transformación  $\mathbf{T}$ , a lo que llamaremos  $\mathbf{A}^K$ , y el segundo término representa la parte de la imagen original que no se perderá al aplicar la transformación  $\mathbf{T}$ , a lo que llamaremos  $\mathbf{A}^I$  (figura 2.10). Las constantes  $\eta_i$  y  $\xi_j$  son los coeficientes de las bases del kernel y de la imagen, y se obtienen de:

$$\begin{aligned} \eta_i &= \langle \mathbf{A}, \mathbf{v}_i \rangle, \forall i \\ \xi_j &= \langle \mathbf{A}, \mathbf{u}_j \rangle, \forall j \end{aligned} \quad (2.17)$$



FIGURA 2.10. Se puede descomponer  $\mathbf{A}$  en la parte que no se perderá al hacer la transformación ( $\mathbf{A}^I$ ) y en la parte que se perderá al hacerla ( $\mathbf{A}^K$ ).

Para encontrar la parte de la imagen que se perdió en la transformación y a la que el SVH es sensible, se define la proyección del  $\text{Ker}(\mathbf{T})$  en el espacio de la parte de la imagen  $\mathbf{A}$  que es perceptualmente relevante como

$$\begin{aligned} \Phi &= \sum_i \alpha_i \mathbf{v}_i \\ \alpha_i &= \langle \mathbf{v}_i, \mathbf{W}^{-1} \Psi \mathbf{W} \mathbf{A}^K \rangle \end{aligned} \quad (2.18)$$

en que  $\mathbf{W}$  es la matriz para calcular la transformada de Fourier (FT) y  $\Psi$  es la CSF, definida en el espacio de Fourier. Además se asume en la ecuación (2.18) que el SVH es únicamente sensible a las frecuencias espaciales, efecto modelado por la CSF. Esta imagen

$\Phi$  es un mapa de diferencias entre la imagen original y la distorsionada, que muestra la parte perceptualmente relevante de la imagen original que se perdió al hacer la transformación.

Como se puede observar, la matriz  $\mathbf{T}$  será de  $N \times N$  elementos, y sin embargo es necesario obtenerla a partir de un par de imágenes de  $N$  elementos, es decir, tenemos un sistema de ecuaciones de  $N^2$  incógnitas y  $N$  ecuaciones, con lo que quedan  $N^2 - N$  grados de libertad. Por esto, es necesario hacer algunos supuestos respecto a  $\mathbf{T}$ . Algunas alternativas se verán a continuación.

### 2.3.1. Transformación que maximice el espacio nulo

Haremos el supuesto que si escogemos la transformación que pierde mayor información entre la imagen original y la distorsionada, estaremos buscando la transformación que tiene el espacio nulo de mayor dimensión, es decir, si  $\mathbf{T}$  es de dimensión  $N$ , la máxima dimensión que puede tener  $\text{Ker}(\mathbf{T})$  es  $N - 1$ . Si asumimos esto, la solución para  $\mathbf{T}$  será

$$\mathbf{T} = \mathbf{D} \frac{\mathbf{e}^*}{\langle \mathbf{e}, \mathbf{A} \rangle} \quad (2.19)$$

que a su vez será la transformación  $\mathbf{T}$  que minimice  $\|\mathbf{T}\|^2$ , que es única. El vector  $\mathbf{e}$  es una imagen de referencia que determina el espacio de la imagen de  $\mathbf{T}$ , es decir, es una forma de determinar la información que no se pierde en la transformación  $\mathbf{T}$ . Un extremo es decir que  $\mathbf{e} = \mathbf{A}$ , con lo que se afirma que no se pierde información en la transformación (pero hemos dicho que se pierde). El otro extremo es decir que  $\mathbf{e} = \mathbf{D}$ , lo que implica que  $\langle \mathbf{e}, \mathbf{A} \rangle = 0$ , pero hemos supuesto que no toda la información se pierde. Entonces queda abierta la elección de  $\mathbf{e}$ . Se podría escoger un  $\mathbf{e}$  suficientemente cercano a  $\mathbf{A}$ , es decir, que formen un ángulo pequeño, lo que equivaldría a establecer una imagen que tuviera valores para los píxeles muy parecidos a la original, por ejemplo que tuviera un valor RMSE pequeño, puesto que el ángulo entre  $\mathbf{A}$  y  $\mathbf{e}$  ( $\theta_{\mathbf{A}\mathbf{e}}$ ) se calcula

$$\theta_{\mathbf{A}\mathbf{e}} = \arccos \left( \frac{\|\mathbf{A}\| \|\mathbf{e}\|}{\langle \mathbf{A}, \mathbf{e} \rangle} \right) \quad (2.20)$$

Una segunda opción es escoger el promedio entre  $\mathbf{A}$  y  $\mathbf{D}$ . En este trabajo se utilizó esta alternativa para determinar  $\mathbf{T}$ .

Finalmente, para obtener un índice numérico de la métrica perceptual presentada en esta sección (TL), calcularemos

$$TL = \frac{\|\Phi\|_2}{N} \quad (2.21)$$

y así se cuantificará la energía de la información perceptualmente relevante que se perdió al hacer la transformación, normalizada por el número de píxeles.

### 2.3.2. Cambio de dominio

Otra alternativa es escoger una transformación  $\mathbf{T}$  que contenga información perceptual. Si  $\mathbf{T}$  es de la forma

$$\mathbf{T} = \mathbf{\Pi}^{-1}\mathbf{G}\mathbf{\Pi} \quad (2.22)$$

en que  $\mathbf{\Pi}$  es una transformación a algún dominio del que se pueda obtener alguna información perceptual de la imagen, podemos evaluar la ecuación (2.3) de la forma

$$\mathbf{\Pi D} = \mathbf{G}\mathbf{\Pi A} \quad (2.23)$$

Además, si decimos que  $\mathbf{G}$  debe ser diagonal, estamos calculando la relación elemento a elemento de este nuevo espacio  $\mathbf{\Pi D}$  con  $\mathbf{\Pi A}$ . Evidentemente se abre un amplio panorama sobre las posibles transformaciones  $\mathbf{\Pi}$  que se pueden escoger. En este trabajo se utilizará el dominio de Wavelets.

En este caso,  $\mathbf{G}$  es de la forma

$$\mathbf{G} = \begin{pmatrix} g_1 & 0 & \dots & 0 \\ 0 & g_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & g_N \end{pmatrix} \quad (2.24)$$

en que cada elemento  $g_i$  de la diagonal es

$$g_i = \frac{C^D(i)}{C^A(i) + C^D(i)}, i = \{1, 2, \dots, N\} \quad (2.25)$$

donde  $C^k(i)$  es el coeficiente  $i$  de la transformación de Wavelets de la imagen  $k$ , expresada como vector de  $1 \times N$ , y  $k \in \{\mathbf{A}, \mathbf{D}\}$ . Con esto, la métrica simplemente se reduce a analizar los coeficientes de Wavelets.

Si se quiere obtener un valor como índice de la métrica perceptual presentada en esta sección (TLD), se puede calcular

$$\text{TLD} = \frac{\|\mathbf{G}\|_2}{N} \quad (2.26)$$

siendo nuevamente la energía de la información relevante normalizada por el número de píxeles.

## 2.4. Experimentos

Para la validación, se realizaron cuatro tipos de experimentos. En primer lugar se realizaron experimentos comparativos del desempeño de cada métrica con el RMSE. En segundo lugar, experimentos que evalúan el comportamiento de cada una de las métricas frente a distintos tipos de distorsiones de las imágenes. En tercer lugar se hizo una experiencia agregando ruido aleatorio en una imagen, y luego ruido estructurado en una imagen médica, simulando la aparición de un tumor, para ver si las métricas son sensibles a esta diferencia. Por último, en cuarto lugar, se hizo un experimento comparativo del desempeño de las distintas métricas con la opinión subjetiva de personas.

Para los experimentos se utilizaron dos imágenes: una imagen fotográfica estándar (figura 2.11 izquierda), que llamaremos *Bárbara*, y otra procedente de MRI (figura 2.11 derecha), que llamaremos *MRI-Cabeza*.

Para obtener las imágenes degradadas en los distintos experimentos, se utilizaron siete tipos de distorsiones que se aplicaron a las imágenes originales en distinto grado. Estas distorsiones son:



FIGURA 2.11. Imágenes originales de los experimentos.

- (i) borrosidad: se le pasa un filtro pasabajos gaussiano a la imagen, en que se modifica la varianza del filtro. Para todos los casos se utilizó el mismo tamaño del filtro (25 píxeles).
- (ii) ruido gaussiano: a la intensidad de cada píxel se le agrega una variable aleatoria que distribuye  $N(0, \sigma^2)$ , en que se modifica el parámetro  $\sigma^2$ .
- (iii) ruido sal y pimienta: a algunos píxeles, elegidos aleatoriamente según una distribución uniforme en toda la imagen, se le agrega una variable aleatoria que distribuye  $U(-b, b)$ . Los parámetros son la cantidad de píxeles ruidosos y la intensidad del ruido ( $b$ ).
- (iv) efecto de Gibbs: simulando el efecto que produce en MRI adquirir sólo una parte del espacio de Fourier de la imagen, este efecto consiste en eliminar parte de la información en Fourier de la imagen, dejando una banda en el centro. El parámetro es la porción del espacio de Fourier que se mantiene, es decir, el ancho de la banda.

- (v) simulación de inhomogeneidades de campo: en MRI algunas secuencias de adquisición de imágenes introducen una fase dependiente del tiempo en los coeficientes de Fourier, y por tanto la secuencia que se utilice para adquirir los datos determinará los tipos de distorsiones que se introducirán. En este trabajo se modeló una secuencia EPI (del inglés *Echo Planar Imaging*) (Stehling et al., 1991).
- (vi) traslación: traslación de unos pocos píxeles.
- (vii) rotación: rotación de unos pocos grados.

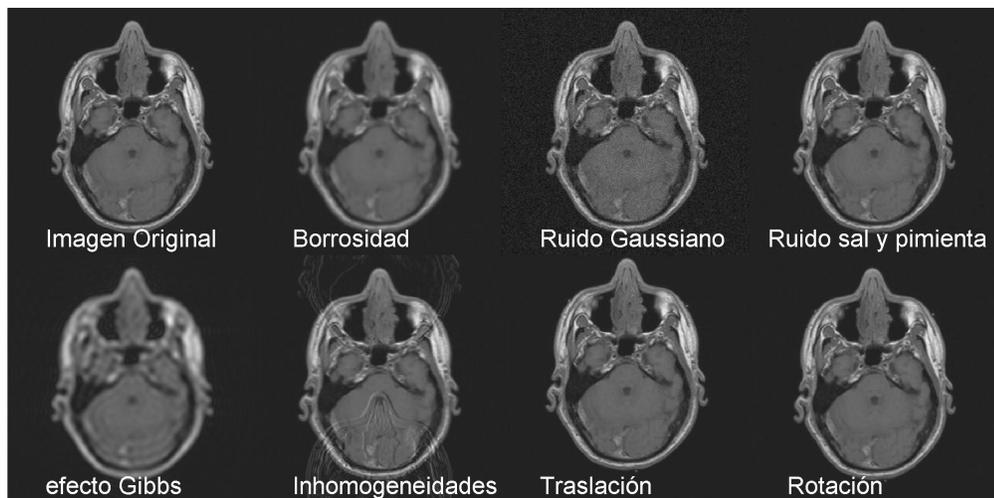


FIGURA 2.12. Imagen original y ejemplos de imágenes distorsionadas con cada tipo de degradación: borrosidad, ruido gaussiano, ruido sal y pimienta, efecto de Gibbs, inhomogeneidades de campo, traslación y rotación.

En adelante, cada vez que nos refiramos al RMSE lo estaremos haciendo al que se calcula según la ecuación (1.6), que hemos nombrado  $RMSE_{rel}$ .

### **2.4.1. Experimento 1: comparación con RMSE**

El objetivo de este experimento es analizar la correlación que podría existir entre las métricas propuestas y el RMSE, para ver si efectivamente tienen un comportamiento distinto de éste frente a las distorsiones de las imágenes. Además de las métricas propuestas, se sometió al experimento a la métrica SSIM, propuesta en (Wang & Bovik, 2002), para ver el comportamiento de otra métrica distinta al RMSE, que también busca modelar la percepción subjetiva del SVH y que ha sido bastante utilizada.

La comparación de las métricas con el RMSE se hizo obteniendo 150 imágenes degradadas a partir de *Bárbara*, de tal modo que el  $RMSE_{rel}$  variara entre 0,05 y 0,2. Este rango es suficientemente grande para abarcar imágenes que son prácticamente iguales a la original hasta imágenes que tienen muy poco en común con ella.

### **2.4.2. Experimento 2: aumento gradual de las distorsiones**

En este experimento se fue degradando la imagen *Bárbara* progresivamente con cada una de las distorsiones descritas anteriormente, con el fin de analizar el comportamiento de las distintas métricas. El objetivo que se persigue es analizar la sensibilidad de cada una de las métricas para cada distorsión, y si un aumento de distorsión implica una peor calificación de la imagen, es decir, estudiar si las métricas tienen un comportamiento monótono creciente ante el aumento de las distorsiones escogidas.

### **2.4.3. Experimento 3: diferencia entre ruido aleatorio y ruido estructurado**

Se realizó un experimento para ver cómo se comportan las distintas métricas ante la presencia de ruido aleatorio o estructurado. A la imagen *MRI-Cabeza*, de 8 bits por píxel, se le distorsionaron sólo 20 píxeles en distintas posiciones de la imagen, aumentando su intensidad en 20 intensidades de gris. Luego se aplicó la misma distorsión a otros píxeles, que se encuentran en el centro de la imagen, que corresponde al canal endimario del cerebro, y se analiza el comportamiento del RMSE y de las métricas propuestas.

El objetivo de este experimento es ver si las métricas son capaces de discriminar entre ruido aleatorio, que en una imagen médica puede tener poca relevancia, y ruido estructurado, que podría ser igual a un tumor, por ejemplo.

#### 2.4.4. Experimento 4: percepción subjetiva

Para comparar todas las métricas con la percepción subjetiva se realizó un experimento en que a 6 observadores se le presentaron 15 imágenes simultáneas en un monitor (HP LP3065 de 30" HD), a las que el usuario debía calificar con una nota entre 1 y 5, siendo nota 1 la peor y 5 la mejor. Cada uno debía realizar el experimento dos veces, una con imágenes distorsionadas a partir de *Bárbara* y otra con imágenes provenientes de *MRI-Cabeza*.

Las imágenes presentadas al observador tienen un valor  $RMSE_{rel}$  similar respecto a la original correspondiente. Para el caso de la figura *Bárbara*, el  $RMSE_{rel}$  de todas las imágenes estaba entre 0,148 y 0,152, es decir, una pequeña banda en torno a 0,15, y para el caso de la figura *MRI-Cabeza*, una banda en torno a 0,45 (entre 0,448 y 0,452).

El rango del  $RMSE_{rel}$  escogido para cada una de las imágenes se seleccionó de tal manera que las diferencias entre las imágenes fueran apreciables, pero que no dañaran demasiado la imagen. Cuando una distorsión degrada completamente la imagen, cualquier métrica debería ser capaz de indicar que se trata de una mala imagen. En cambio, cuando las distorsiones no son apreciables o lo son levemente, sí es relevante establecer criterios de comparación para determinar cuál imagen es mejor que otra, criterios que sean concordantes con la calificación subjetiva de los observadores. Luego, los rangos fueron escogidos de forma heurística para caer en zonas donde se cumpliera esta premisa.

El objetivo de esta experiencia es analizar el comportamiento de cada métrica respecto a la percepción subjetiva, por lo que importará más la tendencia que presentan los datos al grado de ajuste de una recta. Esta experiencia es la más importante de todas las anteriormente propuestas porque refleja precisamente el objetivo de este trabajo.

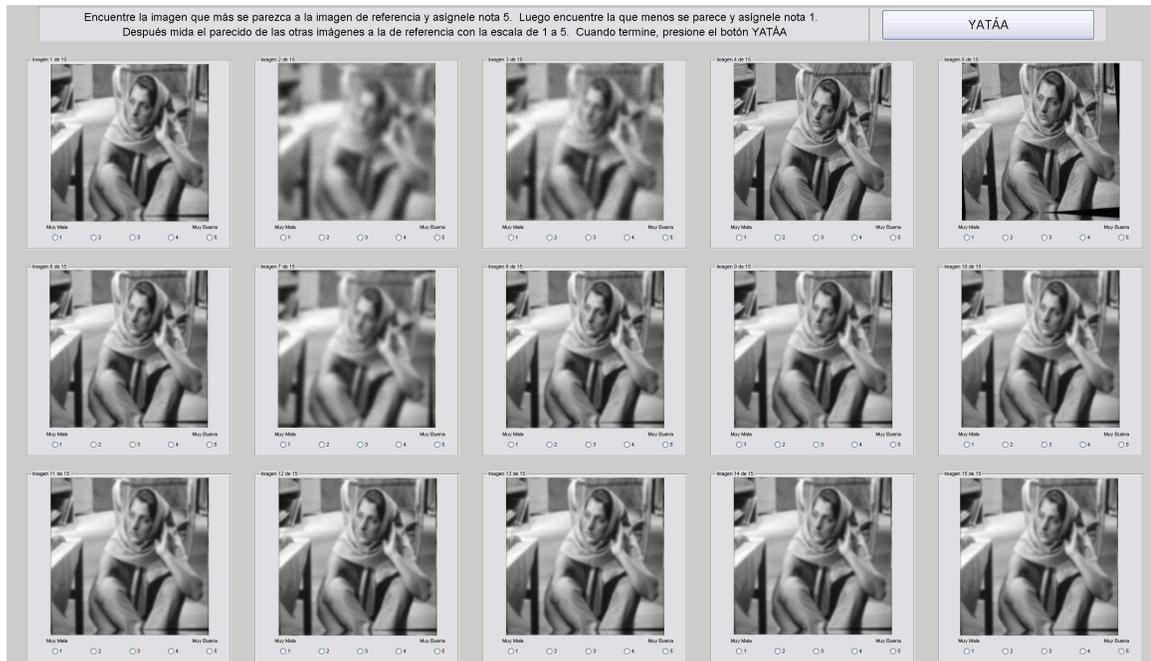


FIGURA 2.13. El experimento consistió en que se presentan 15 imágenes y el observador debe calificarlas entre 1 y 5.

### 3. RESULTADOS EXPERIMENTALES Y VALIDACIÓN

En esta sección se mostrarán los resultados de los experimentos descritos en la sección 2.4, para la validación de las métricas como estimadores de la percepción del SVH.

#### 3.1. Experimento 1: comparación con RMSE

Para las 150 imágenes distorsionadas a partir de *Bárbara* se busca ver si hay correlación entre el RMSE y cada una de las métricas propuestas, además del SSIM.

Al analizar la correlación que existe entre el RMSE y el SSIM (figura 3.1), se puede apreciar que el rango en que varía el SSIM es muy pequeño, siendo que el RMSE varía en un rango amplio. Además se ve una correlación bastante directa con el RMSE, aunque dicha correlación no es lineal.

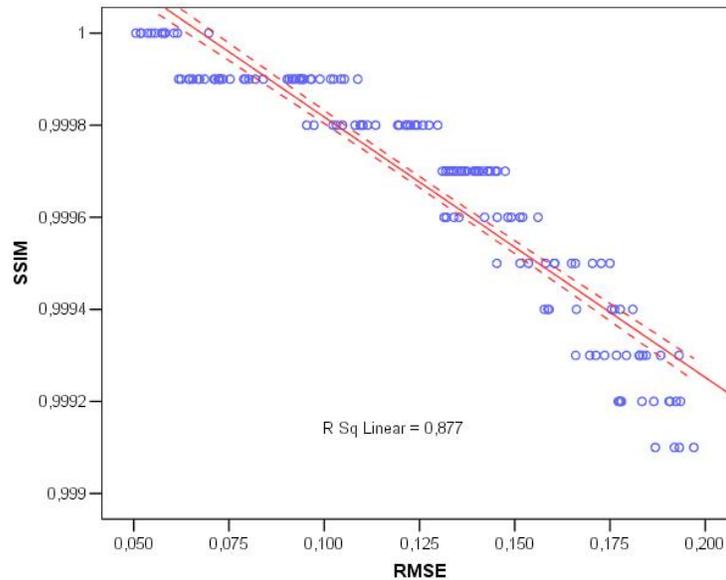


FIGURA 3.1. Correlación entre el RMSE y la métrica SSIM.

Para el caso del EPN (sección 2.1) (figura 3.2), se pueden apreciar zonas del gráfico en que existe una alta dispersión, como la región con  $RMSE_{rel}$  entre 0,05 y 0,125, que da cuenta que en esa zona al menos, la correlación entre ambas métricas es baja.

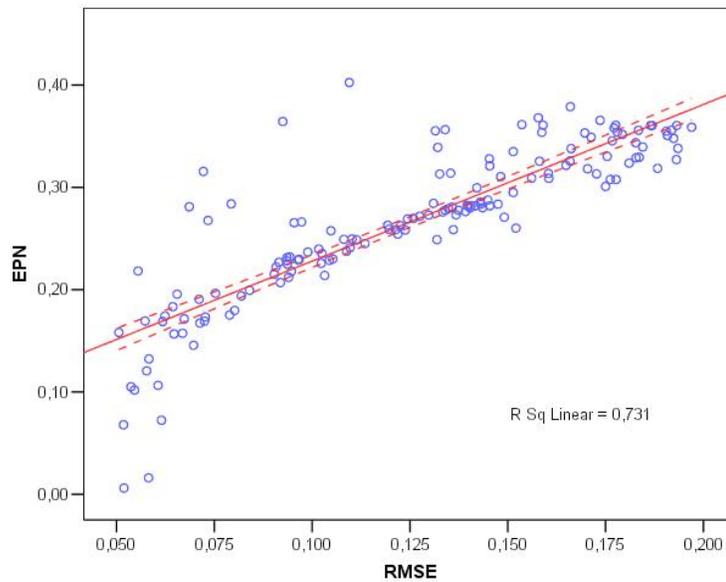


FIGURA 3.2. Correlación entre el RMSE y la métrica EPN.

El caso del IJND\_C (sección 2.2.1) (figura 3.3) presenta una alta dispersión respecto del  $RMSE_{rel}$ , con lo que no se puede afirmar que existe una relación unívoca entre ambas métricas. Lo mismo ocurre para el caso del IJND\_W (sección 2.2.2) (figura 3.4).

La métrica basada en transformaciones lineales buscando el espacio nulo de dimensión máxima (TL) (sección 2.3.1) tiene una relativa correlación con el RMSE (figura 3.5). Esto quiere decir que en términos generales se debieran comportar de manera similar. Aunque de todos modos existe cierta dispersión y la nube de puntos no es homogénea, lo que deja cierto lugar a diferencias entre las apreciaciones de ambas métricas.

Si analizamos el comportamiento de la métrica basada en transformaciones lineales que busca un cambio de dominio (TLD) (sección 2.3.2), vemos que es distinto a los otros casos (figura 3.6). La escasa correlación existente en la parte derecha del gráfico nos indica que ambas métricas tienen comportamientos distintos. Por ejemplo, las imágenes que tienen un  $RMSE_{rel}$  en torno a 0,175 podrían tener un valor TLD que podría variar entre 0,79 y 0,82, que para su escala es un rango bastante amplio.

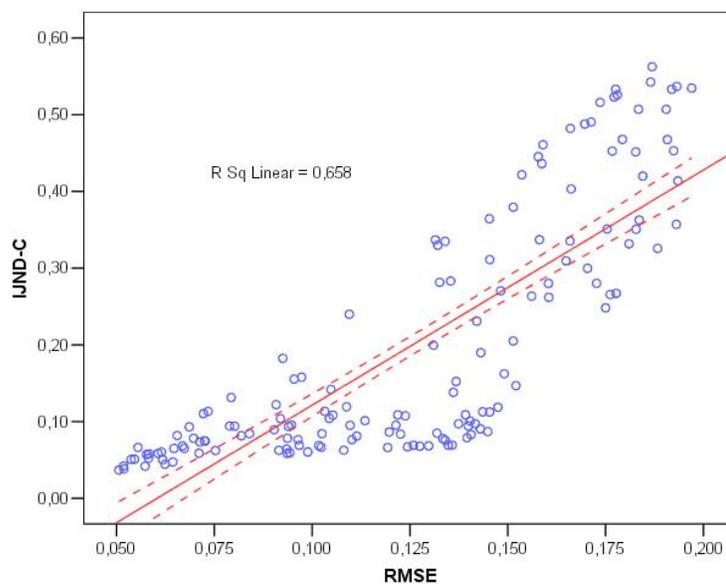


FIGURA 3.3. Correlación entre el RMSE y la métrica IJND\_C.

Métricas	$R^2$
SSIM	0,877
EPN	0,731
IJND_C	0,658
IJND_W	0,659
TL	0,829
TLD	0,466

TABLA 3.1. Valores  $R^2$  de los ajustes lineales entre el RMSE y las métricas. El valor más alto del ajuste lo tiene la métrica SSIM.

### 3.2. Experimento 2: aumento gradual de las distorsiones

A continuación se presentan los comportamientos de cada una de las métricas propuestas ante aumentos graduales de los distintos tipos de distorsiones señalados en la sección 2.4.

La métrica EPN, se comporta monótonicamente creciente para todas las distorsiones, excepto para la rotación en ángulos mayores a 0,5 grados (figura 3.7), en que presenta un comportamiento decreciente.

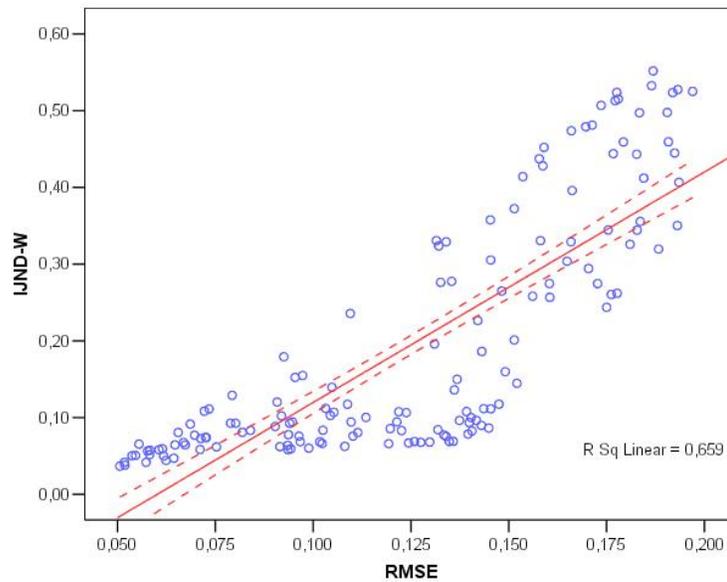


FIGURA 3.4. Correlación entre el RMSE y la métrica IJND-W.

Para recalcar que esta métrica es invariante a la traslación, la curva que representa el desempeño de esta métrica ante el aumento de esta distorsión (línea color rosado) no se normalizó, puesto que permanece prácticamente en cero.

La métrica IJND-C se comporta monótonicamente creciente para todas las distorsiones salvo para el ruido sal y pimienta (figura 3.8). Para la métrica IJND-W vemos que se comporta monótonicamente creciente para todas las distorsiones analizadas.

En el caso de la métrica basada en transformaciones lineales que maximizan el espacio nulo (TL), de la figura 3.10 vemos que ante todas las distorsiones tiene un comportamiento monótonicamente creciente, aunque en la traslación hay un punto que se sale del crecimiento monótonico. La métrica TLD (figura 3.11), en cambio, tiene un comportamiento oscilante ante el ruido, aunque en promedio se puede observar una tendencia ascendente. Por las curvas, se podría decir que la métrica es bastante invariante a las inhomogeneidades y a la rotación, mientras que es dependiente de todas las otras distorsiones.

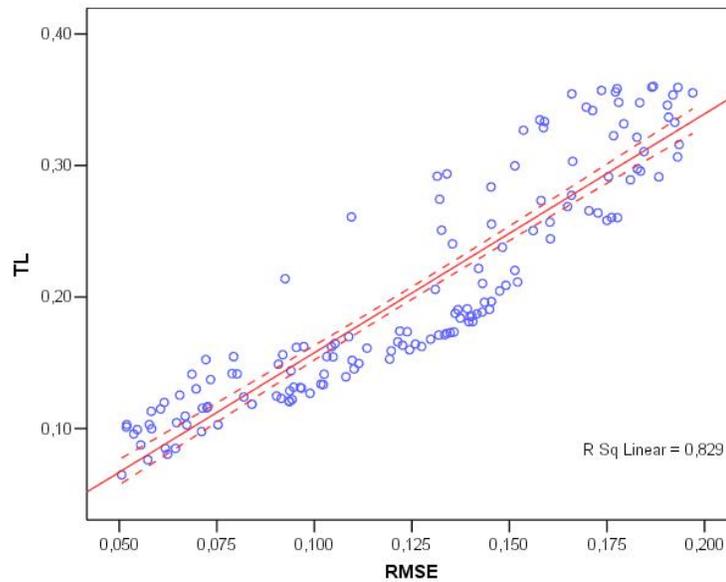


FIGURA 3.5. Correlación entre el RMSE y la métrica TL.

En los casos descritos por las figuras 3.7 y 3.10 se había visto, para el caso del efecto de Gibbs, que la disminución progresiva del ancho de banda del espacio de Fourier que se toma de la imagen implica una distorsión altamente no lineal en la imagen, y eso ocurre porque en el centro del espacio de Fourier hay más información que en la periferia. En este caso, sin embargo, vemos que el comportamiento es bastante más lineal que en los otros (línea color verde), lo que implica que esta métrica es fuertemente no lineal, y eso se debe a usar la transformación de Wavelets.

### 3.3. Experimento 3: diferencia entre ruido aleatorio y ruido estructurado

La imagen original y las distorsionadas se muestran en la figura 3.12, y los resultados se muestran en la tabla 3.2.

Como es previsible, el RMSE no cambia según dónde estén localizadas las distorsiones. El EPN indica que la imagen con ruido estructurado es mejor que la que tiene los píxeles distorsionados repartidos mientras que el TLD indica que son iguales. Las otras

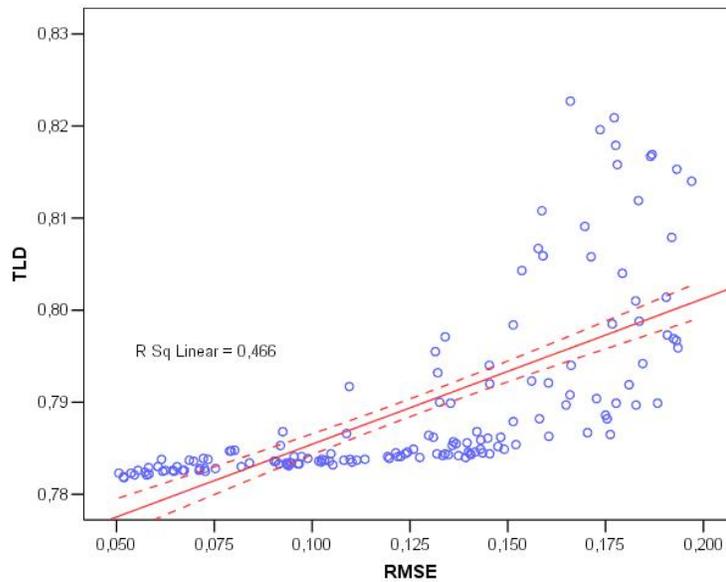


FIGURA 3.6. Correlación entre el RMSE y la métrica TLD.

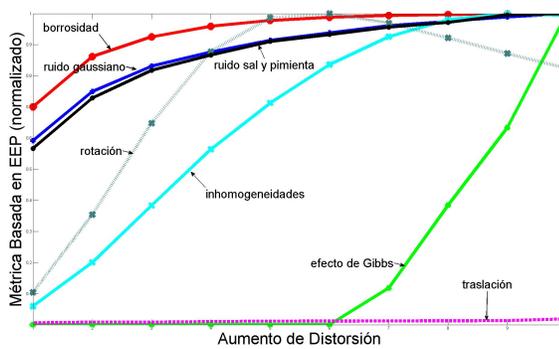


FIGURA 3.7. Desempeño de la métrica basada en el Espacio de Energías Ponderadas Normalizadas para cada tipo de distorsión.

métricas indican que la imagen con ruido estructurado es peor que la imagen con ruido aleatorio.

Con esto se puede afirmar que las métricas TL, IJND\_C e IJND\_W son capaces de discriminar entre imágenes con ruido aleatorio y ruido estructurado, indicando que el ruido estructurado es peor para una imagen que el aleatorio, ambos en iguales cantidades.

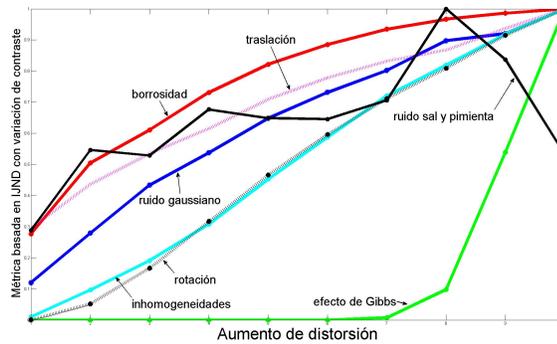


FIGURA 3.8. Desempeño de la métrica basada en IJND con variación de contraste, para cada tipo de distorsión.

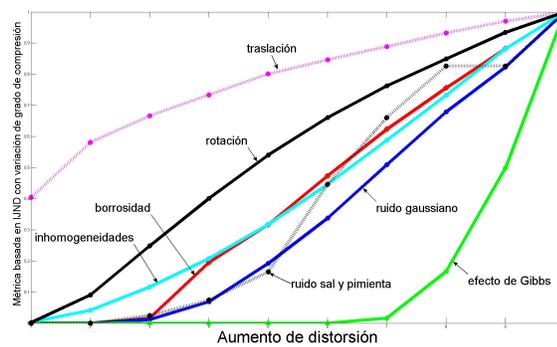


FIGURA 3.9. Desempeño de la métrica basada en IJND con variación de grado de compresión, para cada tipo de distorsión.

		Desempeño métricas					
		RMSE	EPN	TL	TLD	IJND_C	IJND_W
Imagen A		$6,6 \cdot 10^{-3}$	0,26	$0,1 \cdot 10^{-5}$	$15,6 \cdot 10^{-3}$	$0,29 \cdot 10^{-3}$	$0,29 \cdot 10^{-3}$
Imagen B		$6,6 \cdot 10^{-3}$	0,09	$0,12 \cdot 10^{-5}$	$15,6 \cdot 10^{-3}$	$0,31 \cdot 10^{-3}$	$0,31 \cdot 10^{-3}$
$\Delta\%$		0%	-65%	20%	0%	7%	7%

TABLA 3.2. Valores de las distintas métricas para el caso de la figura 3.12.

### 3.4. Experimento 4: percepción subjetiva

Como el experimento se hizo para dos conjuntos de imágenes, provenientes de dos imágenes originales (figura 2.11), en las figuras 3.13 a 3.17 se muestra a la izquierda el resultado del experimento con la imagen *Bárbara* y a la derecha el resultado del experimento con la imagen *MRI-Cabeza*.

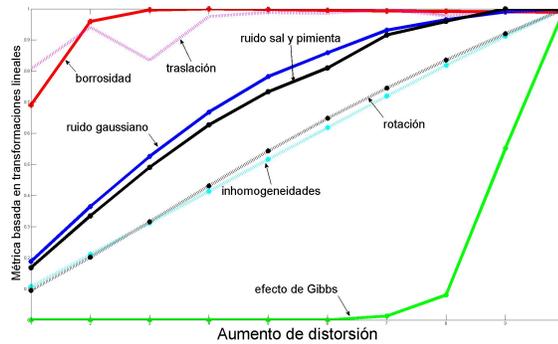


FIGURA 3.10. Desempeño de la métrica basada en transformaciones lineales, en que se busca la transformación que maximice el espacio nulo, para cada tipo de distorsión.

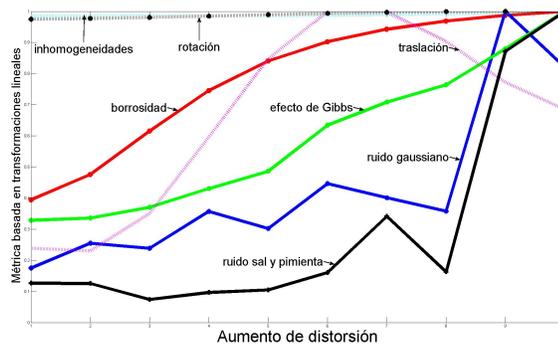


FIGURA 3.11. Desempeño de la métrica basada en transformaciones lineales, en que se busca la transformación que cambie el dominio a uno que tenga información perceptual, para cada tipo de distorsión.

En el caso del RMSE (figura 3.13) se puede apreciar que tanto para la imagen de *Bárbara* y de *MRI-Cabeza* no presenta una correlación entre el valor del índice y la nota subjetiva de los observadores. A modo de comparación, se presenta el caso del SSIM (figura 3.14), en que todas las imágenes tienen un valor similar, casi 1. Es decir, según esta métrica, todas las imágenes son prácticamente iguales a la original, lo que también es concordante con lo que dice el RMSE.

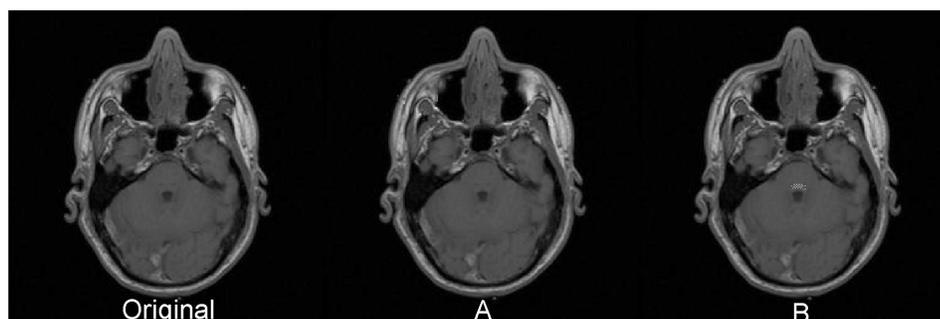


FIGURA 3.12. La imagen de la izquierda es la original. La imagen A está distorsionada con ruido sal y pimienta en posiciones aleatorias dentro de la imagen. La imagen B tiene el mismo grado de distorsión que A, pero todos los píxeles distorsionados están agrupados junto al canal endimario (al centro del cerebro).

Si vemos el caso del EPN (figura 3.15), se aprecia una clara tendencia decreciente, es decir, una tendencia de la métrica acorde a la calificación subjetiva de las personas que realizaron el experimento.

Métricas	Imágenes	
	Bárbara	Cabeza
RMSE	0,118	0,253
SSIM	0,571	0,871
EPN	0,682	0,761
IJND_C	0,707	0,395
IJND_W	0,708	0,282
TL	0,751	0,820
TLD	0,721	0,625

TABLA 3.3. Valores  $R^2$  de los ajustes lineales entre la calificación subjetiva de las imágenes y cada métrica.

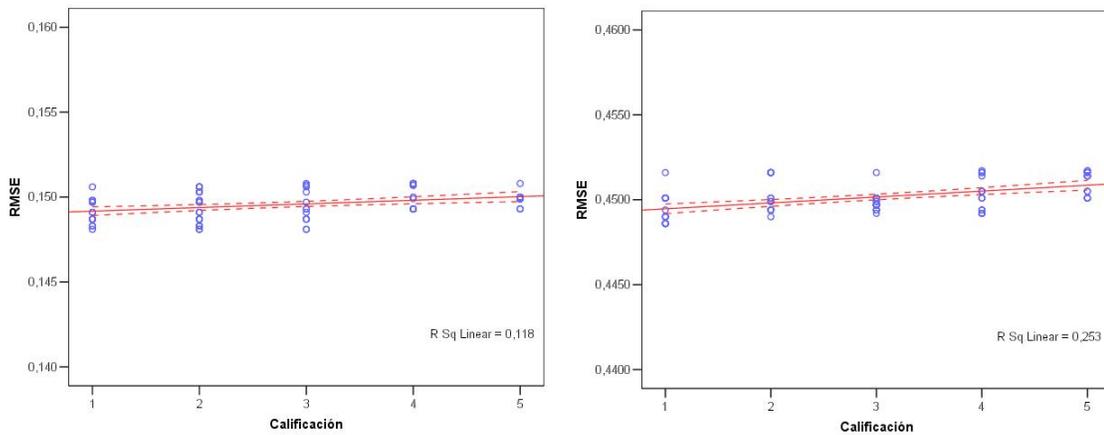


FIGURA 3.13. Relación de la calificación subjetiva con el RMSE. El eje horizontal corresponde a la nota con que fue calificada la imagen, y el eje vertical corresponde al valor de la métrica que se está evaluando. Se presenta la recta ajustada por mínimos cuadrados sólo para ver la tendencia que siguen los puntos (línea sólida). Además, en línea punteada se grafica el intervalo de confianza de la media con un 95% de confianza. Se muestra a la izquierda el resultado del experimento con la imagen *Bárbara* y a la derecha el resultado del experimento con la imagen *MRI-Cabeza*.

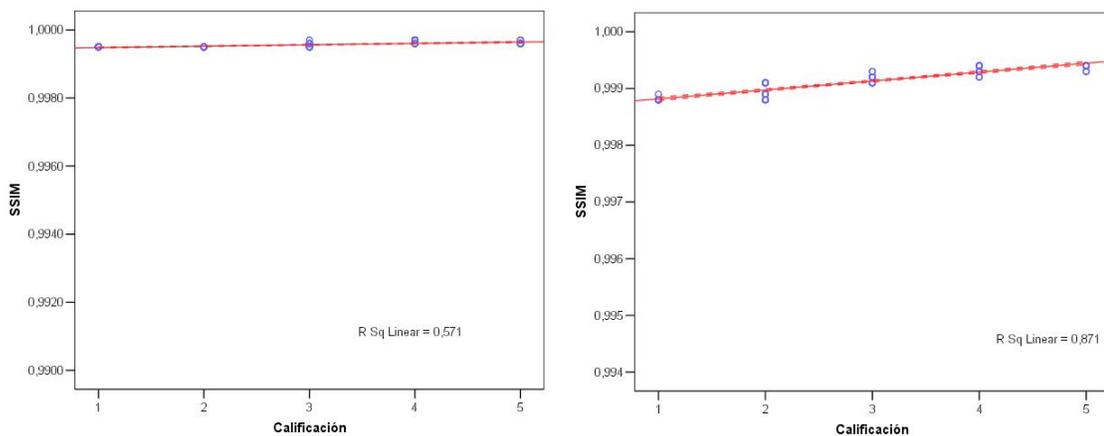


FIGURA 3.14. Relación de la calificación subjetiva con el SSIM. El eje horizontal corresponde a la nota con que fue calificada la imagen, y el eje vertical corresponde al valor de la métrica que se está evaluando. Se presenta la recta ajustada por mínimos cuadrados sólo para ver la tendencia que siguen los puntos (línea sólida). Además, en línea punteada se grafica el intervalo de confianza de la media con un 95% de confianza. Se muestra a la izquierda el resultado del experimento con la imagen *Bárbara* y a la derecha el resultado del experimento con la imagen *MRI-Cabeza*.

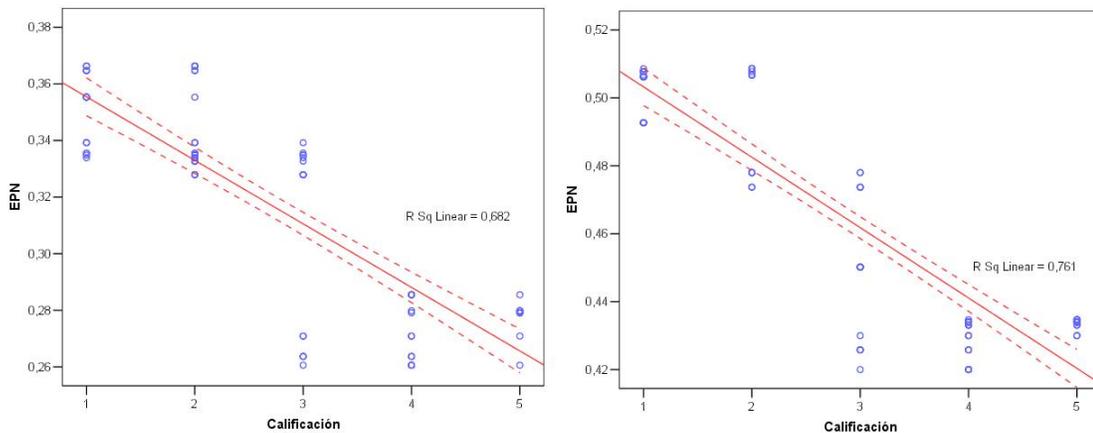


FIGURA 3.15. Relación de la calificación subjetiva con la métrica basada en EPN. El eje horizontal corresponde a la nota con que fue calificada la imagen, y el eje vertical corresponde al valor de la métrica que se está evaluando. Se presenta la recta ajustada por mínimos cuadrados sólo para ver la tendencia que siguen los puntos (línea sólida). Además, en línea punteada se grafica el intervalo de confianza de la media con un 95% de confianza. Se muestra a la izquierda el resultado del experimento con la imagen *Bárbara* y a la derecha el resultado del experimento con la imagen *MRI-Cabeza*.

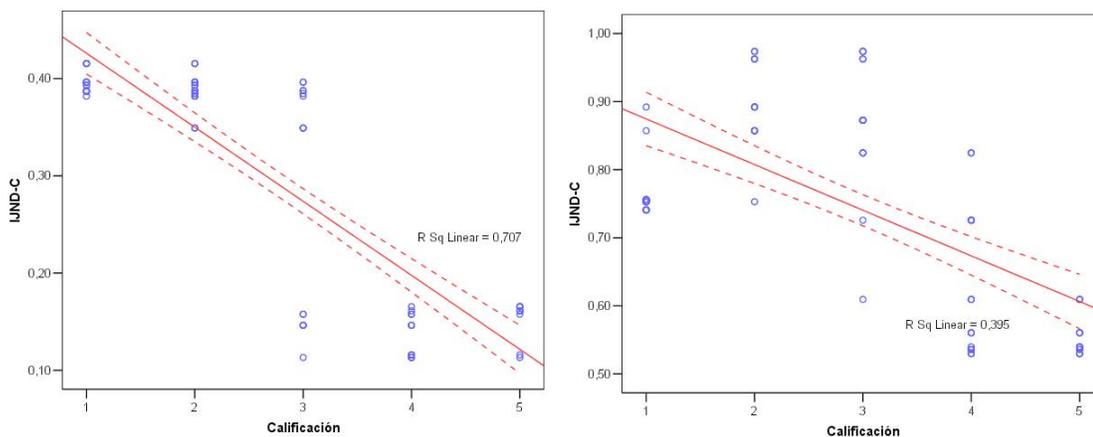


FIGURA 3.16. Relación de la calificación subjetiva con la métrica basada en JND con variación de contraste. El eje horizontal corresponde a la nota con que fue calificada la imagen, y el eje vertical corresponde al valor de la métrica que se está evaluando. Se presenta la recta ajustada por mínimos cuadrados sólo para ver la tendencia que siguen los puntos (línea sólida). Además, en línea punteada se grafica el intervalo de confianza de la media con un 95% de confianza. Se muestra a la izquierda el resultado del experimento con la imagen *Bárbara* y a la derecha el resultado del experimento con la imagen *MRI-Cabeza*.

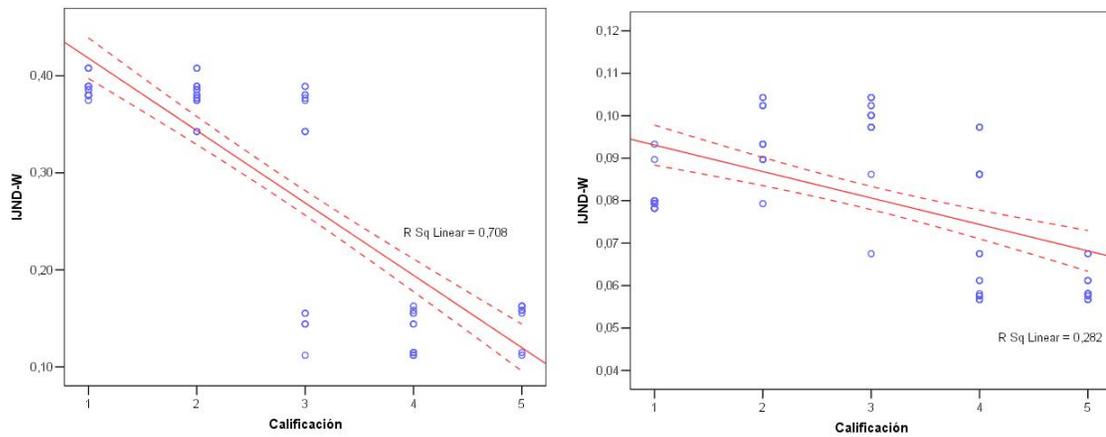


FIGURA 3.17. Relación de la calificación subjetiva con la métrica basada en JND con variación de compresión. El eje horizontal corresponde a la nota con que fue calificada la imagen, y el eje vertical corresponde al valor de la métrica que se está evaluando. Se presenta la recta ajustada por mínimos cuadrados sólo para ver la tendencia que siguen los puntos (línea sólida). Además, en línea punteada se grafica el intervalo de confianza de la media con un 95% de confianza. Se muestra a la izquierda el resultado del experimento con la imagen *Bárbara* y a la derecha el resultado del experimento con la imagen *MRI-Cabeza*.

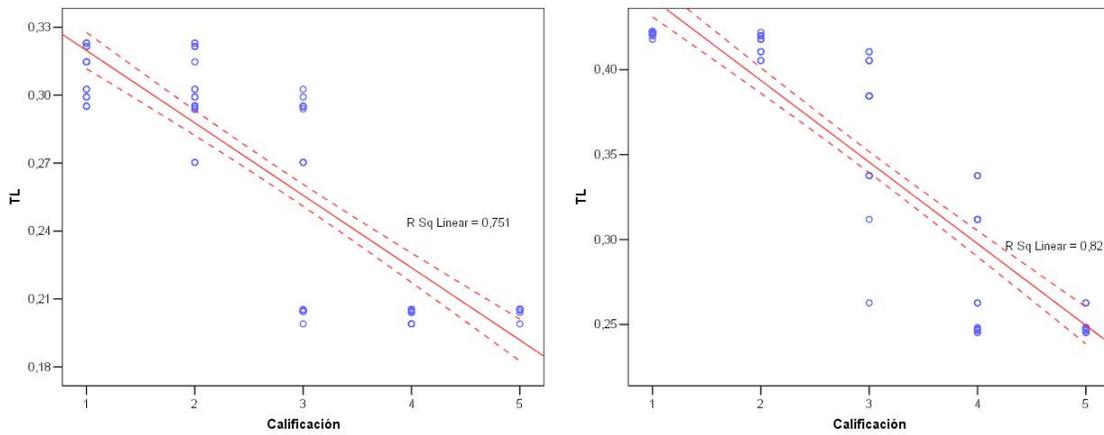


FIGURA 3.18. Relación de la calificación subjetiva con la métrica basada en transformaciones lineales en que se busca maximizar el espacio nulo. El eje horizontal corresponde a la nota con que fue calificada la imagen, y el eje vertical corresponde al valor de la métrica que se está evaluando. Se presenta la recta ajustada por mínimos cuadrados sólo para ver la tendencia que siguen los puntos (línea sólida). Además, en línea punteada se grafica el intervalo de confianza de la media con un 95% de confianza. Se muestra a la izquierda el resultado del experimento con la imagen *Bárbara* y a la derecha el resultado del experimento con la imagen *MRI-Cabeza*.

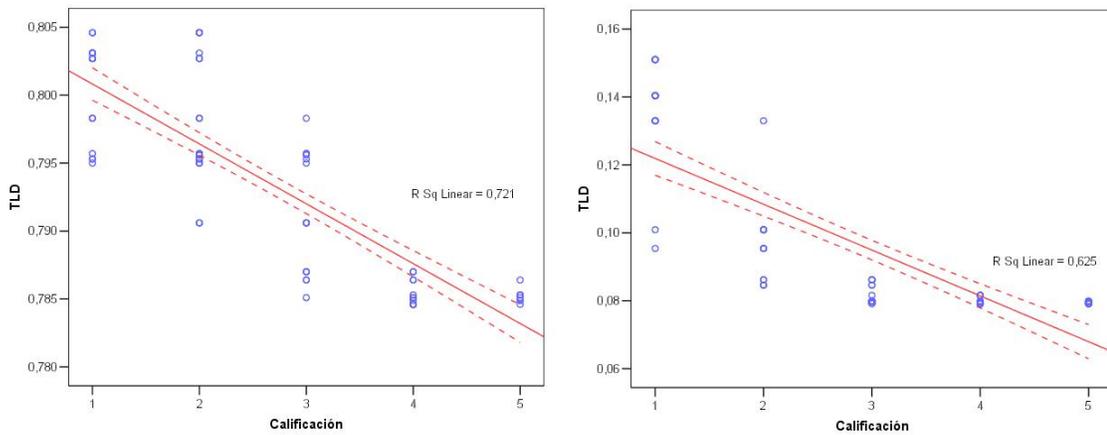


FIGURA 3.19. Relación de la calificación subjetiva con la métrica basada en transformaciones lineales en que se cambia de dominio. El eje horizontal corresponde a la nota con que fue calificada la imagen, y el eje vertical corresponde al valor de la métrica que se está evaluando. Se presenta la recta ajustada por mínimos cuadrados sólo para ver la tendencia que siguen los puntos (línea sólida). Además, en línea punteada se grafica el intervalo de confianza de la media con un 95% de confianza. Se muestra a la izquierda el resultado del experimento con la imagen *Bárbara* y a la derecha el resultado del experimento con la imagen *MRI-Cabeza*.

## **4. CONCLUSIONES Y TRABAJO FUTURO**

### **4.1. Revisión general de los resultados y discusión**

Los distintos experimentos muestran las deficiencias que tiene el RMSE como estimador de la percepción subjetiva del SVH, como quedó comprobado con la experiencia de calificación subjetiva de las imágenes: el SVH califica de modo distinto las imágenes al modo que lo hace el RMSE (figura 3.13).

El desempeño que mostró el SSIM en los distintos experimentos es muy similar al del RMSE, presentando incluso una correlación directa ante una amplia gama de imágenes distorsionadas (figura 3.1), con lo que es difícil que pueda comportarse mucho mejor que el RMSE respecto al SVH, hecho que quedó comprobado en el experimento perceptual (figura 3.14).

Las métricas propuestas en este trabajo presentan correlaciones con el RMSE más bajas que el SSIM (tabla 3.1), lo que quiere decir que no se comportarán necesariamente igual que dicha métrica para evaluar las imágenes. Además, todas ellas presentan tendencias acordes con la calificación subjetiva en los experimentos perceptuales (figuras 3.15 a 3.17), que es la más importante de todas las características.

### **4.2. Revisión y discusión de cada métrica en particular**

Cada una de las métricas propuestas tiene sus propias ventajas y desventajas. El EPN es invariante a la traslación (figura 3.7), pero califica mal la presencia de ruido estructurado (tabla 3.2). Su correlación con el RMSE no es lineal (figura 3.2), lo que es bueno, puesto que hemos demostrado que el RMSE no se correlaciona necesariamente bien con la percepción. La gran ventaja es que el EPN es concordante con la percepción subjetiva (figura 3.15). Su implementación es simple y no contiene conceptos complejos en sus fundamentos, lo que lo hace un potencial candidato para ser utilizado ampliamente.

Las métricas basadas en JND se comportan de modo muy similar en los distintos experimentos. Es interesante la cantidad de información que aporta el histograma (figura 2.9) de los mapas de distorsiones de las imágenes (figura 2.8), aunque el hecho que se sature rápidamente es una desventaja. Son variantes a la traslación y rotación de las imágenes, y la que usa el concepto de contraste ( $I_{JND\_C}$ ) presenta un comportamiento no monótonicamente creciente respecto de la distorsión sal y pimienta (figura 3.8). Respecto a la calificación subjetiva tienen una buena correlación con la percepción (figuras 3.16 y 3.17), aunque también presenta la forma de escalón. La razón puede ser la misma que la explicada para el caso del  $TL$ . Sin embargo, el hecho de usar un concepto concebido originalmente como binario (JND) y hacer de él uno capaz de encontrar graduaciones en la calidad de una imagen mediante un proceso sencillo, como la variación de contraste o compresión, creemos que es una nueva forma de enfrentar el problema de la calidad de imágenes que puede dar luces para nuevas soluciones cada vez mejores a futuro.

El  $TL$  tiene, en promedio, la misma tendencia del RMSE (figura 3.5), pero la relación con la percepción subjetiva es mejor (figura 3.18) aunque se aprecia una discontinuidad en la nota 3, puesto que las imágenes que están calificadas con una nota menor tienen un valor  $TL$  similar entre ellas, y las que están calificadas con una nota mayor también tienen un valor  $TL$  similar entre ellas. La explicación de este fenómeno puede ser que la elección de imágenes mostradas al observador fueran bastante cercanas a la original, y esto se hizo para presentar imágenes reales.

Además, la métrica  $TL$  es sensible al ruido estructurado versus el ruido aleatorio, calificando como peor el ruido estructurado, lo que es bueno para aplicaciones médicas, puesto que se podría confundir el ruido estructurado con alguna patología. Ante todas las distintas distorsiones se comporta monótonicamente creciente (figura 3.10), lo que implica que es variante a la traslación y rotación, cosa que no es deseable.

La métrica  $TLD$  tiene un comportamiento que no es estrictamente monótono ante el aumento de algunas distorsiones (figura 3.11), y como usa el dominio de Wavelets, que es

variante a las traslaciones, TLD también lo es. Al igual que el RMSE no es sensible ante la presencia de ruido estructurado versus aleatorio (tabla 3.2), aunque su correlación con la calificación perceptual es alta (figura 3.19). Su relación con el RMSE es la más baja de todas las métricas presentadas en este trabajo (figura 3.6).

Observando los resultados y la discusión de ellos, se puede afirmar que las métricas propuestas en este trabajo tienen mejor desempeño que el RMSE para imágenes distorsionadas moderadamente, como las que se obtienen de métodos de reconstrucción o compresión. Imágenes apreciablemente distintas que eran calificadas por el RMSE con valores prácticamente iguales, fueron calificadas por las métricas propuestas de un modo más cercano a la calificación subjetiva. Además, los enfoques presentados para evaluar la calidad de una imagen, basados en el álgebra lineal y mediante el aumento progresivo de alguna distorsión, son nuevas ideas que mostraron ser efectivas para las tareas propuestas, teniendo alto potencial de desarrollo futuro.

### **4.3. Trabajo futuro**

En este trabajo se han presentado las bases de nuevos métodos para evaluar la calidad de las imágenes. Si se quiere enfocar en el área de imágenes médicas, sería conveniente adecuar más el concepto de calidad según la opinión de los profesionales de la salud que trabajan con imágenes, los tecnólogos médicos y los radiólogos, puesto que son ellos los que saben cuándo una imagen médica es buena o mala. Si se pudieran modelar matemáticamente esos criterios, se podría elaborar una métrica que se centrara en esos puntos más que en otros.

Respecto a las métricas mismas, en primer lugar se debe establecer un criterio para validar los nuevos métodos de evaluación de calidad de imagen, por ejemplo en que se evalúe la capacidad de la métrica de predecir el desempeño en la ejecución de una tarea por parte del usuario. Posiblemente se puede utilizar el concepto de calidad de imagen que se propone en (Barrett et al., 1993), en que divide la calidad de una imagen médica en dos

ejes: capacidad de estimación de parámetros a partir de la imagen y capacidad de hacer un diagnóstico correcto a partir de la imagen. Con esto, una validación más profunda de los métodos propuestos se hace necesaria.

Las métricas basadas en transformaciones lineales proveen una nueva forma de ver el análisis de la calidad de una imagen. En esta línea se puede seguir profundizando en los supuestos que se hacen, o en las características que tienen que tener las transformaciones mismas para que la información que provean sea consistente con el SVH.

El concepto de variación de degradación como una forma de llevar un sistema que es binario a uno que no lo es presenta un amplio campo de utilización en esta área como en otras. Se pueden buscar otras formas de degradar una imagen, además de la disminución de contraste o el aumento de la compresión, que sean más consistentes con el SVH, y así determinar una distancia perceptual para las imágenes.

El dominio de Wavelets se presenta como una buena alternativa para evaluar la calidad de una imagen, pero todavía no soluciona el problema ante las rotaciones y traslaciones de imágenes. Dado que sólo una de las métricas presentadas es invariante a la traslación, se hace necesario continuar la investigación para incorporarlas al modelo perceptual.

## BIBLIOGRAFÍA

- Barrett, H., Yao, J., Rolland, J., & Myers, K. (1993, November 1, 1993). Model observers for assessment of image quality. *Proceedings of the National Academy of Sciences*, 90(21), 9758-9765.
- Blakemore, C., & Campbell, F. W. (1969, July 1). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *The Journal of Physiology Online*, 203(1), 237-260.
- Bradley, A. P. (1999). *A wavelet visible difference predictor* (No. 5). (ID: 1)
- Cadik, M., & Slavik, P. (2004). *Evaluation of two principal approaches to objective image quality assessment*. (ID: 1)
- Chandler, D. M., Masry, M. A., & Hemami, S. S. (2003). *Quantifying the visual quality of wavelet-compressed images based on local contrast, visual masking, and global precedence*. (ID: 1)
- Chou, C.-H., & Li, Y.-C. (1995). *A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile* (No. 6). (ID: 1)
- Daly, S. J. (1992, August 27, 1992). Visible differences predictor: an algorithm for the assessment of image fidelity. In B. E. Rogowitz (Ed.), (Vol. 1666, p. 2-15). SPIE.
- Engelke, U., & Zepernick, H. J. (2007). *Perceptual-based quality metrics for image and video services: A survey*. (ID: 1)
- Eskicioglu, A. M., & Fisher, P. S. (1995). *Image quality measures and their performance* (No. 12). (ID: 1)
- Freitas-Zampolo, R. de, & Seara, R. (2005). *A comparison of image quality metric performances under practical conditions*. (ID: 1)
- Fronthaler, H., Kollreider, K., & Bigun, J. (2006). Automatic image quality assessment with application in biometrics. *cvprw*, 0, 30.
- Gao, X., Wang, T., & Li, J. (2005). *A content-based image quality metric*.

- Gifford, H. C., King, M. A., Vries, D. J. de, & Soares, E. J. (2000, March 1). Channelized hotelling and human observer correlation for lesion detection in hepatic spect imaging. *The Journal of Nuclear Medicine*, 41(3), 514-521.
- Girod, B. (1993). What's wrong with mean-squared error? In (p. 207-220). MIT Press.
- Gonzalez, R. C., & Woods, R. E. (2001). Digital image fundamentals. In (2 ed., p. 34). New Jersey 07458: Prentice-Hall.
- Hadhoud, M. M. (2001). *X-ray images enhancement using human visual system model properties and adaptive filters*. (ID: 1)
- Heeger, D. J., & Teo, P. C. (1995). *A model of perceptual image fidelity*. (ID: 1)
- Ivkovic, G., & Sankar, R. (2004). *An algorithm for image quality assessment*. (ID: 1)
- Jayant, N. (1992). *Signal compression: technology targets and research directions* (No. 5). (ID: 1)
- Kai, X., Jie, Y., Min, Z. Y., & Liang, L. X. (2005, 7). Hvs-based medical image compression. *European Journal of Radiology*, 55(1), 139-145.
- Kovesi, P. (2000). Phase congruency: A low-level image invariant. *Psychological Research*, 64(2), 136-148.
- Li, B., Meyer, G. W., & Klassen, R. V. (1998, 07/1998). Comparison of two image quality models. *Proceedings SPIE, Human Vision and Electronic Imaging III*, 3299, 98-109.
- Liu, Z., & Laganière, R. (2007, 1/1). Phase congruence measurement for image similarity assessment. *Pattern Recognition Letters*, 28(1), 166-172.
- Lubin, J. (1997). *A human vision system model for objective picture quality measurements*. (ID: 1)
- Manning, D. J., Gale, A., & Krupinski, E. A. (2005, August 1). Perception research in medical imaging. *The British Journal of Radiology*, 78(932), 683-685.

- Mannos, J., & Sakrison, D. (1974). *The effects of a visual fidelity criterion of the encoding of images* (No. 4). (ID: 1)
- Pang, J., Zhang, R., Lu, L., & Liu, Z. (2007, lunes, 03 de diciembre de 2007). Image quality assessment based on energy of structural distortion. *Lecture Notes in Computer Science*, 4810/2007(Advances in Multimedia Information Processing, PCM 2007), 785-785-793.
- Peli, E. (1990, October 1990). Contrast in complex images. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, 7(Issue 10), 2032-2032-2040.
- Peli, E. (2001, 02/01). Contrast sensitivity function and image discrimination. *Journal of the Optical Society of America A*, 18(2), 283-293.
- Sheikh, H. R., & Bovik, A. C. (2006). *Image information and visual quality* (No. 2). (ID: 1)
- Sheikh, H. R., Bovik, A. C., & Veciana, G. de. (2005). *An information fidelity criterion for image quality assessment using natural scene statistics* (No. 12). (ID: 1)
- Shnayderman, A., Gusev, A., & Eskicioglu, A. M. (2006). *An svd-based grayscale image quality measure for local and global assessment* (No. 2). (ID: 1)
- Silva, E. A., Panetta, K., & Aгаian, S. S. (2007, April 27, 2007). Quantifying image similarity using measure of enhancement by entropy. In S. S. Aгаian & S. A. Jassim (Eds.), (Vol. 6579, p. 65790U). SPIE.
- Stehling, M., Turner, R., & Mansfield, P. (1991, October 4). Echo-planar imaging: magnetic resonance imaging in a fraction of a second. *Science*, 254(5028), 43-50.
- Suthaharan, S., Kim, S.-W., & Rao, K. R. (2005). *A new quality metric based on just-noticeable difference, perceptual regions, edge extraction and human vision* (No. 2). (ID: 1)
- Teo, P. C., & Heeger, D. J. (1994). *Perceptual image distortion*. (ID: 1)

- Vujovic, I., Kuzmanic, I., & Krcum, M. (2002). *Experimental results on visibility threshold in human visual perception for application in image/video coding and quality assessment*. (ID: 1)
- Wang, Z., & Bovik, A. C. (2002). *A universal image quality index* (No. 3). (ID: 1)
- Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). *Image quality assessment: from error visibility to structural similarity* (No. 4). (ID: 1)
- Wang, Z., & Simoncelli, E. P. (2005). *Translation insensitive image similarity in complex wavelet domain*. (ID: 1)
- Wei, Z. H., Qin, P., & Fu, Y. Q. (1998). *Perceptual digital watermark of images using wavelet transform* (No. 4). (ID: 1)
- Westen, S. J. P., Lagendijk, R. L., & Biemond, J. (1995). *Perceptual image quality based on a multiple channel hvs model*. (ID: 1)
- Yang, X. K., Lin, W. S., Lu, Z., Ong, E. P., & Yao, S. (2003). *Just-noticeable-distortion profile with nonlinear additivity model for perceptual masking in color images*. (ID: 1)
- Yang, X. K., Ling, W. S., Lu, Z. K., Ong, E. P., & Yao, S. S. (2005, 8). Just noticeable distortion model and its applications in video coding. *Signal Processing: Image Communication*, 20(7), 662-680.
- Yao, S., Lin, W., Lu, Z. K., Ong, E. P., Locke, M. H., & Wu, S. Q. (2007). *Image quality measure using curvature similarity*. (ID: 1)
- Yao, S., Ong, E., & Loke, M. H. (2008). *Perceptual distortion metric based on wavelet frequency sensitivity and multiple visual fixations*. (ID: 1)
- Zhang, X., Lin, W., & Xue, P. (2008, 1). Just-noticeable difference estimation with pixels in images. *Journal of Visual Communication and Image Representation*, 19(1), 30-41.
- Zhang, X., Lin, W. S., & Xue, P. (2005, 4). Improved estimation for just-noticeable visual distortion. *Signal Processing*, 85(4), 795-808.