



## **Doctorado en Ciencias Radiológicas**

### **MEMORIA DE INVESTIGACIÓN**

Presentada al Departamento de Radiología y Medicina Física  
de la Universidad Complutense de Madrid,  
en cumplimiento de los requisitos para la obtención del

### **Diploma de Estudios Avanzados**

Por

#### **Gabriel Prieto Renieblas**

Departamento de Radiología y Medicina Física  
Facultad de Medicina  
Universidad Complutense de Madrid  
España

Dirigido por:

#### **Dra. Margarita Chevalier del Río** **Dr. Eduardo Guibelalde del Castillo**

Departamento de Radiología y Medicina Física  
Facultad de Medicina  
Universidad Complutense de Madrid  
España

**Junio 2009**

# ÍNDICE

	<b>Pág.</b>
1. INTRODUCCIÓN	3
2. ANTECEDENTES: MODELOS DE OBSERVADOR	6
2.1. El modelo del Observador Ideal	6
2.2. El Observador Fisher-Hotelling	8
2.3. El Observador Filtro Adaptado de No Blanqueo (NPWMF)	12
2.4. El modelo JNDMetrix	13
3. FUNDAMENTOS TEÓRICOS	15
3.1. El índice SSIM	15
3.2. El índice de correlación cruzada R*	16
4. MATERIAL Y MÉTODOS	18
4.1. El maniquí CDMAM	18
4.2. Evaluación de las imágenes	18
4.2.1. Localización de la rejilla	19
4.2.2. Localización de los discos y cálculo del índice R*	21
5. RESULTADOS	25
6. DISCUSIÓN	28
7. CONCLUSIONES	30
Anexo A. El programa CDCOM	31
REFERENCIAS	33

# 1. INTRODUCCIÓN

El análisis de la calidad de imagen tiene un papel central en el diseño de los sistemas de obtención de imágenes de diagnóstico médico. Se ha realizado un gran esfuerzo para desarrollar métricas de evaluación de la calidad objetivas que estén correlacionadas con la “calidad” percibida por el radiólogo, es decir, métricas de calidad de imagen que se correlacionen con la capacidad clínica del sistema. Hasta ahora sólo se ha alcanzado un éxito parcial. Ciertas métricas muy extendidas, como la Relación Señal-Ruido de Pico (PSNR) o el Error Cuadrático Medio (MSE) son muy simples de calcular, pero no presentan una buena correlación con la calidad de imagen percibida por un observador humano<sup>1,2</sup> y son incluso mucho peores si se intenta deducir de ellas la utilidad diagnóstica de una imagen<sup>3</sup>. Existen métricas más cercanas a la calidad real de un sistema, como la Función de Transferencia de Modulación (MTF), Espectro de Potencia del Ruido (NPS), Fotonos Equivalentes a Ruido (NEQ) o Eficiencia Cuántica de Detección (DQE)<sup>4</sup>. Estas métricas describen mucho mejor el proceso de formación de imagen del sistema y pueden ser usadas para dos objetivos: mejorar la calidad de imagen del sistema y realizar predicciones de su comportamiento utilizando modelos ideales de observador, principalmente con el modelo de Observador Ideal<sup>5</sup>, basado en la Teoría Estadística de la Decisión (SDT).

Sin embargo, el cálculo de estas magnitudes no es sencillo y el modelo de Observador Ideal sólo es aplicable en tareas muy simples y específicas del tipo Señal Conocida Exactamente / Fondo Conocido Exactamente (SKE/BKE)<sup>6</sup>. Además, la sensibilidad del modelo del Observador Ideal es muy superior a la del observador humano, por lo que no hay una relación directa entre parámetros como MTF, NPS, NEQ y DQE y la calidad de imagen percibida.

Existen otros modelos que presentan una mejor correlación con el observador humano y se pueden aplicar a tareas de mayor complejidad que las tipo SKE/BKE. Entre ellos cabe destacar principalmente los modelos de Fisher-Hotelling<sup>7</sup> y el Filtro Adaptado de No Blanqueo (*Non PreWhitening Matched Filter*, NPWMF)<sup>8</sup>. El cálculo asociado a estos modelos es complejo y se basa en presupuestos estadísticos muy estrictos acerca de las características de la señal y el ruido de las imágenes que se analizan. En el caso de que se desee utilizar el modelo en el dominio de la frecuencia espacial, se añaden

premisas tales como que el sistema de imagen sea invariable a pequeñas traslaciones, que el sistema sea lineal y que la distribución de ruido sea ergódica. Estas restricciones hacen que estos modelos sean aplicables sobre imágenes de objetos muy simples o bien hacen necesario el análisis de una enorme cantidad de imágenes para crear indicadores estadísticos adecuados<sup>9</sup>.

Otros modelos de percepción no basados en la SDT, como el índice JNDMetrix de Sarnoff Corporation<sup>10</sup> pueden ser aplicados a todo tipo de imágenes, pero son modelos enormemente complejos y, en el caso del modelo JNDMetrix, su adaptación al campo de la imagen médica ha sido cancelado por parte de la citada corporación.

El objetivo de este trabajo ha sido estudiar el comportamiento del Índice de Similitud Estructural (SSIM)<sup>11</sup>, en su versión Índice de Correlación Cruzada de Similitud Estructural Multiescala ( $R^*$ )<sup>12,13</sup>, en la evaluación de imágenes médicas. Este modelo se basa en la hipótesis de que el sistema visual humano está muy adaptado para extraer información estructural de las imágenes, de tal forma que una medida de la información estructural puede dar una buena aproximación de la calidad de imagen percibida. Así mismo, presenta otras características que pueden hacerlo adecuado para el objetivo de este estudio:

- 1) SSIM está diseñado y probado para analizar escenas naturales, cuya complejidad es del orden, o incluso mayor que la de las imágenes médicas. No existe ninguna premisa previa que obligue a simplificar la complejidad de la imagen.
- 2) SSIM se ha utilizado para el aseguramiento de la calidad y de la fidelidad de la imagen en un gran número de aplicaciones comerciales y de investigación: reconocimiento de texto, de huellas dactilares, de rostro, aplicaciones de control de calidad en compresión de vídeo e imagen fija, eliminación de ruido, análisis de calidad de color en imagen, vigilancia por vídeo, imagen médica, etc.
- 3) SSIM supera a la mayoría de las métricas utilizadas en la actualidad en análisis de vídeo e imagen fija<sup>11</sup>.
- 4) En el caso de  $R^*$  su sensibilidad es análoga, en los experimentos realizados hasta ahora, a la de un observador humano en el análisis de imágenes cercanas al umbral de percepción<sup>13</sup>.
- 5) Su complejidad computacional es extremadamente baja.

En el presente estudio, como primer paso, se ha aplicado esta métrica a un problema relativamente sencillo: la localización de estructuras simples sobre fondos relativamente uniformes. Para ello, hemos utilizado como referencia imágenes de un maniquí del tipo contraste-detalle, en concreto imágenes del maniquí CDMAM, ampliamente utilizado en el campo de la mamografía. Este maniquí proporciona imágenes con una estructura relativamente sencilla, ya que está constituido por simples discos de oro de diferente diámetro y espesor que se han depositado sobre una lámina uniforme de metacrilato. Por otro lado existen desarrollos informáticos de evaluación automática de las imágenes del citado maniquí, siendo el más extendido y probado de ellos el programa CDCOM<sup>14</sup>, que utiliza como base de detección el modelo NPWMF. Si comparamos los resultados de nuestro algoritmo con los resultados obtenidos por observadores humanos, por el programa CDCOM y por algunas de las aproximaciones que se realizan basándose en sus cálculos, podemos tener una buena referencia de la efectividad del método de evaluación propuesto.

El producto de este trabajo ha sido el desarrollo de un sistema automático de evaluación del maniquí CDMAM que produce, como salida, la detección esperable por un observador humano de los discos de este maniquí. El programa basa el cálculo de la probabilidad de detección de los discos en el algoritmo R\*. Se ha construido como un algoritmo informático desarrollado con el lenguaje de programación Java y como programa integrado (*plugin*) dentro del sistema de visualización y proceso de imágenes ImageJ<sup>15</sup>. Este plugin estará próximamente disponible en forma de código abierto de libre distribución (OSS) como herramienta de libre uso para toda la comunidad científica, tal y como ya hemos publicado anteriormente resultados intermedios del mismo.

## 2. ANTECEDENTES: MODELOS DE OBSERVADOR

### 2.1. El modelo del Observador Ideal

El Observador Ideal se deriva de los estudios de Bayes en Teoría Estadística de la Decisión (SDT). Este modelo calcula la probabilidad de ocurrencia de dos hipótesis alternativas y decide a favor de la más probable de las dos usando toda la información disponible. Aplicado a Teoría de la Imagen, utiliza toda la información disponible en las dos imágenes alternativas que constituyen las hipótesis que se deben valorar. Aunque este modelo puede aplicarse a distintas tareas de decisión en el campo de la imagen, se aplica con más facilidad a las del tipo SKE/BKE<sup>16</sup>, que se basan en la detección de una señal completamente especificada y conocida contra un fondo también completamente especificado. En términos clínicos, esta tarea se concreta en la detección de lesiones de características bien conocidas (posible localización, tamaño, forma, etc.), sobre fondos de características también conocidas por el radiólogo. La tarea que debe efectuar el radiólogo es decidir si la imagen incluye o no la lesión.

La figura de mérito más ampliamente utilizada para el Observador Ideal es el índice de detección, directamente relacionado con la probabilidad de detección de la señal. La expresión matemática del citado índice es<sup>17</sup>:

$$SNR_I^2 = K^2 \int \frac{|\Delta \tilde{f}(v)|^2 \cdot MTF^2(v)}{W_n(v)} \cdot dv \quad (\text{Ec. 1})$$

- $K$  = constante característica de transferencia del sistema a gran escala.
- $|\Delta \tilde{f}(v)|$  = diferencia de las Transformadas de Fourier (FT) de las dos señales que se están analizando, es decir, FT (señal + fondo) – FT (fondo).
- $MTF(v)$  : función de transferencia de modulación.
- $W_n(v)$  : es el NPS (espectro de potencia del ruido) del sistema.

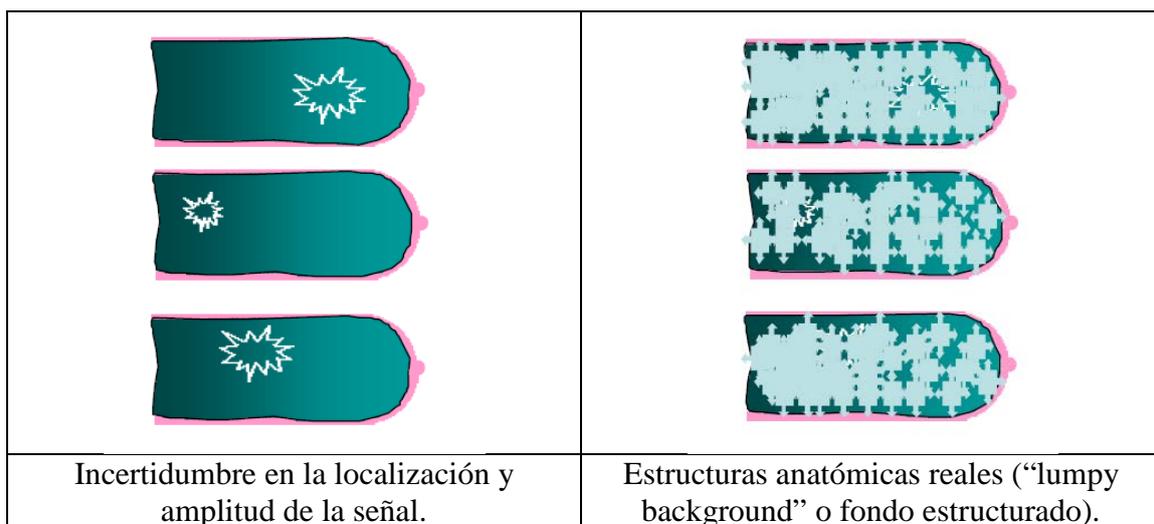
Como puede verse, la probabilidad depende de la frecuencia espacial ( $v$ ) y debe calcularse en este dominio. Hay que destacar que se exige el completo conocimiento de estos parámetros para poder calcular la estrategia del Observador Ideal. La presencia del

factor  $W_n(v)$  en el denominador asegura que si existe ruido correlacionado en el sistema, este modelo de observador lo elimina, ya que está también presente en el factor  $\Delta f(v)$ . Por ello a veces este modelo es conocido como el modelo Filtro Adaptado de No Blanqueo Ideal (*Ideal PreWhitening Matched Filter*), ya que “blanquea” la señal del ruido correlacionado volviéndolo blanco.

El cálculo del Observador Ideal se basa en varias asunciones:

- La señal y el fondo se conocen de forma exacta (SKE/BKE).
- El sistema de imagen ha de ser lineal o, al menos, linealizable.
- El sistema de imagen ha de ser invariante al desplazamiento.
- Las señales han de ser de bajo contraste.
- El ruido ha de ser gaussiano y aditivo.
- El ruido ha de ser ergódico.

El cálculo del Observador Ideal presenta ciertos inconvenientes. Si bien muchos sistemas de bajo contraste son lineales o fácilmente linealizables, la mayor parte de ellos no tienen una distribución de ruido ergódica. Este problema es obvio trabajando en el dominio espacial en lugar del dominio de frecuencias espaciales, operando directamente con las matrices de transferencia del sistema y con las matrices de datos de la imagen. Sin embargo, la principal dificultad es el paradigma SKE/BKE. Obsérvese la Fig. 1.



**Fig. 1.** Paradigma SKE/BKE y la realidad radiológica.

En este caso, el paradigma SKE/BKE se viola claramente, ya que la señal y el fondo se conocen sólo de forma aproximada. Es necesario utilizar aproximaciones estadísticas que incluyen un número elevadísimo de muestras. El Observador Ideal se vuelve virtualmente imposible de calcular. Además, el uso de aproximaciones estadísticas obliga a que el modelo del Observador Ideal introduzca operaciones no lineales sobre los datos. En este sentido, existen fuertes dudas de que el observador humano pueda ejecutar dicho tipo de operaciones sobre las imágenes.

Por último, y aun en los casos en que el modelo pueda aplicarse (caso SKE/BKE), su capacidad de detección es mucho mejor que la de un observador humano y no hay una relación sencilla entre estas capacidades de detección que permita deducir una de otra. El modelo del Observador Ideal permanece como referencia de la mejor capacidad de detección a la que puede llegar un observador, pero sin una clara correlación con la calidad de imagen percibida.

## 2.2. El Observador Fisher-Hotelling

El Observador Fisher-Hotelling<sup>7</sup> muestra la mayor capacidad de percepción de una señal entre todos los posibles modelos de observador que estén limitados a realizar sólo operaciones lineales sobre los datos. Este modelo nos proporciona un formalismo para determinar la separabilidad de las imágenes en dos estados (presencia o no de la imagen buscada) en el caso de que el fondo o los objetos que han de ser detectados presenten variabilidad. Para manejar esta variabilidad, este modelo es estadístico. Si las características estadísticas de los fondos se conocen, como es el caso en el uso de fondos sintéticos, el cálculo del modelo es sencillo.

El índice de detección para el Observador Hotelling es:

$$SNR_{HOT}^2 = \int \frac{|\Delta \bar{f}(v)|^2 \cdot |\bar{H}(v)|^2}{W_g(v)} \cdot dv \quad (\text{Ec. 2})$$

donde:

- $\Delta f(\nu)$  = diferencia de las transformadas de Fourier de las dos señales *medias* que se están analizando, es decir, FT (valor medio de (señal + fondo)) – FT (valor medio de fondo).
- $\tilde{H}(\nu)$ : transformada de Fourier de la matriz de transferencia espacial del sistema.
- $\tilde{W}_g(\nu)$ : es la transformada de Fourier de la covarianza *media* de las imágenes usadas para obtener datos estadísticos del fondo y de las imágenes utilizadas para obtener datos estadísticos de la señal.

Nótese que si existe ruido correlacionado en el sistema, éste se ve compensado por el factor  $\tilde{W}_g(\nu)$  ya que está presente también en el factor  $\Delta f(\nu)$ . De esta forma, este observador puede convertir el ruido del sistema en ruido blanco puro, eliminando la componente correlacionada. En este sentido, comparte con el Observador Ideal el filtro que “blanquea” el ruido del sistema. Por lo tanto, también se considera como un Filtro Adaptado de No Blanqueo (*PreWhitening Matched Filter*), aunque en este caso es no-ideal. Pese a su carácter no-ideal, se mantiene como límite superior de las prestaciones humanas para sistemas de imagen médica, con prestaciones menores o iguales a las del Observador Ideal.

El modelo presenta ciertos problemas:

- 1) **Enmascaramiento.** Si bien el modelo es válido para distribuciones de ruido no gaussianas, en presencia de fondo estructurado sus prestaciones sobrepasan a las del observador humano. Este hecho es debido en parte al fenómeno de enmascaramiento<sup>18</sup>, presente en humanos y no reproducido por este modelo. En este fenómeno, cuando las estructuras del fondo tienen tamaños parecidos a los de la señal, ésta se “enmascara” y se reduce su probabilidad de detección, dando lugar a curvas detección/tamaño de señal con pendiente negativa, es decir, a mayor tamaño de señal, menor probabilidad de detección.

- 2) **El problema estadístico.** Si se utilizan fondos reales que no puedan ser descritos por parámetros estadísticos, como es el caso de los fondos muy estructurados, la cantidad de muestras que se ha de analizar para tener estimaciones estadísticas del modelo es inmanejable. El número teórico de imágenes que se necesita para calcular la matriz de covarianza media es del orden de 10 a 100 veces el número de píxeles de la imagen, es decir, en el caso de una mamografía de 25 Megapíxeles implica el análisis de un mínimo de 250 millones de imágenes.

### **Soluciones propuestas al problema del enmascaramiento**

Este fenómeno está relacionado con el efecto de blanqueo de la señal a infinitas frecuencias espaciales que proporciona el Observador Hotelling, dado que compensa el ruido en todas las frecuencias. Los estudios psicofísicos y fisiológicos de la función visual indican que el observador humano carece de esa capacidad. Para ello, se han introducido modelos basados en la respuesta del sistema visual humano a canales discretos de frecuencia espacial “*channelized models*”<sup>19</sup>. En ellos se descomponen las frecuencias espaciales de entrada en canales discretos y sobre ellos se calcula la respuesta del sistema. Esta canalización es compleja y en los estudios realizados hasta ahora puede realizarse de dos formas que han presentado un relativo éxito.

- a) Modelo *Fisher-Hotelling with Rect Function Channels and Averaging* (FHC<sub>AVG</sub>)<sup>20</sup>. Divide la señal de entrada en canales de frecuencia completamente separados por simplicidad matemática. Sin embargo, la aplicación directa de este proceso se basa en una premisa irreal, ya que los estudios psicofísicos indican que la respuesta del observador humano a las frecuencias espaciales no actúa separadamente sobre cada una de ellas. Además, y más importante, los estudios experimentales muestran que no existe correlación entre este planteamiento y los resultados obtenidos por humanos. Para evitar estos problemas, se añade una fuente de ruido que simula el ruido neuronal interno del observador. Se incluye un filtro (*eye filter*) cuya respuesta en frecuencias espaciales es similar a la del ojo humano. Por último se descompone la señal en canales de frecuencia y se introduce un factor de ponderación sobre todos los canales igual a la respuesta media sobre un rango de una octava, valor aproximadamente igual a la separación entre canales obtenida para observadores humanos.

- b) Modelo *Fisher-Hotelling with Difference-of-Mesa Filter Channels* ( $FHC_{DOM}$ )<sup>20</sup>. Utiliza canales de frecuencia no separados. Esta aproximación casa bien con los análisis teóricos sobre el comportamiento visual humano y con los resultados experimentales. Como se ha comentado antes, su modelación matemática no es sencilla.

Los modelos matemáticos son complejos y dependen del ajuste de bastantes parámetros, a los que los dos modelos son sensibles, particularmente el modelo  $FHC_{DOM}$ .

### **Soluciones propuestas al problema estadístico**

Existen dos técnicas principales. La primera es limitar la “ventana” de búsqueda<sup>21</sup>, comparando zonas específicas de la imagen en lugar de la imagen completa. Esto implica definir ventanas de búsqueda a priori de la señal, lo que no coincide con la labor del radiólogo y devalúa el modelo. Aun con esta reducción, evaluar ventanas mayores de 20x20 píxeles es computacionalmente inmanejable, ya que exige el cálculo de la inversa de la matriz de la covarianza media de 4.000 imágenes.

La segunda técnica<sup>20</sup> también implica limitar la ventana de búsqueda y además tomar un número menor de muestras y simular el resto de las adquisiciones mediante técnicas de Monte Carlo o de “bootstrapping”, que no dejan de ser aproximaciones estadísticas con sus inconvenientes y limitaciones. En cualquier caso, el número de muestras que se debe adquirir puede oscilar entre 70 y 700 para ventanas que coincidan en tamaño con lesiones significativas, y no es realista adquirir ese número de lecturas en un sistema de mamografía en un entorno hospitalario cada vez que se necesite realizar pruebas de calidad de imagen.

Los modelos Hotelling-Fisher con la adición de canales presentan un buen ajuste a la calidad percibida por el observador humano en fondos estructurados relativamente sencillos, pero fallan a la hora de predecir el comportamiento del observador humano en fondos altamente estructurados en los que se incluyen agrupaciones locales de estructuras diseminadas de forma aleatoria (*clustered lumpy backgrounds*), tipo de fondos que coinciden con los que presentan las mamografías reales<sup>22</sup>.

### 2.3. El Observador Filtro Adaptado de No Blanqueo (NPWMF)

El observador NPWMF8 es un observador sub-óptimo. Comparte con el modelo del Observador Ideal el uso de toda la información disponible en los parámetros de la imagen, pero a diferencia del Observador Ideal es incapaz de realizar ninguna correlación en el ruido presente en la imagen y convertirlo en ruido blanco. Por tanto, tiene menor capacidad de discriminación en las situaciones en las que el ruido está “coloreado”, es decir, en los casos en que el NPS es dependiente de la frecuencia espacial.

El interés en este observador se debe a que existen evidencias de que el observador humano no puede decorrelacionar el ruido presente en una imagen. Además, en tareas no SKE/BKE su cálculo resulta mucho más sencillo que el del Observador Ideal, siempre que los parámetros de la señal estén especificados de manera estadística.

La expresión matemática del índice de detección para el Observador NPWMF es:

$$SNR_{NPW}^2 = \frac{[\int |\Delta f(v)|^2 \cdot |H(v)|^2 dv]^2}{\int |\Delta f(v)|^2 \cdot |H(v)|^2 W_g(v) dv} \quad (\text{Ec. 3})$$

Nótese que el factor  $\Delta f(v)$  en el numerador y el denominador impide decorrelar la señal de ruido correlacionado si éste estuviera presente.

En el espacio discreto es mucho más clara la simplicidad de este modelo. El índice de detección cuando analizamos la imagen como una matriz de píxeles es:

$$SNR_{NPW}^2 = (\Delta \mathbf{g}_{med})^t \mathbf{g} \quad (\text{Ec. 4})$$

Donde  $\Delta \mathbf{g}_{med}$  es la diferencia de los valores medios de las matrices de píxeles señal+fondo y el fondo, y  $\mathbf{g}$  es la matriz de píxeles de la imagen que se desea analizar.

El modelo NPWMF no necesita tantas muestras estadísticas como el modelo Hotelling para presentar correlaciones aceptables con el observador humano (obsérvese que en la expresión discreta de la Ec. 4 se elimina el cálculo de la matriz de covarianza). La mayor parte de los autores usa del orden de 60 a 70 imágenes. Sin embargo, en fondos estructurados presenta problemas de falta de ajuste con los resultados obtenidos por humanos por el problema del enmascaramiento, que no es detectado por este modelo. Para ello, se han introducido *eye filters*<sup>23</sup> que ajustan la respuesta en frecuencias del sistema a las del ojo humano. La inclusión de este filtro mejora la respuesta del modelo. Sin embargo, estudios de Burgess et al.<sup>20</sup> demuestran que no existe correlación entre este modelo corregido e imágenes con fondos estructurados más cercanos a los que se manejan en la realidad del entorno hospitalario.

#### **2.4. El modelo JNDMetrix<sup>24</sup>**

Este modelo de observador fue desarrollado por la empresa Sarnoff Corporation y enfocado en su inicio a tareas de reconocimiento de imágenes militares. Más tarde fue aplicado con éxito a tareas industriales y un equipo especialista de esta empresa lo adaptó al ámbito de la imagen médica.

A diferencia de los modelos anteriores, su aproximación no se fundamenta en la Teoría Estadística de la Decisión, siendo su aproximación de tipo *bottom-up* basada en modelos psicofisiológicos del proceso de la visión. Parte de una imagen test y una imagen de referencia e indica el nivel de visibilidad para un humano de la imagen de referencia en la imagen test. Presenta diez pasos en el proceso de análisis de las imágenes:

- 1) Óptico. Función que aproxima la dispersión de punto del ojo humano.
- 2) Muestreo. Función que simula el muestreo realizado por los conos del ojo humano.
- 3) Proceso de la luminosidad de la imagen. Normaliza la señal en unidades de contraste y divide la señal de entrada en 7 frecuencias espaciales.
- 4) Niveles piramidales. Convoluciona la señal con 8 pares de filtros espacialmente orientados derivados de datos psicofísicos.

- 5) Filtrado de pares de imágenes. Simula la transformación que se realiza en el córtex al pasar de células de respuesta lineal a células con respuesta a los niveles de energía.
- 6) Fase de transducción. Para cada uno de los niveles piramidales aplica la función específica de contraste del humano para ese nivel de contraste y luminosidad.
- 7) Normalización.
- 8) Salida de transducción, en la que se ajusta la señal a la sensibilidad de la fovea.
- 9) Análisis de distancia. Análisis de la distancia de los vectores espacialmente orientados para cada una de las posibles posiciones espaciales.
- 10) Mapa espacial JND. Produce un único valor que indica el nivel de discriminación de la señal.

La complejidad del sistema es evidente, aunque presenta la enorme ventaja de no plantear ninguna exigencia a la imagen que se desea analizar, ni de tipo de señal ni de tipo estadístico. Las correlaciones presentadas con el observador humano son excelentes para cualquier señal y en cualquier entorno.

Sin embargo, aparte de la complejidad señalada, los detalles del algoritmo no están disponibles por completo para la comunidad científica. Por último, destacamos que su adaptación médica se ha cancelado y la empresa ya no proporciona este programa.

### 3. FUNDAMENTOS TEÓRICOS

#### 3.1. El índice SSIM

Sean  $\mathbf{x} = \{x_i / i = 1, 2, \dots, N\}$  e  $\mathbf{y} = \{y_i / i = 1, 2, \dots, N\}$  dos zonas extraídas de la misma localización espacial de las dos imágenes X e Y que se desea comparar<sup>9</sup>. Sean  $\mu_x, \mu_y$  los valores medios de píxel en las zonas  $\mathbf{x}$  e  $\mathbf{y}$ . Sean  $\sigma_x, \sigma_y$  y  $\sigma_{xy}$  la desviación típica de  $\mathbf{x}$ , la desviación típica de  $\mathbf{y}$  y la covarianza de  $\mathbf{x}$  e  $\mathbf{y}$  respectivamente.  $\mu_x$  está directamente relacionada con la luminosidad media de los píxeles de la imagen  $\mathbf{x}$ ,  $\sigma_x$  es una estimación del contraste entre los píxeles de la imagen  $\mathbf{x}$ . La covarianza  $\sigma_{xy}$  nos proporciona una estimación de la tendencia de  $\mathbf{x}$  e  $\mathbf{y}$  a variar conjuntamente. Este parámetro nos da una indicación de la similitud estructural de ambas imágenes. Los índices de comparación de la luminosidad, contraste y estructura de las zonas extraídas  $\mathbf{x}$  e  $\mathbf{y}$  se definen<sup>11</sup> como:

$$m(x,y) = (2 \mu_x \mu_y + C_1) / (\mu_x^2 + \mu_y^2 + C_1) \quad (\text{Ec. 5}) \quad (\text{índice de luminosidad})$$

$$v(x,y) = (2 \sigma_x \sigma_y + C_2) / (\sigma_x^2 + \sigma_y^2 + C_2) \quad (\text{Ec. 6}) \quad (\text{índice de contraste})$$

$$r(x,y) = (\sigma_{xy} + C_3) / (\sigma_x \sigma_y + C_3) \quad (\text{Ec. 7}) \quad (\text{índice de estructura})$$

donde  $C_1, C_2$  y  $C_3$  son constantes que se introducen para evitar inestabilidades en las ecuaciones cuando los factores  $(\mu_x^2 + \mu_y^2)$ ,  $(\sigma_x^2 + \sigma_y^2)$  o  $\sigma_x \cdot \sigma_y$  son próximos a cero. La forma general del índice SSIM, que mide la similitud entre la zona  $\mathbf{x}$  e  $\mathbf{y}$  en una escala de valores de 0 (completamente diferentes) a 1 (imágenes iguales), se define como el producto ponderado de los tres índices de similitud:

$$\text{SSIM}(x, y) = [m(x, y)]^\alpha \cdot [v(x, y)]^\beta \cdot [r(x, y)]^\gamma \quad (\text{Ec. 8})$$

donde  $\alpha, \beta$  y  $\gamma$  son parámetros utilizados para definir la importancia relativa de cada uno de los componentes.

### 3.2. El índice de correlación cruzada $R^*$

Cuando  $\sigma_{xy} \ll C_3$  y  $\sigma_x \sigma_y \ll C_3$ , el índice de estructura  $r(x,y)$  es muy cercano a 1, indicando una gran similitud entre las imágenes  $x$  e  $y$ . Sin embargo esta indicación es falsa, ya que este comportamiento puede ocurrir cuando se está próximo al nivel umbral de percepción, es decir, cuando la señal evaluada es difícilmente distinguible del fondo  $y$ , por tanto, la diferencia de esta imagen es muy grande comparada con la imagen de referencia. Este es el caso de las imágenes en radiodiagnóstico y, en concreto, de las imágenes del maniquí CDMAM. Este problema se debe a la introducción de la constante  $C_3$  en la Ec. 7. El uso de las constantes  $C_i$  tiene sentido en el cálculo del índice SSIM, ya que este índice está enfocado a la evaluación de la percepción para imágenes muy alejadas del umbral de percepción.

En una versión alternativa del índice SSIM, denominada MS-SSIM<sup>\*13</sup> las constantes  $C_i$  tienen un valor nulo. Con ello el valor del índice no es cercano a 1 (máxima similitud) cuando nos acercamos al umbral de percepción. La componente de correlación cruzada del índice MS-SSIM<sup>\*</sup> ( $R^*$ ), relacionada con el índice de similitud estructural  $r(x,y)$ , muestra una fuerte correlación con los niveles de percepción umbral<sup>12</sup>. Sin embargo, los experimentos realizados hasta el momento<sup>13</sup> indican que el componente asociado al valor medio (índice de luminosidad) y el asociado a la desviación típica (índice de contraste) no son predictores válidos del problema del umbral de percepción de una imagen. Por tanto, estos dos últimos valores no son utilizados para evaluar dicho umbral de percepción.

Para calcular el índice  $R^*$  se define en primer lugar el índice  $r^*$ , que es el valor de  $R^*$  para la visión a una distancia y escala determinadas. Este valor se deduce del valor de  $r(x,y)$  eliminando la constante  $C_3$  e introduciendo valores específicos de  $r(x,y)$  para aquellos casos en los que  $\sigma_x \cdot \sigma_y = 0$  y que dividen por cero la Ec. 7. Supongamos que  $\sigma_x > 0$  y que el valor de la luminosidad en la zona  $y$  ( $\mu_y$ ), es constante. En este caso, la desviación típica de  $y$ ,  $\sigma_y$ , es cero. En estas circunstancias,  $y$  no está correlacionada con  $x$ , luego el valor del componente estructural  $r^*(x,y)$  debe ser cero. Cuando las dos zonas tienen la misma desviación típica ( $\sigma_x = \sigma_y$ ), el componente estructural debe adquirir un valor de 1, ya que las imágenes son idénticas estructuralmente ( $r^*$  tiene un valor de cero

para una desemejanza completa y un valor igual a 1 para imágenes idénticas). El índice  $r^*$ , por tanto, queda definido como:

$$r^*(x,y) = \begin{cases} 0 & \sigma_x > \sigma_y = 0, \text{ o bien, } \sigma_y > \sigma_x = 0 \\ 1 & \sigma_x = \sigma_y = 0 \\ r(x,y) & \text{en cualquier otro caso} \end{cases} \quad (\text{Ec. 9})$$

para  $r(x, y)$  como se definió en la Ec. (7).

La percepción de los detalles de una imagen depende, entre otros factores, de la resolución de dicha imagen y de la distancia de la imagen al observador. Un método de escala simple, como lo es el índice SSIM, es sólo apropiado para distancias de visualización y resoluciones específicas. Un método multiescala es un medio conveniente para incorporar a los índices de comparación de dos imágenes distintas resoluciones que simulan distintas escalas de visión.

Tomando la imagen de referencia y la imagen que se desea evaluar como entradas del sistema, el algoritmo aplica iterativamente un filtro paso-bajo y reduce el tamaño de la imagen de referencia y de la imagen que se evalúa en un factor de 2 M-1 veces<sup>9</sup>. La evaluación global aplicando este nuevo índice ( $R^{*13}$ ) se obtiene combinando las medidas a diferentes escalas de acuerdo con la siguiente expresión:

$$R^* = \prod_{j=1}^M r_j^*(X,Y) \quad (\text{Ec. 10})$$

donde  $r_j^*(x,y)$  se define de acuerdo con la Ec. 9, evaluando el valor de cada extracto de imagen  $(\mathbf{x}, \mathbf{y})$  sobre las imágenes completas y para cada j-ésima escala. Para la escala  $j=1$  no se aplica ningún factor de reducción. El índice  $R^*$  muestra una fuerte correlación con el umbral de percepción de una imagen<sup>13</sup>.

El valor máximo del factor de escala M en el presente trabajo es 5, que es el resultado de considerar  $M = \log_2$  (máximo diámetro en píxeles de los discos) + 1, y tomando 20 píxeles como el máximo valor del diámetro. Mayores reducciones de la imagen hacen desaparecer toda la información de los detalles que se evalúan.

## 4. MATERIAL Y MÉTODOS

### 4.1. El maniquí CDMAM

El maniquí CDMAM está formado por una base de aluminio con discos de oro de diferentes diámetros y espesores. Esta base de aluminio está recubierta con una funda de metacrilato (PMMA). Los discos de oro están colocados en una matriz de 16 filas por 16 columnas. Dentro de una fila, el diámetro de los discos es constante, con incrementos logarítmicos del espesor. Dentro de cada columna, el espesor de los discos es constante, con un incremento logarítmico del diámetro. Cada celda contiene dos discos idénticos, uno en el centro y otro en una esquina seleccionada al azar. La tarea del observador es detectar ambos discos, el central y el excéntrico.

Las imágenes del maniquí CDMAM utilizadas en este estudio han sido descargadas del web de *European Reference Organisation for Quality Assured Breast Screening and Diagnostic Services* (EUREF), sección *CDMAM Readout*<sup>25</sup>. Es un conjunto de 8 imágenes adquirido con un sistema GE Senograph 2000D, a 27 kVp, 125 mAs y con una resolución de 1 píxel por cada 100  $\mu\text{m}$ . Las imágenes fueron evaluadas por cuatro observadores humanos expertos y estas evaluaciones están disponibles en el mismo web y en la misma sección de EUREF. En el presente trabajo las imágenes han sido evaluadas con el algoritmo propuesto ( $R^*$ ) y con el programa CDCOM. Se comparan los resultados obtenidos con ambos algoritmos, así como con los que resultan de aplicar a los resultados del programa CDCOM los métodos propuestos (PRCDCOM) para tratar de aproximar los resultados a los de los observadores humanos<sup>26</sup> (Ver Anexo A para una descripción detallada de los algoritmos CDCOM y PRCDCOM).

### 4.2. Evaluación de las imágenes

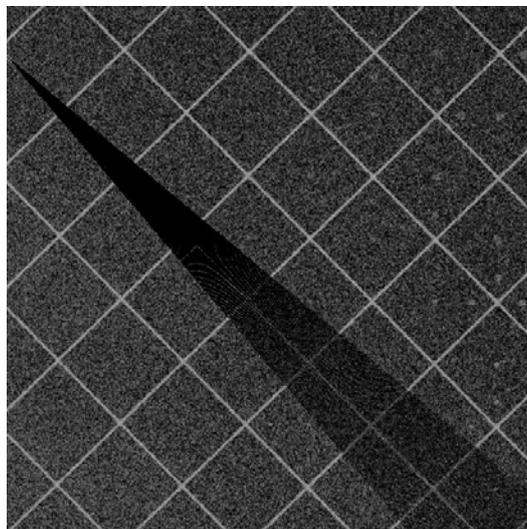
La aplicación del algoritmo  $R^*$  requiere la realización de dos tareas independientes. La primera de ellas es la localización de la rejilla que forman las líneas del maniquí CDMAM, ya que los discos se encuentran a distancias conocidas de los puntos de cruce de las líneas que constituyen esta rejilla.

La segunda tarea es localizar de forma exacta los discos y aplicar el algoritmo de percepción propuesto para evaluar qué esquina presenta mayor probabilidad de contener un disco.

#### 4.2.1. Localización de la rejilla

Para poder encontrar la posición de los discos dentro de la imagen del maniquí CDMAM es necesario determinar la posición de la rejilla que forman las filas y columnas. Han sido propuestos varios métodos para detectar dichos puntos<sup>6</sup> y en este trabajo se ha usado una metodología propia que localiza los puntos de cruce de la rejilla con un error del orden de  $\pm 2$  píxeles con respecto a la posición real<sup>27</sup>.

El primer paso es seleccionar una región de interés (ROI) cuadrada en el centro de la imagen, con la dimensión del lado igual a un tercio de la imagen completa. De esta forma se puede asegurar que en este ROI no existe información alfanumérica o zonas no expuestas cuya luminosidad pueda afectar al algoritmo de búsqueda.



**Fig. 2.** Abanico de rectas dentro del ROI central.

Dentro de este área se hace un barrido sobre todos los píxeles de la primera y última columna. Para cada píxel se considera un abanico de líneas rectas, cuyo origen está en el citado píxel y el final en el otro extremo del ROI. El abanico de rectas se calcula con un margen de inclinación entre  $35^\circ$  y  $55^\circ$ , espaciada cada recta en escalones de un cuarto de grado (Fig. 2). Para cada una de las rectas del abanico se calcula la suma del valor de

todos los píxeles. La recta que presenta el valor máximo de todas estas sumas es aquella que coincide en ángulo y posición con una de las líneas de la rejilla. Los valores máximos obtenidos por la aplicación de este algoritmo a la columna de la izquierda permiten detectar las líneas de rejilla con pendiente negativa. Este cálculo de máximos aplicado a la última columna permite detectar las líneas de rejilla con pendiente positiva. Los valores encontrados para una y otra columna permiten calcular la dimensión de las diagonales (D) que forman los rombos de la rejilla. Este valor D es distinto en las columnas de la derecha que en las columnas de la izquierda. Esta diferencia se debe probablemente a distorsiones geométricas introducidas por los sistemas radiográficos.

Usando los datos de ángulo y valor de la diagonal, se extrapolan las líneas de rejilla hasta los bordes de la imagen completa. Para asegurar que los resultados obtenidos coinciden con la extrapolación hasta el borde, se exploran los puntos teóricos de intersección de las líneas de la rejilla con el borde del maniquí. Se repite el proceso en pasos de  $D \pm 10$  píxeles alrededor de los puntos señalados, buscando el punto de mejor encaje entre la recta calculada y la posición real de la rejilla. Este proceso se repite para los dos lados del maniquí. De acuerdo con los datos obtenidos para cada una de las rectas, se calculan los puntos de cruce. En los experimentos realizados, la distancia encontrada entre los puntos calculados y los reales oscilan entre cero y un píxel. Sólo en un subconjunto reducido de puntos de cruce ( $< 1\%$ ) la distancia fue igual o mayor a 2 píxeles.

Las principales propiedades de este algoritmo son:

- 1) Baja complejidad computacional. El índice de complejidad computacional para los cálculos que se realizan en el ROI central es del orden de  $30 \times n$ , donde  $n$  es el número de píxeles de la imagen completa. Esta complejidad es del orden de  $n/10$  para los cálculos asociados al resto de la imagen<sup>1</sup>.
- 2) No existe ninguna necesidad de preproceso, incluso para imágenes muy ruidosas.
- 3) El algoritmo es muy robusto. El porcentaje de éxito en la detección de la rejilla en los experimentos realizados fue del 100%. Se comprobó el algoritmo con un total

---

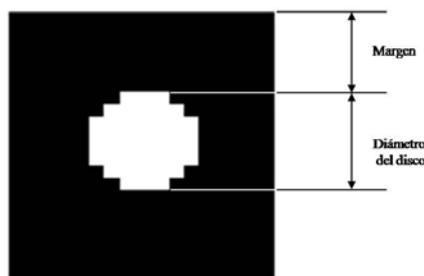
<sup>1</sup> El tiempo empleado en encontrar la rejilla en una imagen de  $1628 \times 2280$  píxeles, con un nivel de grises de 16 bits por píxel, es aproximadamente de 0,5 segundos, evaluado en un ordenador portátil Dell Inspiron 4400, con procesador Intel Centrino Core2 Duo T7200, 2 Ghz, 2 Gb RAM.

de 100 imágenes, obtenidas con diferentes equipos de distintos fabricantes (LORAD-HOLOGIC, GE MEDICAL SYSTEMS, AGFA, FUJI) y con diferentes niveles de ruido. El índice de ruido (desviación estándar / valor medio del píxel) medido en una esquina sin ningún tipo de señal (fondo uniforme) oscilaba entre valores de 0,010 y 0,025. El ángulo de las líneas de rejilla se encontraba entre 43° y 47°.

#### 4.2.2. Localización de los discos y cálculo del índice R\*

Utilizando la información obtenida sobre la posición de la rejilla y el conocimiento previo del diámetro y la localización exacta (con referencia a la citada rejilla) de los discos de cada celda, se construye una máscara que se ajusta a la posición real de los discos en cada una de las cuatro posibles esquinas de cada celda. Dicha máscara constituye la imagen de referencia que se utiliza para aplicar el algoritmo R\*.

Ya que el índice R\* proporciona una medida de la similitud estructural, para examinar las posiciones de los discos decidimos usar como máscaras estructuras con grandes diferencias visibles. La estructura de los discos del maniquí CDMAM es, llevada a su extremo, la de un disco completamente blanco sobre un fondo completamente negro. Por tanto, las imágenes de referencia son discos blancos con un diámetro igual al del disco que se evalúa y con un borde negro de 3 píxeles (Fig. 3).

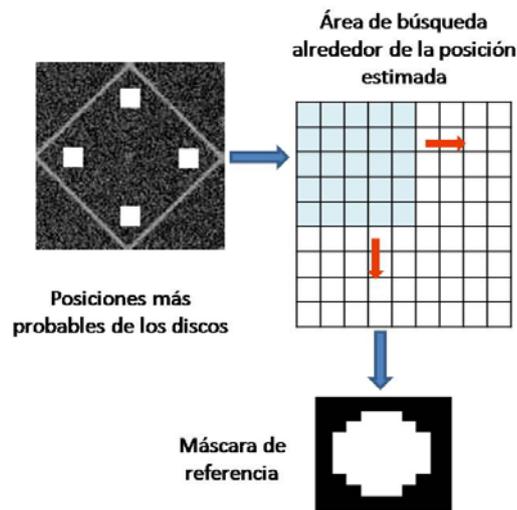


**Fig. 3.** Imagen de referencia.

Debido a la geometría del sistema de adquisición, las celdas del maniquí no son cuadrados reales, sino más bien trapecios y, además, la longitud de sus diagonales varía a lo largo de la imagen del maniquí. Incluso hemos encontrado que los centros de las imágenes de los discos no se encuentran a lo largo de las diagonales de las celdas, llegándose a desplazar hasta 4 o 5 píxeles de estos ejes de teórica simetría. Por tanto, es

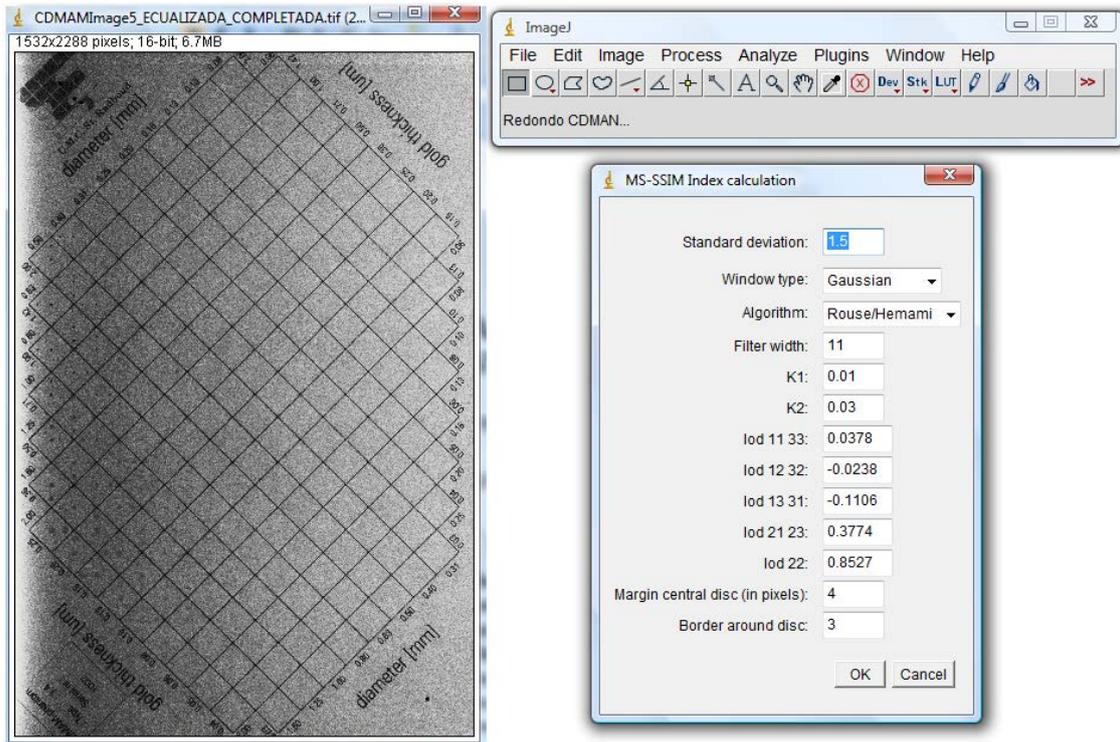
necesario fijar una zona de búsqueda alrededor de la posición teórica esperada de los discos para poder encontrarlos con cierto margen de seguridad. Dado el nivel de efectividad conseguido en encontrar la posición de los puntos de cruce de la rejilla ( $\pm 2$  píxeles de error máximo en general), esta región de seguridad es de  $\pm 3$  píxeles para las imágenes utilizadas en este trabajo, excepto en dos imágenes; en ellas el margen de seguridad se ha elevado a  $\pm 4$  píxeles, ya que el error de posición en alguno de los puntos de cruce de la rejilla era de  $\pm 3$  píxeles.

Se estima el valor del índice  $R^*$  dentro de cada esquina de la celda hasta en 25 posiciones alrededor de la posición esperada del disco (Fig. 4). El máximo valor de  $R^*$  obtenido de las distintas posiciones dentro de cada esquina se toma como el valor  $R^*$  de esa esquina. La esquina seleccionada como la de mayor probabilidad para encontrar el disco es aquella con el mayor valor  $R^*$ . Siguiendo este procedimiento, se obtiene una matriz de posiciones calculadas que puede ser comparada con la máscara previamente construida (Fig. 4).

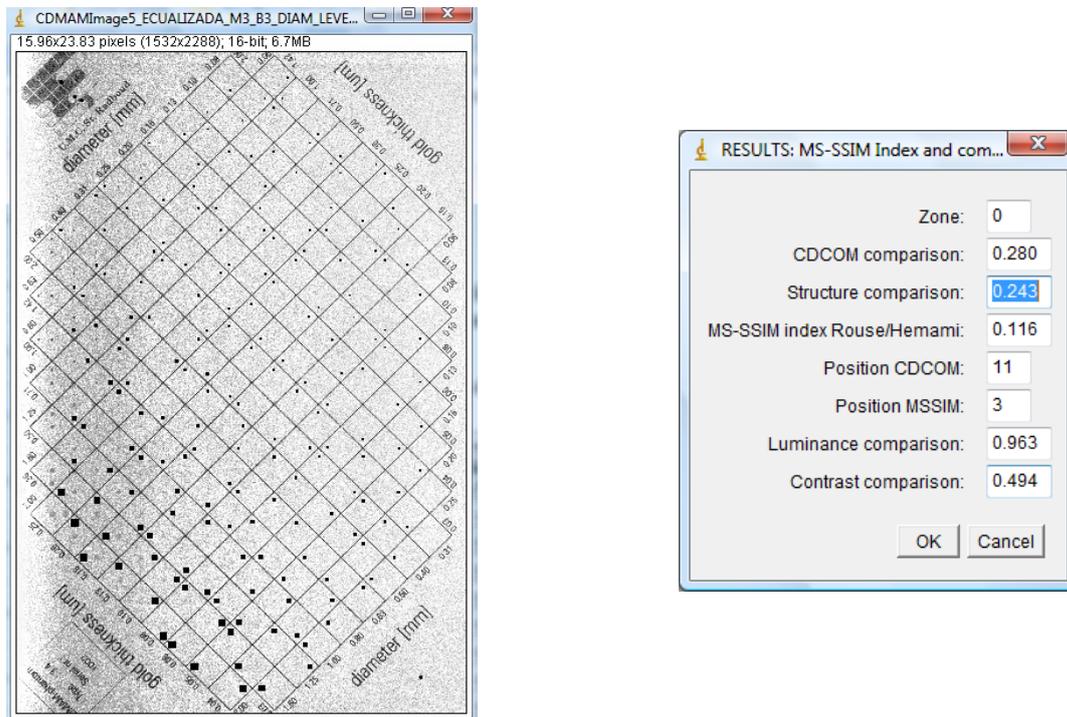


**Fig. 4.** Método de búsqueda.

El programa completo se desarrolló en lenguaje Java como un plugin de ImageJ<sup>15</sup>. En el futuro se publicará como OSS en la web de ImageJ. En la sección de plugins de este web<sup>28</sup> y en el web de nuestro grupo de trabajo<sup>29</sup> ya se han publicado resultados intermedios (índices SSIM, MS-SSIM, rotación del CDMAM) en forma de código abierto.



**Fig. 5.** Pantalla de selección de parámetros del programa.



**Fig. 6.** Determinación gráfica automática de las esquinas con mayor probabilidad de contener un disco.

**Fig. 7.** Valores obtenidos para la esquina con mayor probabilidad de contener un disco. El programa considera la zona con el valor máximo de  $R^*$  ("Structure comparison" en la figura) como zona (esquina) de posición más probable.

La Fig. 5 muestra la interfaz de usuario y los parámetros que pueden modificarse en la presente versión. La Fig. 6 muestra la salida gráfica obtenida como resultado de aplicar el algoritmo  $R^*$ . Como puede verse, quedan marcadas aquellas esquinas que el programa considera como más probables para la localización de un disco. La Fig. 7 nos indica los valores cuantitativos obtenidos para la exploración de cada celda, comparando el valor de distintos índices de calidad y percepción. Para cada imagen se han aplicado las reglas NNC (Corrección del Colindante más Cercano)<sup>30</sup>. Estas reglas tienen el efecto de suavizar el borde entre celdas que han sido evaluadas con acierto y aquellas que no lo han sido. Estas mismas reglas se recomiendan en el manual de utilización del maniquí CDMAM para que sean aplicadas a los resultados obtenidos en la evaluación de las imágenes por observadores humanos. A continuación se calcula el valor medio de los umbrales de espesor asociados a cada diámetro tras evaluar varias imágenes del CDMAM adquiridas en idénticas condiciones.

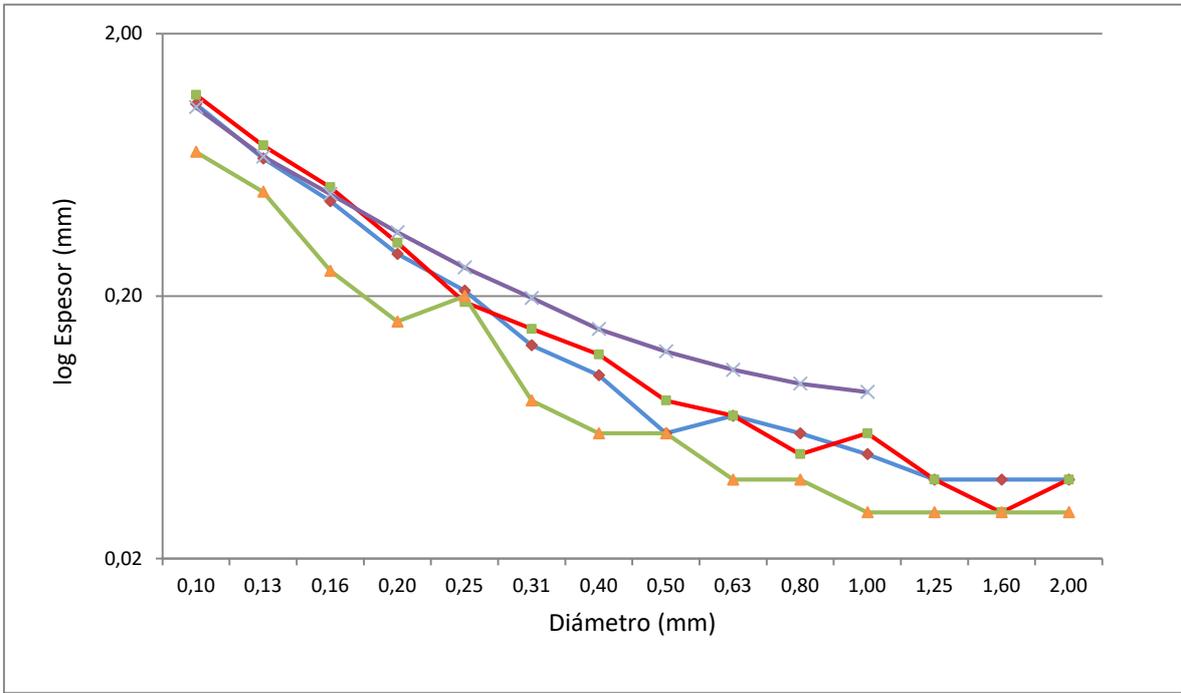
Se señala que la simplicidad estructural del modelo de obtención de resultados basado en  $R^*$  reduce posibles fuentes de error producidas por el tratamiento matemático y por la necesaria introducción de constantes, incluso en el caso de que éstas hayan sido obtenidas como resultados experimentales y no como valores ad hoc. En este sentido,  $R^*$  utiliza muy pocas constantes para producir sus resultados.

## 5. RESULTADOS

En la tabla 1 y en la figura 8 se muestra el espesor umbral necesario para percibir los discos a diferentes diámetros. Se presentan los valores obtenidos con el algoritmo propuesto y con otros tres métodos: observadores humanos (OH), CDCOM y PRCDKOM. La tabla 1 muestra asimismo las desviaciones existentes entre los tres métodos automáticos analizados y las observaciones realizadas por observadores humanos. No se incluyen resultados para los discos inferiores a 0,1 mm, debido a que están fuera del límite de detección de la mayoría de los observadores humanos evaluados. La figura 9 presenta los espesores umbrales obtenidos por el observador humano y mediante el uso del algoritmo R\*. Sobre estos datos se han sobrepuesto barras de error a +/- 1 desviación estándar. En la figura 10 se presentan las líneas de tendencia asociadas a los resultados de los observadores humanos y del índice R\*. Ambas líneas corresponden a polinomios de tercer grado.

**Tabla 1.** Valores medios de espesor umbral obtenidos con el índice R\*, observadores humanos (OH), CDCOM y PRCDKOM y desviaciones (%) de los espesores predichos por el índice R\*, CDCOM y PRCDKOM con respecto a los resultados obtenidos por observadores humanos.

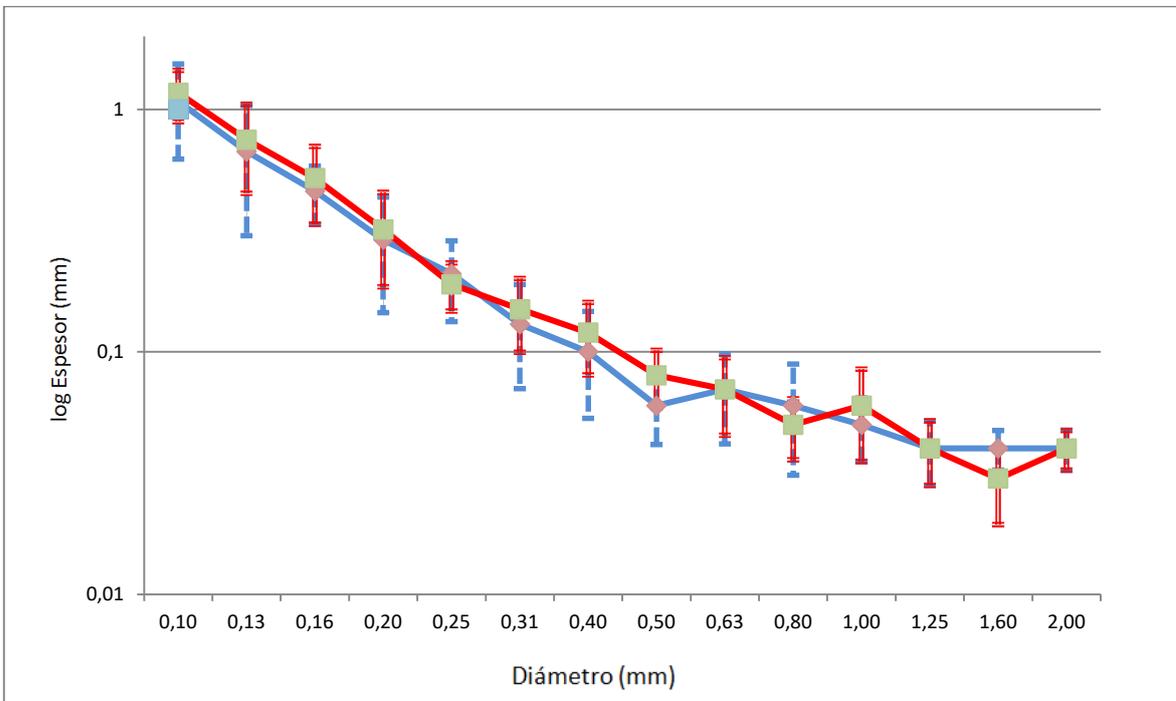
Ø mm	OH	Índice R*	Desviación R* / OH	CDCOM	Desviación CDCOM / OH	PRCDKOM	Desviación PRCDKOM / OH
<b>0,10</b>	1,17	1,08	-7,7%	0,71	-39,3%	1,05	-10,2%
<b>0,13</b>	0,75	0,67	-10,7%	0,50	-33,3%	0,68	-9,2%
<b>0,16</b>	0,52	0,46	-11,5%	0,25	-51,9%	0,49	-5,7%
<b>0,20</b>	0,32	0,29	-9,4%	0,16	-50,0%	0,35	9,6%
<b>0,25</b>	0,19	0,21	10,5%	0,20	5,3%	0,26	35,5%
<b>0,31</b>	0,15	0,13	-13,3%	0,08	-46,7%	0,20	31,3%
<b>0,40</b>	0,12	0,10	-16,7%	0,06	-50,0%	0,15	24,9%
<b>0,50</b>	0,08	0,06	-25,0%	0,06	-25,0%	0,12	53,9%
<b>0,63</b>	0,07	0,07	0,0%	0,04	-42,9%	0,10	49,8%
<b>0,80</b>	0,05	0,06	20,0%	0,04	-20,0%	0,09	85,8%
<b>1,00</b>	0,06	0,05	-16,7%	0,03	-50,0%	0,09	43,9%
<b>1,25</b>	0,04	0,04	0,0%	0,03	-25,0%	NA	NA
<b>1,60</b>	0,03	0,04	33,3%	0,03	0,0%	NA	NA
<b>2,00</b>	0,04	0,04	0,0%	0,03	-25,0%	NA	NA
<b>Promedio</b>			<b>-3,4%</b>		<b>-36,7%</b>		<b>28,1%</b>



- ◆ Índice R\*
- OH
- ▲ CDCOM
- × PRCDCOM

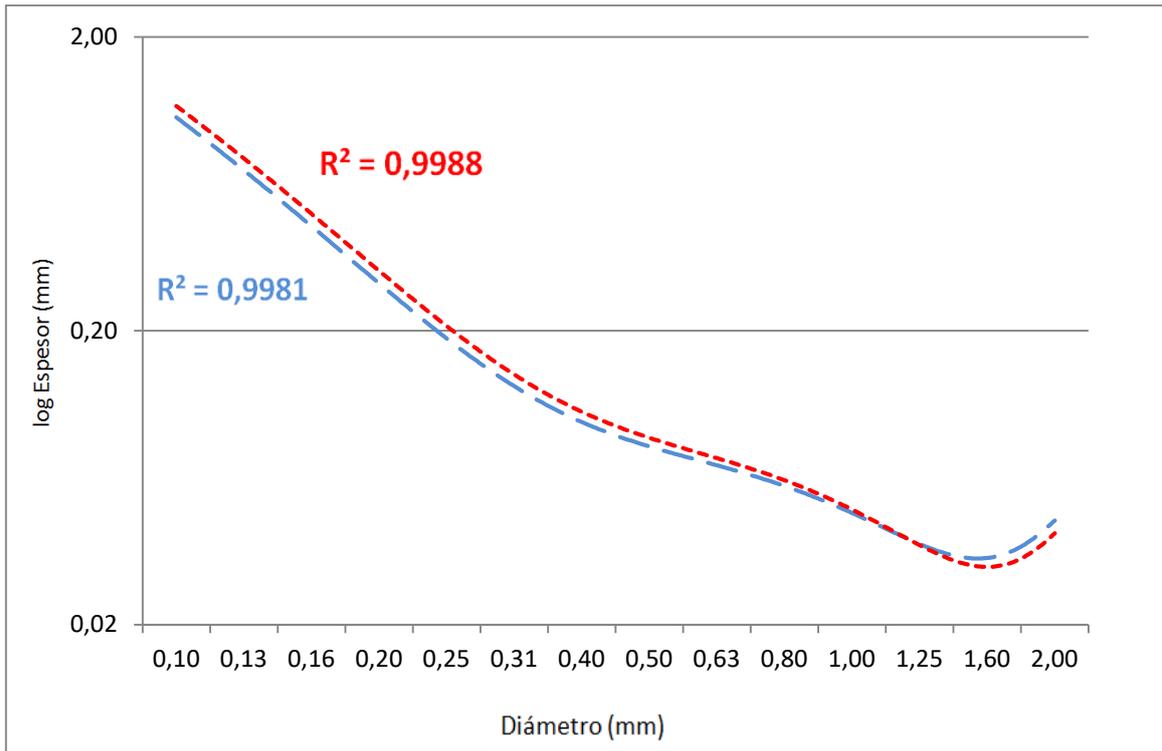
**Fig. 8.** Curva contraste detalle para las imágenes del maniquí CDMAM.

Coefficiente de Pearson OH/CDCOM = 0,980.  
 Coeficiente de Pearson OH/PRCDCOM = 0,998.  
 Coeficiente de Pearson OH/R\* = 0,999.



- ◆ Índice R\*
- OH

**Fig 9.** Curva contraste-detalle para las imágenes del maniquí CDMAM. Datos para observadores humanos (OH) e índice R\* con barras de error a +/- 1 desviación estándar.



— Polinómica (Índice R\*)  
 - - - Polinómica (OH)

**Fig. 10.** Líneas de tendencia de los resultados obtenidos por observadores humanos y con el índice R\*, con indicación del coeficiente de regresión R<sup>2</sup>. Ambas líneas corresponden a polinomios de tercer grado.

## 6. DISCUSIÓN

Los resultados de los observadores humanos fueron obtenidos por expertos entrenados en la tarea de evaluar el maniquí CDMAM y están disponibles en el web de EUREF. Estos resultados son la referencia para los usuarios del maniquí, de forma que los observadores se validan como evaluadores de las imágenes del CDMAM utilizando estos datos. En este sentido, los citados expertos se pueden considerar como observadores de referencia y sería deseable que cualquier proceso automático de evaluación ajustara sus resultados a los obtenidos por ellos. Por ello a continuación se analizan los datos obtenidos en el transcurso de esta investigación frente a los resultados obtenidos por esos expertos.

La figura 8 muestra una buena correlación entre los niveles de espesor umbral calculados mediante la aplicación del índice  $R^*$  y los detectados por los observadores de referencia. El coeficiente de Pearson (0,999) muestra un estrecho ajuste entre ambas series de datos.

La figura 9 indica una alta correlación no sólo en las graficas de resultados medios, sino en las desviaciones estándar. Las barras de error se sobreponen de manera constante a lo largo de toda la gráfica y muestran un elevado grado de solapamiento.

Las líneas de tendencia de la figura 10 presentan un excelente valor del coeficiente de regresión  $R^2$ , lo que indica la fiabilidad de su ajuste a los datos obtenidos con el índice  $R^*$  y por los observadores humanos. La correlación que presentan ambas líneas de tendencia entre sí muestra un elevado nivel de ajuste entre los dos conjuntos de resultados.

Señalamos que el único procesamiento aplicado posteriormente al algoritmo  $R^*$  es el de las reglas NNC (Corrección del Colindante más Cercano). Estas reglas tienen el efecto de suavizar el borde entre celdas que han sido evaluadas con acierto y aquellas que no lo han sido. Sin embargo, estas mismas reglas se aplican a los maniqués corregidos por observadores humanos, tal y como se indica en el manual de utilización del maniquí CDMAM<sup>30</sup>, luego no añaden ningún procesamiento añadido a aquel seguido en las evaluaciones manuales. Aparte de este proceso, no se ha utilizado ningún algoritmo de

suavizado específico para eliminar valores aleatorios ni se ha realizado ningún ajuste a ninguna función.

La comparación de la curva contraste-detalle de los distintos sistemas automáticos indica que el rango de validez del índice  $R^*$  es mayor que el del algoritmo PRCDCOM. La correlación entre  $R^*$  y el observador humano se mantiene para diámetros mayores de 1 mm, donde el programa PRCDCOM no es aplicable. La desviación media del algoritmo  $R^*$  con respecto a los resultados obtenidos por los observadores humanos ronda el 4%, mientras que la del algoritmo PRCDCOM se sitúa en un 28% (tabla 1), diferencia debida principalmente a las desviaciones de este último algoritmo para los diámetros mayores. También  $R^*$  está mejor relacionado con el observador humano que el programa CDCOM. Este algoritmo presenta en todo el rango de diámetros valores de sensibilidad mayores que los de los observadores humanos, resultado coherente con los estudios existentes<sup>26</sup>.

Sin embargo, para determinados diámetros las correlaciones entre el observador humano y  $R^*$  son hasta cierto punto aleatorias (tabla 1) y otros estudios han obtenido con otros conjuntos de imágenes mejores correlaciones entre las salidas del programa PRCDCOM y las evaluaciones realizadas por observadores humanos. Aunque estas correlaciones son algo peores que las que presenta  $R^*$  en este estudio (medias de desviación en los citados estudios del 7% de PRCDCOM frente a medias de  $R^*$  de 3,5%), su variabilidad con respecto a la observación humana a lo largo de los distintos diámetros es menor que la que presenta  $R^*$ . Por ello, es necesario desarrollar experimentos que impliquen un mayor número de imágenes y observadores.

Además, los siguientes pasos de esta investigación pasan por evaluar imágenes progresivamente más complejas. En un primer escenario, se analizarán imágenes formadas por fondos y señales no uniformes, sintetizados ambos por software. Estas imágenes simularán fondos y lesiones mamográficas reales. En un segundo paso, el análisis se realizará sobre fondos y lesiones mamográficas reales, analizando el grado de validez del índice  $R^*$  en este entorno.

## 7. CONCLUSIONES

- 1) Se han encontrado excelentes correlaciones medias entre los resultados obtenidos con el índice  $R^*$  y los obtenidos por observadores humanos.
- 2) Se han encontrado excelentes correlaciones medias entre la variabilidad de los resultados obtenidos con el índice  $R^*$  y la variabilidad de los resultados obtenidos por observadores humanos.
- 3) Se ha elaborado un programa que podría superar a otros desarrollos actuales que son estándares de facto del maniquí CDMAM.
- 4) Se han logrado altas velocidades de proceso computacional en la aplicación del citado programa.

## ANEXO A. El programa CDCOM

En el desarrollo de este programa se han utilizado dos modelos de percepción. Inicialmente<sup>31</sup> se utilizó el modelo de Observador Ideal para evaluar la posición de los discos. Más tarde<sup>32</sup> se utilizó el modelo NPWMF. Este modelo también ha sido utilizado por otros autores en otros sistemas de evaluación de este maniquí<sup>33</sup>.

Cuatro métodos distintos de procesamiento de los resultados del programa CDCOM (PRCDCOM) han sido propuestos y analizados en la literatura<sup>26</sup>. El método que mejor resultado presenta en la actualidad sigue los siguientes pasos:

- 1) Se combinan los resultados obtenidos por el programa CDCOM para un mínimo de 8 imágenes (recomendable 16) en una matriz de probabilidad. En ella consta el porcentaje de discos del citado conjunto de 8 (o 16) imágenes que han sido leídos correctamente para cada celda. Es decir, en cada celda de la matriz se presenta la probabilidad de que un disco haya sido correctamente detectado en alguna de las lecturas de las citadas imágenes.
- 2) Se acepta como lectura segura la de aquellas celdas cuya probabilidad de detección supera el 62,5%. Esto evita el efecto de lectura al azar, que es de un 25%. El punto medio de probabilidad es  $(100 + 25)/2 = 62,5$ .
- 3) Se aplica un filtro de suavizado de 3x3 a la matriz de probabilidades (tabla A.1). El valor de cada celda se reemplaza por la media ponderada de su valor y el de las celdas adyacentes usando los pesos relativos que se muestran en la tabla. Este suavizado tiene el efecto de reducir las fluctuaciones aleatorias de los valores de las celdas, lo que hace más sencillo introducir interpolaciones posteriores.

1	2	1
2	4	2
1	2	1

**Tabla A.1.** Algoritmo de suavizado.

- 4) Finalmente, se ajustan los resultados a una curva psicométrica que relaciona los valores obtenidos en los pasos anteriores con la percepción real de un observador

humano. Para cada diámetro, la curva adopta la siguiente expresión, basada en los modelos de percepción realizados por Georg Rasch en 1960<sup>34</sup>:

$$P_i(t) = 0,25 + \frac{0,75}{1 + e^{f[-(C(t)-C_T)]}} \quad (\text{Ec. A.1})$$

- $t$  es el espesor del disco
- $P_i(t)$  es la probabilidad de detección de un disco de espesor  $t$
- $C(t) = \ln(1 - e^{-(\mu t)})$  se corresponde con el contraste.  $\mu$  es el parámetro de atenuación lineal del oro y se fija en  $0,190 \text{ mm}^{-1}$
- $C_T$  es el contraste correspondiente a  $P(t)=0,625$  (el umbral de contraste)
- $f$  es un parámetro libre que se determina en el ajuste y que ha de ser constante. Se trata del poder discriminante que es el parámetro de dificultad de percepción del disco de diámetro  $t$  y representa el punto en la escala de aptitud en el que la probabilidad de responder correctamente al ítem  $i$  es igual a 0,5

5) Después de estos cuatro ajustes, las lecturas automáticas procesadas presentan correlaciones apreciables con las lecturas efectuadas por observadores humanos. Sin embargo, este análisis automático sigue presentando valores de contraste umbral notablemente menores que los detectados por observadores humanos. K Young y otros<sup>26</sup> desarrollaron una segunda curva de ajuste de tipo potencial que correlacionaba los valores del umbral de contraste de un conjunto de observadores humanos con los valores obtenidos después de aplicar los pasos del anterior procedimiento. Esta correlación no lineal es del tipo:

$$T_{\text{Chumano}} = a[T_{\text{Cauto}}]^n \quad (\text{Ec. A.2})$$

Donde  $T_{\text{Chumano}}$  es el valor del umbral de contraste medio para un conjunto de observaciones humanas,  $T_{\text{Cauto}}$  es el valor del contraste umbral medio obtenido con el algoritmo descrito y  $a$  y  $n$  son coeficientes que se fijan para obtener la mejor correlación.

## REFERENCIAS

- <sup>1</sup> B. Girod, “What’s wrong with mean-squared error”, in *Digital Images and Human Vision*, A. B. Watson, ed., pp. 207–220, the MIT press, 1993.
- <sup>2</sup> Z. Wang, A. C. Bovik, and L. Lu, “Why is image quality assessment so difficult”, in *Proc. IEEE Int. Conf. Acoust. , Speech, and Signal Processing*, vol. 4, (Orlando), pp. 3313–3316, May 2002.
- <sup>3</sup> A.E. Burgess, “The Rose model, revisited”, *J Opt Soc Am. A Opt Image Sci Vis* 1999; 16: pp. 633– 46, 1999.
- <sup>4</sup> R. Shaw, “The equivalent quantum efficiency of the photographic process”, *J. Photog. Sci.* 11, pp. 199-204, 1963.
- <sup>5</sup> K. J. Myers, “Ideal observer models of visual signal detection”, in *Handbook of Medical Imaging, Physics and Psychophysics*, edited by J. Beutel, H. Kundel, R. Van Metter (SPIE, Bellingham, WA, 2000), Vol 1, Physics and Psychophysics, pp. 558-592, 2000.
- <sup>6</sup> H. H. Barrett, K. J. Myers, and R. F. Wagner, “Beyond signal detection theory”, *Application of Optical Instrumentation in Medicine XIV and Picture Archiving and Communications (PACS IV) for Medical Applications*, Newport Beach, CA Proceedings of the Society of Photo-optical Instrumentation Engineers, Bellingham, WA, Vol. 626, pp. 231–239, 1986.
- <sup>7</sup> R. D. Fiete, H. H. Barrett, W. E. Smith, and K. J. Myers, “Hotelling trace criterion and its correlation with human observer performance”, *J. Opt. Soc. Am. A* 4, pp. 945–953, 1987.
- <sup>8</sup> R. F. Wagner, D. G. Brown, M. S. Pastel, “Application of information theory to the assessment of computed tomography”, *Med. Phys.* 6, pp. 83–94, 1979.
- <sup>9</sup> ICRU Report 54, “Medical Imaging – The Assessment of Image Quality”, Bethesda, MD: International Commission on Radiation Units and Measurements, 1996.
- <sup>10</sup> W J. Lubin, “The use of psychophysical data and models in the analysis of display system performance”, *Digital Images and Human Vision*, ed. A. B. Watson (MIT Press, Cambridge, MA, 1993), pp. 163-178, 1993.
- <sup>11</sup> Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity”, *IEEE Trans. Image Processing*, vol. 13, pp. 600–612, Apr. 2004.
- <sup>12</sup> Z. Wang, E. P. Simoncelli, A. C. Bovick, “Multi-scale structural similarity for image quality assessment,” 37th IEEE Asilomar Conf. on Signals, Systems and Computers 37, 2003.

---

<sup>13</sup> D. M. Rouse and S. S. Hemami, “Analyzing the Role of Visual Structure in the Recognition of Natural Image Content with Multi-Scale SSIM”, *Proc. SPIE* Vol. 6806, Human Vision and Electronic Imaging 2008.

<sup>14</sup> R. Visser and N. Karssemeijer, “CDCOM Manual: software for automated readout of CDMAM 3.4 images”; (nota: CDCOM software, manual e imágenes de muestra están disponibles en [www.euref.org](http://www.euref.org) ), 14<sup>th</sup> May 2009.

<sup>15</sup> W. S. Rasband, ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, <http://rsb.info.nih.gov/ij/>. 1997-2007, 14<sup>th</sup> May 2009.

<sup>16</sup> A. E. Burgess, R. F. Wagner, R. J. Jennings, H. B. Barlow, “Efficiency of human visual discrimination”, *Science* 214, pp. 93-94, 1981.

<sup>17</sup> ICRU Report 54, “Medical Imaging – The Assessment of Image Quality,” Bethesda, MD: International Commission on Radiation Units and Measurements, 1996.

<sup>18</sup> G. Revesz, H. L. Kundel, M. A. Graber, “The influence of structured noise on detection of radiologic abnormalities”, *Invest. Radiol.* 9, pp. 479 – 486, 1974.

<sup>19</sup> K. J. Myers, H. H. Barrett. “Addition of a channel mechanism to the ideal-observer model”, *J. Opt. Soc. Am. A* 4, pp. 2447-2457, 1987.

<sup>20</sup> A. E. Burgess, X. Li, C. K. Abbey, “Visual signal detectability with two noise components: anomalous masking effects”. *J. Opt. Soc. Am. A* Vol. 14, Issue 9, pp. 2420-2442, 1997.

<sup>21</sup> M. P. Eckstein, C. K. Abbey, F. O. Bochud, “A practical guide to model observers for visual detection in synthetic and natural noisy images”, in *Handbook of Medical Imaging, Physics and Psychophysics*, edited by J. Beutel, H. Kundel, R. Van Metter (SPIE, Bellingham, WA, 2000), Vol 1, Physics and Psychophysics, pp. 593-626, 2000.

<sup>22</sup> F. O. Bochud, C. K. Abbey, M. P. Eckstein, “Visual signal detection in structured backgrounds. III. Calculation of figures of merit for model observers in statistically nonstationary backgrounds”, *J. Opt. Soc. Am. A* 17, 193–205, 2000.

<sup>23</sup> A. E. Burgess, “Statically defined backgrounds: performance of a modified Prewhitening matched filter model”, *J. Opt. Soc. Am. A* 11, pp. 1237-1242, 1994.

<sup>24</sup> W J. Lubin, “The use of psychophysical data and models in the analysis of display system performance” *Digital Images and Human Vision*, ed. A. B. Watson (MIT Press, Cambridge, MA, 1993), pp. 163-178, 1993.

<sup>25</sup> [www.euref.org](http://www.euref.org) ), 14<sup>th</sup> May 2009.

<sup>26</sup> K.C. Young, A. Alsager, J. M. Oduko, H. Bosmans, B. Verbrugge, T. Geertse, R. van Engen. “Evaluation of software for reading images of the CDMAM test object to assess digital mammography systems”, *Medical Imaging 2008: Physics of Medical Imaging*. Edited by Hsieh, Jiang; Samei, Ehsan. Proceedings of the SPIE, Volume 6913, pp. 69131C-69131C-11, 2008.

- 
- <sup>27</sup> G. Prieto, M. Chevalier, E. Guibelalde, “A CDMAM Image Phantom Software Improvement for Human Observer Assessment”, E.A. Krupinski (Ed.): *IWDM 2008*, LNCS 5116, pp. 181–187, Springer-Verlag Berlin Heidelberg 2008.
- <sup>28</sup> W. S. Rasband, ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA <http://rsb.info.nih.gov/ij/plugins/index.html> 1997-2007, 14<sup>th</sup> May 2009.
- <sup>29</sup> G. Prieto, M. Chevalier, E. Guibelalde, Departamento de Radiología de la Universidad Complutense de Madrid, SPAIN, 2008-2009. [http://www.ucm.es/info/fismed/Digital\\_Imaging/Digital\\_imaging.htm](http://www.ucm.es/info/fismed/Digital_Imaging/Digital_imaging.htm), 14<sup>th</sup> May 2009.
- <sup>30</sup> K. R. Bijkerk, M. A.O. Thijssen, Th. J. M. Arnoldussen, “Manual CDMAM-Phantom Type 3.4” (translation from the Dutch by S. van Woudenberg), University Medical Centre Nijmegen, July 2000.
- <sup>31</sup> N. Karssemeijer, M.A.O. Thijssen, “Determination of contrast-detail curves of mammography systems by automated image analysis” in *Digital Mammography*, ed. Doi K, Giger R, Nishikawa, Schmidt R A. Elsevier, Amsterdam, pp. 155-160, 1996.
- <sup>32</sup> W. J. H. Veldkamp et al., “The value of scatter removal by a grid in full field digital mammography”, *Med. Phys.* 30, pp. 1712-1718 , 2003.
- <sup>33</sup> R. Rico, S.L. Muller, G. Peter, “Automatic scoring of CDMAN a dose study”, *Proc. SPIE* 5034, pp. 164-173, 2003.
- <sup>34</sup> G. Rasch, “Probabilistic models for some intelligence and attainment tests”, in *Studies in Mathematical Psychology I*. Danmarks paedagogiske Institut, Copenhagen, 1960.