



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



⑪ Número de publicación: **2 310 157**

⑫ Número de solicitud: 200801943

⑬ Int. Cl.:

H04N 7/30 (2006.01)

H03M 7/30 (2006.01)

G06T 9/00 (2006.01)

H04N 7/32 (2006.01)

H04N 7/34 (2006.01)

⑭

PATENTE DE INVENCION

B1

⑮ Fecha de presentación: **20.06.2008**

⑯ Fecha de publicación de la solicitud: **16.12.2008**

Fecha de la concesión: **19.08.2009**

⑰ Fecha de anuncio de la concesión: **03.09.2009**

⑱ Fecha de publicación del folleto de la patente:
03.09.2009

⑲ Titular/es: **Universitat de València, Estudi General**
Avda. Blasco Ibáñez, nº 13
46010 Valencia, ES

⑳ Inventor/es: **Malo López, Jesús;**
Gutiérrez Aguado, Juan;
Camps Valls, Gustavo y
Luque Cobija, María Josefa

㉑ Agente: **No consta**

㉒ Título: **Compresión de imágenes en color basada en representaciones perceptuales no lineales y en aprendizaje de máquina.**

㉓ Resumen:

Compresión de imágenes en color basada en representaciones perceptuales no lineales y en aprendizaje de máquina.

Se propone un método para la compresión de imágenes en color que comprende los siguientes pasos: (1) expresar los coeficientes de una representación local espacio-frecuencial de los canales oponentes cromático y acromáticos en unidades de contraste perceptual, (2) aplicar una transformada no lineal con normalización divisiva por canal de forma que incluyan relaciones entre los coeficientes de los canales de contraste acromático y cromáticos, y (3) usar algoritmos de aprendizaje estadístico para seleccionar los datos relevantes de los canales normalizados no-linealmente.

Se han llevado a cabo experimentos psicofísicos y numéricos con el fin de estimar y validar los parámetros de las transformaciones de contraste y de normalización divisiva. Además, se compara el método propuesto experimentalmente con una implementación JPEG. En ambos casos se emplearon bloques de 16x16 en DCT, y en cada caso, se emplearon valores comparables para los parámetros relacionados con las transformadas cromáticas y dependientes de la frecuencia. En estas condiciones, los resultados experimentales proporcionan una ganancia media en la compresión respecto a JPEG del 50% para distintas medidas de distorsión (RMSE, SSIM y S-CIELab) en el rango realista y comercial de [1, 2.2] bits/pix.

También se reivindican el aparato y el código para la implementación del método propuesto.

Aviso: Se puede realizar consulta prevista por el art. 37.3.8 LP.

DESCRIPCIÓN

Compresión de imágenes en color basada en representaciones perceptuales no lineales y en aprendizaje de máquina.

Campo de la invención

La presente invención comprende los métodos, aparato y programa para comprimir imágenes en color, en particular a imágenes y a secuencias de imágenes en color.

Contexto de la invención

Actualmente el volumen de datos de imágenes se incrementa exponencialmente en un amplio rango de aplicaciones tales como remote sensing, fotografía digital y cámaras de vídeo, imagen médica, bibliotecas digitales y documentos, películas y vídeo conferencia. Ello plantea diferentes problemas y necesidades para transmitir, almacenar y recuperar imágenes. En consecuencia, la compresión de imágenes digitales es una tecnología crucial. Es importante resaltar que comprimir una imagen es significativamente diferente de comprimir otros tipos de datos, debido a sus propiedades estadísticas particulares, y por lo tanto la aplicación directa de otros métodos de compresión está lejos de producir resultados óptimos. Por lo tanto, el conocimiento estadístico del problema resulta de importancia extrema para desarrollar esquemas de codificación eficientes.

El estándar ISO/IEC 10918-1, conocido como JPEG [1], se ha convertido en el método de compresión más popular. Diferentes métodos se han propuesto para incrementar la compresión manteniendo el mismo nivel de calidad, bien basados en técnicas de aprendizaje máquina tales como redes neuronales o bien mediante la inclusión de conocimiento perceptual en el proceso de codificación.

La patente US Patent No. 5,005,206 [2] describe el uso de redes neuronales para codificación de imágenes. En el esquema que proponen, la imagen es definida mediante los pesos asociados a una red neuronal entrenada en un dominio lineal particular.

La patente US Patent No. 6,798,914 [3] también describe el uso de redes neuronales para compresión de imágenes. El método combina inteligencia artificial y redes neuronales para convertir los datos digitales en datos simbólicos que son comprimidos utilizando una codificación *run-length*. En todos estos esquemas, están presentes los problemas propios de utilizar redes neuronales, a saber: inestabilidad en el entrenamiento, selección del algoritmo apropiado para el entrenamiento, estructura de la red, y la elección de los parámetros libres.

WO 03/050959 [4] describe el uso de otro algoritmo de aprendizaje máquina, la máquina de vectores soporte (o SVM), trabajando sobre una transformación lineal -la Transformada Discreta de Coseno (DCT)- para compresión de datos con pérdida. El método es utilizado en particular para comprimir imágenes en niveles de gris. La introducción de SVM ofrece varias ventajas y control sobre el proceso de aprendizaje: interpretación de los parámetros libre, raleza en el espacio de aplicación, y no necesidad de heurísticas en el entrenamiento ya que el problema de optimización resultante es convexo y la solución única. Sin embargo, no se conoce la aplicación del método a imágenes en color o cómo el esquema podría tratarlas. Además, la aplicación directa de la SVM estándar en el dominio de trabajo especificado (el dominio lineal DCT) no es una buena elección. En [5] este último problema es resuelto introduciendo una SVM adaptativa que tiene en cuenta la diferente relevancia perceptual de los coeficientes en el dominio lineal. Ciertamente la inclusión de conocimiento perceptual (o estadístico) *a priori* es de capital importancia para la codificación de imágenes. En este sentido, la patente US Patent No. 5,426,512 [6] describe la tecnología DCTune para optimizar la compresión de imágenes estáticas. DCTune calcula las mejores matrices de cuantificación JPEG para lograr la máxima compresión posible para un error perceptual establecido, dadas una imagen particular y un conjunto de condiciones de visualización. Este método se basa en un modelo no lineal puntual de enmascaramiento (auto-enmascaramiento), pero no considera relaciones entre coeficientes (enmascaramiento cruzado) ni tampoco propone el uso de algún método de aprendizaje máquina.

Estos métodos producen resultados buenos y competitivos comparados con los métodos estándar JPEG. Sin embargo, tienen al menos las siguientes limitaciones: (i) hasta la fecha ninguno de los métodos integra conocimiento perceptual y métodos de aprendizaje máquina para compresión de imágenes en color; (ii) los métodos existentes no proponen los esquemas de transformación para trabajar con contraste en imágenes en color, o cómo pueden ser calculados e introducidos en los esquemas ya conocidos; (iii) los métodos existentes no introducen transformaciones no-lineales perceptuales con interacciones entre coeficientes antes de aplicar eventualmente un método de aprendizaje máquina; y (iv) el aprendizaje máquina no se formula para ser independiente del dominio de aplicación.

La presente invención pretende resolver estos inconvenientes.

El método y aparato de esta invención extiende las propuestas realizadas en [6, 7, 8] aplicando una transformada no-lineal perceptual que incluye interacción entre coeficientes tras una transformada espacio-frecuencial. Además, respecto a JPEG2000 [9], permite una matriz de interacción entre coeficientes general (no causal) incluyendo relaciones entre canales cromáticos. Además, el método propuesto resuelve los problemas encontrados en [8, 10] ya que en éstos no se aplica una etapa de obtención del contraste cromático con el consiguiente error en la transformada no

lineal (ya sea puntual o basada en la normalización divisiva). El método también difiere de los algoritmos basados en redes neuronales propuesto hasta el momento en las patentes US Patent No. 5005206 [2], y No. 6798914 [3], ya que se puede aplicar un algoritmo de aprendizaje de propósito general, que debe ser necesariamente adaptado a la relevancia estadística o perceptual de los coeficientes en el dominio de trabajo donde se realiza la codificación. En [5] el método de regresión con vectores soporte (SVR) se aplicó en el dominio DCT, como en [4, 12]. Sin embargo, dadas las características del dominio, el algoritmo fue adaptado para incorporar la relevancia perceptual de los coeficientes en este dominio, lo cual no había sido considerado en [4, 12]. La SVR ha sido aplicada también en un dominio perceptual no lineal para codificación de imágenes acromáticas en [13], y debido a las características del dominio usado (perceptualmente Euclídeo tal y como se analiza en [14]) se puede aplicar correctamente una SVR estática. Sin embargo, cuando se trabaja en otros dominios es necesario utilizar una SVR adaptativa como se ilustra en [5] y se analiza con más detalle en [14].

Resumen de la invención

Es un objeto de la presente invención proporcionar métodos, aparato y programas que definen un nuevo marco de propósito general para la codificación de imágenes en color.

Otro objeto de la presente invención es proporcionar, métodos, aparato y programas para mejorar la performance en compresión de imágenes en color para un esfuerzo computacional comparable con los métodos existentes.

En un primer aspecto, estos y otros objetos se consiguen con un método de compresión de imágenes en color que comprende los siguientes pasos:

- Recibir de los datos de la imagen en color expresados en canales RGB.

- Transformar de los canales RGB a un espacio de color oponente obteniendo un canal A, un canal RG y un canal YB.

- Transformar de los canales A, RG y YB a un dominio espacio frecuencia como una suma de funciones base multiplicadas por el conjunto correspondiente de coeficientes y codificar de los residuos de baja frecuencia.

- Expresar los coeficientes de la mencionada representación espacio-frecuencial local de los canales acromático y cromáticos oponentes en unidades de contraste con significado perceptual.

- Aplicar una transformada no-lineal paramétrica a los canales acromático y cromáticos expresados en contraste.

- Estimar los coeficientes de la mencionada transformada perceptual no-lineal usando un procedimiento de aprendizaje.

- Realizar la cuantificación de la descripción obtenida y codificar los resultados.

Preferentemente el método usa una de las siguientes transformadas espacio-frecuenciales locales: la DCT o cualquier forma de transformada discreta basada en funciones de Gabor o en wavelets.

Preferentemente el método calcula el contraste del canal acromático como una función del número de parámetros relevantes y el contraste de los canales cromáticos como una función del número de parámetros relevantes incluyendo parámetros obtenidos mediante experimentos sicofísicos y numéricos.

Preferentemente el método incluye la obtención de parámetros respecto a las proporciones de luminancia media de las funciones cromáticas mediante experimentos psicofísicos y la obtención de parámetros respecto a la modulación máxima del color mediante el uso de experimentos numéricos.

Preferentemente la transformada no lineal aplicada a los canales cromáticos es la normalización divisiva general, y las transformadas no lineales perceptuales aplicadas a los canales cromáticos son transformadas de normalización divisiva con sensibilidades frecuenciales adaptadas a estos canales.

Preferentemente el método usa un procedimiento de aprendizaje para identificar la secuencia de funciones base y sus pesos correspondientes, con un criterio apropiado para optimizar la compacidad de la secuencia de datos codificados y/o la calidad de la imagen codificada.

Preferentemente el método usa como procedimiento de aprendizaje la SVR o un procedimiento implementado con alguno de los siguientes procedimientos de aprendizaje máquina: redes neuronales, técnicas basadas en splines, métodos basados en kernel, redes Bayesianes, procesos Gaussianos o lógica borrosa.

Preferentemente el método codifica el flujo de datos usando codificación entrópica.

Opcionalmente el método incluye pasos adicionales para procesar la información de fase de forma separada de la información de amplitud de los coeficientes.

En un segundo aspecto, los objetos de esta invención son alcanzados mediante un programa almacenado en un medio que cuando es ejecutado en un dispositivo (computador, PDA, teléfono móvil, ...) lleva a cabo el método descrito anteriormente.

En un tercer aspecto, los objetos de esta invención son alcanzados mediante un aparato que produce datos comprimidos de imágenes en color comprendiendo el mecanismo para la recepción de los datos de la imagen en color y el mecanismo para realizar el procesado para llevar a cabo los siguientes pasos:

- Transformar de los canales RGB a un espacio de color oponente obteniendo un canal A, un canal RG y un canal YB.

- Transformar de los canales A, RG y YB a un dominio espacio frecuencia como una suma de funciones base multiplicadas por el conjunto correspondiente de coeficientes y codificar de los residuos de baja frecuencia.

- Expresar los coeficientes de la mencionada representación espacio-frecuencial local de los canales acromático y cromáticos oponentes en unidades de contraste con significado perceptual.

- Aplicar una transformada no-lineal paramétrica a los canales acromático y cromáticos expresados en contraste.

- Estimar los coeficientes de la mencionada transformada perceptual no-lineal usando un procedimiento de aprendizaje.

- Realizar la cuantificación de la descripción obtenida y codificar los resultados.

Otras características y ventajas de la invención presente quedarán claros a partir de la siguiente descripción detallada de las concreciones ilustrativas de sus objetos en relación con las figuras proporcionadas.

Descripción de las figuras

Figura 1 muestra un diagrama de bloques de un método acorde con la presente invención. Véase página 10 y siguientes donde se describe cada bloque.

Figura 2 muestra un ejemplo de una red sinusoidal puramente cromática en la dirección cromática amarillo-azul (YB). En este caso, se muestra una función base DCT de tamaño 16×16 con frecuencia $f_x = 8$ y $f_y = 0$ ciclos por grado. El color medio es $T_0 = [80 \ 80 \ 80]^T$ en RGB NTSC o $T_0 = [80 \ 0 \ 0]^T$ en YUV lineal, el contraste cromático es $C_{chrom}^{(2)} = 0.85$ asumiendo una modulación máxima $\Delta T_{2max} = 45$ en el canal U. La fila superior muestra los dos redes en contra fase de cromaticidad extrema $e_1^{(2)}$ (amarillo a la izquierda), y $e_2^{(2)}$ (azul en el centro), usados para generar la red cromática puro (a la derecha). Los valores medios de luminancia (ordenadas) de las redes amarilla y azul han sido escogidos para obtener luminosidad constante en la red final ($\eta_{YB} B = 0.14$, ver sección 2.2), el cual aísla la respuesta en el canal Amarillo-Azul ($\Delta T_2 \neq 0$ y $\Delta T_3 = 0$). La fila central muestra la luminancia de las redes mostrados en la primera fila. La gráfica inferior muestra las coordenadas cromáticas CIE xy de los colores extremos (luminancia unidad) $e_j^{(2)}$ con $j = 1, 2$ (estrellas) y los colores actuales en la red (círculos). Incrementar el contraste cromático significa incrementar simultáneamente las amplitudes de las luminancias $\Delta Y_j^{(2)}$, por tanto incrementando el rango de cromaticidad cubierto por los círculos en la gráfica más inferior. En el caso extremo (contraste unidad), el rango cromático podría ser determinado exactamente por los colores extremos $e_j^{(2)}$. Reducir el contraste cromático significa reducir las amplitudes de las luminancias $\Delta Y_j^{(2)}$, y por tanto se reduce el rango cromático. En el límite de contraste cero, el único color de la malla sería el color medio T_0 (el círculo central en el diagrama más inferior).

La figura 3 muestra imágenes de muestra que limitan la modulación máxima en las direcciones U y V (arriba) y los colores correspondientes (abajo).

La figura 4 muestra los parámetros $si(f)$ para cada canal en la ecuación 35.

La figura 5 muestra el parámetro $0(f)$ de la ecuación 35.

La figura 6 muestra tres ejemplos de vecindades de interacción entre coeficientes (filas de la matriz $h(f, f')$ reordenadas en 2-D) en la ecuación 35. Cada superficie corresponde con las vecindades asociadas a diferentes frecuencias.

La figura 7 muestra las respuestas y umbrales de visibilidad asociados a dos sensores ajustados a frecuencias 4 y 10 ciclos por grado en condiciones de auto-enmascaramiento (el fondo es la función base para la que está ajustado el sensor). También se muestra la cantidad de distorsión Δc que se debe aplicar para obtener un incremento de respuesta τ para diferentes contrastes (abscisas) del patrón de entrada.

La figura 8 muestra las respuestas y los umbrales de visibilidad asociados para dos sensores ajustados a frecuencias 4 y 10 ciclos por grado cuando existe un patrón de enmascaramiento (ordenadas) de diferente frecuencia (6 ciclos por grado) a diferentes contrastes (abscisas): 0 (auto-enmascaramiento, línea sólida) y 0.5 (línea punteada). En este caso la cantidad de distorsión requerida Δc para obtener un incremento de respuesta τ a un contraste del estímulo se incrementa cuando el contraste del enmascarador es incrementado. A: Estímulo; B: Máscara

La figura 9 muestra un conjunto de imágenes consideradas para evaluar las capacidades de compresión del método propuesto.

Las figuras 10a, 10b y 10c muestran las curvas medias de ratio de compresión frente a distorsión (abcisas: medida como entropía en bpp) sobre 25 imágenes usando diferentes medidas de distorsión (ordenadas): (a) RMSE, (b) S-CIELab [28], y (c) Structural SIMilarity index (SSIM) [27], para los métodos considerados: 16x16 JPEG (punteado) y el método propuesto C-NL-SVR (línea sólida).

La figura 11a muestra las curvas de distorsión (abcisas: medida como entropía en bpp) para dos imágenes de ejemplo (arriba y abajo) usando RMSE para los métodos 16 x 16 JPEG (punteada) y el C-NL-SVR (sólida). La figura 11b ilustra un esquema para medir la ganancia del método C-NL-SVR (línea sólida) frente a 16x16 JPEG (línea punteada). Para un nivel de distorsión, la diferencia en ratio entre los dos métodos es medida y comparada.

La figura 12 muestra la media en ganancia (con barras de desviación estándar) del método C-NL-SVR frente a 16x16 JPEG en términos de diferentes medidas de distorsión (abcisas): (a) RMSE, (b) SSIM [27] y (c) S-CIELab [28]; ordenadas: entropía (bpp).

La figura 13 muestra ejemplos de imágenes “Parrot” (zona del ojo) decodificadas con 16x16 JPEG (fila superior) y C-NL-SVR (fila inferior) a diferentes ratios de compresión (de izquierda a derecha: {1.0,1.4,1.8,2.2} bits por pixel, bpp).

La figura 14 muestra ejemplos de imágenes “Lena” decodificadas con 16x16 JPEG (fila superior) y C-NL-SVR (fila inferior) a diferentes ratios de compresión (de izquierda a derecha: {1.0,1.4,1.8,2.2} bits por pixel, bpp).

La figura 15 muestra ejemplos de imágenes “Roof” decodificadas con 16x16 JPEG (fila superior) y C-NL-SVR (fila inferior) a diferentes ratios de compresión (de izquierda a derecha: {1.0,1.4,1.8,2.2} bits por pixel, bpp).

La figura 16 muestra ejemplos de imágenes “Flower 3” decodificadas con 16x16 JPEG (fila superior) y C-NL-SVR (fila inferior) a diferentes ratios de compresión (de izquierda a derecha: {1.0,1.4,1.8,2.2} bits por pixel, bpp).

La figura 17 muestra ejemplos de imágenes “Face 1” decodificadas con 16x16 JPEG (fila superior) y C-NL-SVR (fila inferior) a diferentes ratios de compresión (de izquierda a derecha: {1.0,1.4,1.8,2.2} bits por pixel, bpp).

Descripción detallada de la invención

Con el fin de comprender mejor esta invención, y teniendo en cuenta que proporciona un marco de propósito general novedoso para codificar imágenes en color que combina técnicas conocidas y nuevas, vamos a describir en primer lugar un esquema completo de la invención, en segundo lugar describiremos en detalles los bloques clave centrándonos en los aspectos noveles e inventivos de esta invención, y en tercer lugar describiremos los resultados experimentales. Finalmente nos referiremos a la implementación del método.

1. Esquema de la invención

La figura 1 muestra un esquema general de la invención que incluye los siguientes bloques:

Bloque 1.- Iniciar los valores de los parámetros.

Bloque 1.A.- Selección por parte del usuario de la ratio de compresión (tamaño de la imagen codificada) o de la distorsión (nivel de calidad de la imagen codificada).

Bloque 1.B.- Selección de los parámetros propios del procedimiento de aprendizaje y de la resolución de la cuantificación aplicable a la descripción obtenida por el proceso de aprendizaje. Para ello, se tienen en cuenta las elecciones consideradas en el Bloque 1.A, por ejemplo mediante una look-up tabla (LUT) que relacione los parámetros libres y la distorsión/ratio sobre un conjunto de imágenes apropiado. En pasos posteriores, el Bloque 1.B permite la adaptación de los parámetros preestablecidos teniendo en cuenta la desviación respecto a la ratio/distorsión deseada y los valores actuales obtenidos en los Bloques 16 y 17. Por ejemplo, si el aprendizaje se realiza mediante una SVR, se deben establecer o modificar los siguientes parámetros: los valores de ϵ , el factor de penalización λ y los parámetros asociados al kernel K (véase el Bloque 9 y la sección 4.2) y el número de bits usados en la cuantificación de los pesos asociados a los vectores soporte seleccionados (véase el Bloque 10).

Bloque 2.- Recepción de los datos de la imagen en color expresados en canales digitales RGB (denotados como RGB).

Bloque 3.- Transformación de la señal RGB a un espacio de color oponente, comenzando con la caracterización del color dependiente del dispositivo digital, obteniendo en primer lugar una descripción del color en un espacio triestímulo independiente del dispositivo (tal como CIE XYZ). Ello requiere la los parámetros de calibración del dispositivo de adquisición. Si tales parámetros no están disponibles, se pueden utilizar curvas gamma y primarios

ES 2 310 157 B1

genéricos. En segundo lugar, las imágenes X, Y, Z son combinadas en los siguientes tres canales: Acromático (A), Rojo-Verde (RG), Amarillo-Azul (YB) según una transformación opONENTE lineal o no lineal.

Bloques 4.A, 4.B, 4.C.- Transformación de cada canal de color opONENTE, A, RG, YB a un dominio espacio-frecuencial local, usando por ejemplo, la transformada discreta del coseno (DCT), la transformada wavelet o cualquier otra transformada lineal o no lineal apropiada.

Bloques 5.A, 5.B, 5.C.- Obtención del contraste de cada canal con transformadas apropiadas (basadas en percepción), esto es, expresar las amplitudes de las señales de frecuencia local en las unidades de contraste apropiadas para usar modelos de percepción en los bloques 7.A, 7.B y 7.C. Para el canal A se puede utilizar el contraste de Michelson, pero para cada uno de los canales RG y YB es necesario realizar tanto experimentos sicofísicos con las funciones base cromáticas de las transformadas como experimentos numéricos sobre una base de datos de imágenes de color (véase sección 2).

Bloque 6.- Codificación de los residuos de baja frecuencia de los canales A, RG y YB. Esta codificación puede consistir en la aplicación de un DPCM seguido por un codificador entrópico.

Bloques 7.A, 7.B, 7.C.- Aplicación de una transformada no lineal perceptual a cada canal. En el caso del canal acromático y asumiendo que se ha utilizado la DCT en el bloque 4.A, se puede aplicar la transformación no lineal introducida en [11, 14]. En el caso de los canales cromáticos la transformación a aplicar debe ser diferente (véase la sección 3 para formas funcionales explícitas).

Bloque 8.- (Opcional) Almacenamiento de la información de fase original de los canales transformados no linealmente. Este bloque es necesario solamente si la información de la fase es codificada de forma independiente a la información sobre amplitud (relacionado con los Bloques 11 y 12).

Bloque 9.- Aplicación y entrenamiento de un procedimiento de aprendizaje máquina con los parámetros establecidos en el Bloque 1.B. Este procedimiento selecciona los coeficientes más representativos e informativos que mejor aproximan a la señal en cada canal. En el caso de que se siga el camino opcional hacia el Bloque 8, el proceso de aprendizaje se aplica sobre los valores absolutos de las señales. Si no es así, puesto que la señal de fase no se procesa de forma separada a la señal, el proceso de aprendizaje se aplica directamente a las señales que proceden de los Bloques 7.A, 7.B y 7.C. Véase la sección 4 para un ejemplo de entrenamiento usando SVR y guardando los pesos y vectores soporte asociados.

Bloque 10.- Realizar una cuantización de la descripción obtenida en el bloque 9 usando para ello los parámetros de cuantización obtenidos/ajustados en el bloque 1.B. En el caso de usar una SVR, se puede emplear una cuantización uniforme de los pesos asociados a los vectores soporte seleccionados.

Bloque 11 (Opcional)- Realización de una reconstrucción canal a canal a partir de los pesos cuantizados del Bloque 10. El presente Bloque se necesita tan solo si se codifica la información de fase de forma independiente de la información sobre amplitud (Bloque 8). El resultado de esta reconstrucción se utiliza para descartar la información de fase innecesaria que proviene del Bloque 8 en el Bloque 12. Para invertir las transformaciones no lineales de los Bloques 7.A, 7.B y 7.C hace falta un procedimiento adecuado. En el caso de que haya realizado una transformada con normalización divisiva [15], se puede utilizar un procedimiento recursivo [11]. Se debe asegurar que las no linealidades en 7.A, 7.B y 7.C sean invertibles. En [11] se probó que la transformada de 7.A lo era. El estudio de las condiciones necesarias para invertir las transformadas propuestas en los Bloques 7.B y 7.C puede verse en la sección 3.

Bloque 12 (Opcional)- Proceso de eliminación de la información de fase procedente del Bloque 8 (opcional), basado en los coeficientes descartados en el proceso de reconstrucción (Bloque 11, opcional), después de la cuantización realizada en el Bloque 10.

Bloque 13.- Siempre que se procesen por separado la información de fase y la de amplitud, el Bloque 13 toma la información de amplitud seleccionada (procedente del Bloque 10) y la de fase (procedente del Bloque 12) y las combina para su uso posterior en el Bloque 14. En caso contrario, el Bloque 13 se limita a transmitir al Bloque 14 los pesos cuantizados con su correspondiente información de fase.

Bloque 14.- Codificación de la entropía de las señales. Existe una variedad de mecanismos de codificación de la entropía que podrían utilizarse con este fin.

Bloque 15.- Generación de un flujo de datos a partir de las señales de entropía codificadas de los Bloques 6 y 14. El decodificador, que no se muestra aquí, debería invertir todo el proceso hasta reconstruir la imagen. Como se hizo notar en el Bloque 11, el único proceso de inversión que no es trivial es el asociado con las transformadas no lineales de los bloques 7.A, 7.B y 7.C. En la sección 3 se estudia la invertibilidad de las transformadas propuestas.

Bloque 16.- Cálculo de la entropía o de la distorsión.

Bloque 17.- Vuelta al Bloque 1.B, salvo que ya se hayan alcanzado los parámetros de calidad (distorsión) o tamaño (entropía) especificados previamente.

Bloque 18.- Almacenamiento o transmisión de la imagen en color codificada.

2. Cambio de la métrica de contraste (Bloques 5.A, 5.B, 5.C)

5 El primer elemento clave es el uso de una métrica de contraste adecuada en la representación local en frecuencias de los canales A, RG, y YB. Este procesamiento inicial es un paso previo necesario antes de aplicar las transformadas perceptuales: para que sea posible aplicar parámetros obtenidos de experimentos psicofísicos a las transformadas no lineales de los Bloques 7.A, 7.B y 7.C, la señal de entrada debe expresarse en las unidades de *contraste* adecuadas.

10 La literatura sobre experimentos psicofísicos [15, 16] describe la respuesta no lineal de los sensores del sistema visual humano frente a señales periódicas expresadas en unidades de contraste. Además, se ha mostrado que la forma de la curva de respuesta no-lineal de los sensores cromáticos, sintonizados a distintas posiciones y frecuencias, es similar a la encontrada para sensores acromáticos [16], siempre que el estímulo se describa en las unidades de contraste cromático adecuadas. Sin embargo, no existe un procedimiento general para definir dichas unidades para un conjunto particular de funciones base en el dominio espacial (tales como la DCT en bloques o una transformación *wavelet* particular) cuando la modulación del estímulo es puramente cromática. Por tanto, la transformación de los coeficientes de amplitud a unidades de contraste es el factor clave que permitirá: (1) realizar el escalado relativo de los parámetros de los canales acromático y cromáticos (por ejemplo, la sensibilidad relativa en función de la frecuencia [17]), y (2) diseñar los parámetros que controlarán la forma de las no linealidades cromáticas, tomando como punto de partida los del caso acromático. Las transformadas de contraste de los Bloques 5.A, 5.B y 5.C son claves para simplificar la formulación de los Bloques 7.A, 7.B y 7.C.

La definición de contraste en la literatura psicofísica está íntimamente ligada al procedimiento que se haya seguido para construir los estímulos usados en el experimento. Por ejemplo, el contraste cromático puede definirse a partir de la modulación de luminancia de dos funciones base de dos colores, que llamaremos “extremos”, $e_1^{(i)}$ y $e_2^{(i)}$, sumados en contrate para generar redes de luminosidad percibida constante [17]. Las variaciones máximas de luminancia y de color (la amplitud máxima de los estímulos, que definirían la unidad de contraste) se ve limitada por la gama de colores generables por el dispositivo de generación de colores particular utilizado en el experimento.

30 Para aplicar una definición de contraste comparable deberemos: (1) obtener las ecuaciones que relacionan las amplitudes de los coeficientes locales de frecuencia y la modulación máxima del espacio de color con la luminancia y los colores extremos de redes equiluminosas equivalentes, y (2) asegurar que los colores extremos calculados, con las ecuaciones descritas arriba, a partir de una base de datos de imágenes suficientemente amplia, se encuentran dentro de la gama de colores reproducibles por los dispositivos de reproducción habituales. La idea subyacente es definir la unidad de contraste a partir de las variaciones máximas de luminancia y color que se encuentran en las imágenes naturales y al mismo tiempo ser consistente con la definición de contraste cromático de la literatura psicofísica.

Según lo anterior, la definición de contraste cromático y acromático de los Bloques 5.A, 5.B y 5.C, se basan en dos pasos:

40 • En primer lugar, obtención de las ecuaciones que relacionan el contraste cromático y acromático con las amplitudes de la transformada de frecuencia local de los canales acromático y cromáticos (ver la subsección 2.1). Estas ecuaciones, para un canal cromático i dependen de la pareja de colores extremos $e_1^{(i)}$ y $e_2^{(i)}$, obtenidos a partir de la variación máxima, $\Delta T_{\max}^{(i)}$, en la dirección que aísla ese canal. Además, si se utiliza un modelo lineal de visión del color, para generar una red equiluminosa, la luminancia media de los colores que se mezcla no puede ser la misma, ya que colores de diferente cromaticidad tienen distintas luminosidades a igual luminancia. Por tanto, se precisa de un experimento psicofísico para determinar la proporción adecuada de luminancias en la mezcla. En la Sub-sección 2.2 se describe este procedimiento, así como los resultados particulares de un experimento diseñado para las funciones base de la DCT, utilizados en una implementación posible del invento.

50 • En Segundo lugar, estudio del rango de valores de la variación máxima $\Delta T_{\max}^{(i)}$, en una base de datos de imágenes representativa. A partir del estudio empírico mostrado en la sub-sección 2.3, estimamos la variación máxima que permite una definición de contraste unidad consistente con la literatura psicofísica. Estos resultados son un ejemplo para la transformada DCT particular elegida y para el modelo YUV lineal.

55 Por último, en la sub-sección 2.4 se resume la transformada resultante con la definición de contraste adecuada usada en los Bloques 5.A, 5.B y 5.C, cuando se toma como sistema de representación de imágenes y del color la DCT y el espacio YUV lineal.

60 2.1. Formulación: Contraste de redes acromáticas y cromáticas, obtenido a partir de los coeficientes de las transformadas de imágenes descritas en espacios de color oponentes

Las redes (funciones base particulares en el dominio espacial) de frecuencia f empleadas en experimentos psicofísicos, tienen un color diferente, descrito mediante un vector tridimensional T , en cada posición espacial, x :

$$65 \quad T(x) = T_0 + \Delta T(f) \cdot B(f, x) \quad (1)$$

donde T_0 es el color promedio, $\Delta T(f)$ representa la variación cromática máxima (pico) para esa frecuencia respecto al color promedio, y $B(f, x)$ es la función base seleccionada de f , p.e. una de las de la DCT. El índice f , que identifica a la función base, puede simbolizar algo más general que la frecuencia: en el caso de una transformada wavelet, f tendría tanto significado espacial como frecuencial. En la ec. 1, hemos añadido el índice f también a la variación cromática $\Delta T(f)$, porque en imágenes que contienen más de una función base la amplitud de la modulación cromática puede ser diferente para cada función base, $B(f, x)$.

En un espacio de color oponente y lineal, como, por ejemplo, el espacio YUV lineal, las tres componentes de los vectores $T = [T_1, T_2, T_3]^T$, son la luminancia, $T_1 = Y$, la componente Amarillo-Azul (YB), $T_2 = U$, y la componente Rojo-Verde (RG), $T_3 = V$. Las redes descritas arriba se utilizan en experimentos psicofísicos para estimular sensores acromáticos o cromáticos concretos, seleccionados modificando las componentes de la modulación $\Delta T(f)$. Las modulaciones $\Delta T_i(f)$, se pueden relacionar de forma sencilla con los coeficientes de la transformada espacial discreta de la imagen correspondiente con la ec. 1 (véanse las ecs. 20 más abajo). El problema radica en que los contrastes cromáticos en las direcciones Amarillo-Azul y Rojo-Verde, tal y como se encuentran definidos en la literatura psicofísica, [17], no se relacionan de forma trivial con $\Delta T_2(f)$ y $\Delta T_3(f)$. Por claridad, en lo que sigue, omitiremos de forma temporal el índice f en las modulaciones de color $\Delta T_i(f)$.

Comencemos por el caso más simple, el acromático. En este caso, no hay modulación en los canales cromáticos ($\Delta T_2 = \Delta T_3 = 0$), por lo que la red que aislaría al canal acromático es:

$$T^{(1)}(x) = \begin{pmatrix} T_{01} \\ T_{02} \\ T_{03} \end{pmatrix} + \begin{pmatrix} \Delta T_1 \\ 0 \\ 0 \end{pmatrix} \cdot B(f, x) \quad (2)$$

y el contraste acromático se define simplemente como el contraste de Michelson de la red acromática (sólo de luminancia):

$$C_{achrom} = \frac{\Delta T_1}{T_{01}} \quad (3)$$

donde T_{01} es la luminancia promedio (o la luminancia del color promedio), y la amplitud ΔT_1 se puede obtener a partir de los coeficientes de la transformada de frecuencia local del canal de luminancia correspondientes utilizando las Ecs. 20. Esta definición es extensible a bases wavelet más complejas, sin más que dividir cada coeficiente de cada sub-banda por el residuo pasa-baja correspondiente a esa resolución. Como resultado, el contraste acromático se encuentra en el rango $C_{achrom} \in [0, 1]$ ya que $\Delta T_1 \in [0, T_{01}]$.

Sin embargo, en [17], la definición del contraste cromático tiene en cuenta la forma en la que se diseñan experimentalmente las redes cromáticas. Para generar una modulación puramente cromática entorno a un cierto color promedio, T_0 , a lo largo de una dirección dada T_i , con $i = 2, 3$, se suman en contrafase (ver línea superior de la fig. 3) dos redes de luminancia con la cromaticidad de los colores extremos $e_1^{(i)}$ y $e_2^{(i)}$ (la luminancia de los cuales se considera unidad):

$$T^{(i)}(x) = e_1^{(i)} \left(\frac{T_{01}}{2} (1 + \eta_i) + \Delta Y_1^{(i)} \cdot B(f, x) \right) + e_2^{(i)} \left(\frac{T_{01}}{2} (1 - \eta_i) - \Delta Y_2^{(i)} \cdot B(f, x) \right) \quad (4)$$

donde las modulaciones de luminancia $\Delta Y_j^{(i)}$ están en los rangos:

$$\Delta Y_1^{(i)} \in [0, \frac{T_{01}}{2} (1 + \eta_i)] \text{ y } \Delta Y_2^{(i)} \in [0, \frac{T_{01}}{2} (1 - \eta_i)] \text{ , con las restricciones siguientes:}$$

- El color promedio se obtiene sumando los colores extremos $e_j^{(i)}$ con sus luminancias escaladas en una proporción particular (dada por el factor η_i):

$$T_0 = e_1^{(i)} \frac{T_{01}}{2} (1 + \eta_i) + e_2^{(i)} \frac{T_{01}}{2} (1 - \eta_i) \quad (5)$$

- Las modulaciones de luminancia $\Delta Y_j^{(i)}$, están ligadas entre sí:

$$\Delta Y_2^{(i)} = \frac{(1 - \eta_i)}{(1 + \eta_i)} \Delta Y_1^{(i)} \quad (6)$$

- La variación total en la dirección cromática i , $\Delta T^{(i)}$ no induce ninguna modulación en el canal cromático ortogonal. Esto es:

$$\Delta T^{(2)} = \begin{pmatrix} \eta_2 T_{01} \\ \Delta T_2 \\ 0 \end{pmatrix} \quad (7)$$

$$\Delta T^{(3)} = \begin{pmatrix} \eta_3 T_{01} \\ 0 \\ \Delta T_3 \end{pmatrix} \quad (8)$$

Nótese que, dada la diferencia entre *luminosidad* y *luminancia*, para obtener una red equiluminosa con dos redes de diferente color, estas deben tener distintas luminancias medias (lo que induce una modulación residual de luminancia $\eta_i T_{01}$). La variación respecto a la luminancia media (el factor η_i) debe determinarse experimentalmente para cada canal cromático.

La Sección 2.2 está dedicada a la obtención de este factor para funciones base de la DCT con frecuencia y color promedio dados.

En la literatura psicofísica, el contraste cromático en el canal i se define como el contraste de Michelson de cualquiera de las dos redes de luminancia sumadas en contratase que generan la red puramente cromática:

$$C_{chrom}^{(i)} = \frac{\Delta Y_1^{(i)}}{\frac{T_{01}}{2}(1 + \eta_i)} = \frac{\Delta Y_2^{(i)}}{\frac{T_{01}}{2}(1 - \eta_i)} \quad (9)$$

A partir de las Ecs. (1), (4) y (9), puede deducirse la relación entre la modulación cromática en la dirección cromática i , $\Delta T^{(i)}$, y el correspondiente contraste cromático:

$$\Delta T^{(i)} = \left(\mathbf{e}_1^{(i)} - \frac{1 - \eta_i}{1 + \eta_i} \mathbf{e}_2^{(i)} \right) \frac{T_{01}}{2} (1 + \eta_i) C_{chrom}^{(i)} \quad (10)$$

Si consideramos el caso de contraste cromático unidad (haciendo que $C_{chrom}^{(i)} = 1$) obtenemos la relación entre lo que se considera modulación máxima en ese canal, $\Delta T_{max}^{(i)}$, y los colores extremos, $\mathbf{e}_j^{(i)}$:

$$\mathbf{e}_1^{(i)} = \frac{1}{T_{01}(1 + \eta_i)} \left(\Delta T_{max}^{(i)} + T_0 \right) \quad (11)$$

$$\mathbf{e}_2^{(i)} = \frac{2}{T_{01}(1 - \eta_i)} \left(T_0 - \frac{T_{01}}{2} (1 + \eta_i) \mathbf{e}_1^{(i)} \right) \quad (12)$$

En los experimentos psicofísicos, la modulación máxima se elige de manera que los colores extremos queden dentro de la gama de colores generables por el dispositivo de reproducción utilizado. Siguiendo este criterio, en la sección 2.3 exploramos una base de datos de imágenes representativa para obtener una estimación de la modulación máxima posible entorno al color promedio de cada imagen, de forma que los colores extremos estén siempre dentro del rango de colores generables en monitores estándar (por ejemplo, dentro del triángulo que aparece representado en línea de puntos dentro del diagrama cromático de la fig. 3).

Usando la ec. 10, se puede derivar el contraste cromático $C_{chrom}^{(i)}$ a partir de la modulación ΔT_i :

$$C_{chrom}^{(i)} = \frac{\Delta T_i}{\left(c_1^{(i)} - \frac{1-\eta_i}{1+\eta_i} c_2^{(i)} \right)_i \frac{T_{01}}{2} (1 + \eta_i)} = \frac{\Delta T_i}{(\Delta T_{max}^{(i)})_i} \quad (13)$$

Conocidos los contrastes acromático y cromáticos en función de la modulación en cada canal oponente (ecs. 3 y 13), el último dato requerido para poder expresar las amplitudes de las transformadas de frecuencia local en función de los contrastes, es la relación entre los coeficientes de la transformada discreta y las modulaciones correspondientes.

En el caso de la DCT, esta relación puede encontrarse en la literatura [18]. Aquí, los índices x, f utilizados anteriormente se declaran de forma explícita como $x = (m, n)$ y $f = (p, q)$. Con esta notación, los coeficientes $a(p, q)$ de la DCT bidimensional, correspondientes a la representación de un bloque de imagen $M \times N$ en el dominio espacial, $A(m, n)$, son [18]:

$$a(p, q) = \beta_p \beta_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A(m, n) \cos \left(\frac{\pi(2m+1)p}{2M} \right) \cos \left(\frac{\pi(2n+1)q}{2N} \right) \quad (14)$$

donde $\beta_p = 1/\sqrt{M}$ for $p = 0$ and $\beta_p = \sqrt{2}/\sqrt{M}$ for $p > 0$

$\beta_q = 1/\sqrt{N}$ for $q = 0$ and $\beta_q = \sqrt{2}/\sqrt{N}$ for $q > 0$.

La inversa de la DCT se escribe como:

$$A(m, n) = \sum_{p=0}^{M-1} \sum_{q=0}^{N-1} \beta_p \beta_q a(p, q) \cos \left(\frac{\pi(2m+1)p}{2M} \right) \cos \left(\frac{\pi(2n+1)q}{2N} \right) \quad (15)$$

Si la imagen adquirida es una simple función base (en el canal cromático i) definida como sigue:

$$A_i(m, n) = T_{0i} + \Delta T_i(p, q) \cos \left(\frac{\pi(2m+1)p}{2M} \right) \cos \left(\frac{\pi(2n+1)q}{2N} \right), \quad (16)$$

entonces es posible escribir la relación apropiada entre los coeficientes de la imagen en el cana i , $a_i(p, q)$ y la modulación cromática real para esa frecuencia $\Delta T_i(p, q)$:

$$T_{0i} = \beta_o \beta_o a_i(o, o) = \frac{1}{\sqrt{MN}} a_i(o, o) \quad (17)$$

$$\Delta T_i(o, q) = \beta_o \beta_q a_i(o, q) = \frac{\sqrt{2}}{\sqrt{MN}} a_i(o, q) \quad (18)$$

$$\Delta T_i(p, o) = \beta_p \beta_o a_i(p, o) = \frac{\sqrt{2}}{\sqrt{MN}} a_i(p, o) \quad (19)$$

$$\Delta T_i(p, q) = \beta_p \beta_q a_i(p, q) = \frac{2}{\sqrt{MN}} a_i(p, q) \quad (20)$$

2.2. Experimento 1: Proporciones Azul-Amarillo y Rojo-Verde en redes de luminosidad constante

En esta sección, describimos los experimentos psicofísicos realizados para evaluar en qué relación deben estar las luminancias medias de las redes acromáticas utilizadas para generar redes puramente cromáticas -con luminosidad uniforme- en las direcciones U y V, esto es, los factores η_i de las ecuaciones anteriores.

Estímulos. Los estímulos se generaron en CRTs controlado por ordenador, con una tarjeta gráfica de 8 bits de resolución. Los sistemas estaban colorimétricamente calibrados y se controlaban usando Matlab®, con la librería de funciones COLORLAB (disponible en <http://www.uv.es/vista/vistavalencia>).

Los colores $e_j^{(i)}$ se escogieron en las direcciones del espacio de color que anulaban uno de los canales cromáticos del espacio YUV, verificándose que:

- Las modulaciones que producen a lo largo de la dirección i , a partir de un color promedio T_0 , sólo difieren en signo. El color promedio seleccionado era el promedio de una base de datos de imágenes pública [19]. En el espacio de color NTSC RGB tenemos que $T_0 = [132 \ 116 \ 90]^T$, con una luminancia de 117.8 cd/m^2 .

- La modulación inducida era la máxima permitida por la gama de colores del monitor. Dos redes DCT puras de luminancia, ambas de 8 cpd, y cada una con la cromaticidad de un color extremo, se generaron en contrafase, inicialmente con la misma luminancia media, y se sumaron. La imagen resultante, que subtendía 0.25 grados, se presentó sobre un fondo gris de 50 cd/m^2 en una habitación a oscuras.

Antes de comenzar las tandas de medidas, propiamente dichas, se modificó mediante sucesivos ajuste, la luminancia del punto acromático, T_0 , de forma que:

- La luminancia se mantuviese dentro del rango de valores posibles del monitor, en todos los píxeles de la imagen resultante.

- El rango de variación de luminancias de las redes, fijado por el valor promedio, fuese suficiente para que los observadores pudiesen modificar la luminancia de una de ellas hasta alcanzar la condición de luminosidad constante.

Con estas restricciones, la luminancia escogida para el color promedio fue de 80 cd/m^2 .

Medida. Las redes descritas anteriormente se mostraban al observador, quien las fijaba fovealmente, tras un período de adaptación de un minuto al fondo gris. La tarea del observador era ajustar la variable η_i , de modo que los dos semiperíodos de la red resultante pareciesen de igual luminosidad. Para ello, se utilizó el método de ajuste (MOA) [20]. Para cada observador, el resultado del experimento se obtuvo como el promedio de cinco tandas de medida. El resultado final es el promedio de un conjunto de 4 observadores no entrenados, en el rango de edades entre 20 y 40 años.

Resultados. Los factores encontrados para las redes Azul-Amarillo y Rojo-Verde del experimento descrito son:

$$\eta_{YB} = \eta_2 = \eta_U = 0.14 \pm 0.06 \quad (21)$$

$$\eta_{RG} = \eta_3 = \eta_V = 0.02 \pm 0.07 \quad (22)$$

Puesto que los valores de η_i dependen de (1) la cromaticidad de los colores extremos $e_j^{(i)}$, (2) la frecuencia de la red, y (3) la naturaleza de la función base, los valores mostrados son válidos estrictamente sólo para las condiciones experimentales descritas. El cambio de funciones base (p.e. si se utilizan wavelets en vez de DCT por bloques) o del espacio de color oponente, requeriría realizar experimentos específicos siguiendo el procedimiento descrito. Tal como está actualmente implementado el invento, asumimos que las proporciones que hemos encontrado con las condiciones arriba descritas son aproximadamente válidas para diferentes colores promedio y funciones base.

Debe hacerse notar que el uso de modelos no lineales de la visión de color, en los que el canal acromático correspondiente describa mejor la percepción de luminosidad, podría aliviar la necesidad de este paso experimental para generar redes puramente cromáticas.

2.3. Experimento 2: Modulación máxima en imágenes naturales representadas en espacios oponentes

En esta sección, describimos el experimento numérico mediante el cuál se estimaron valores razonables de la modulación máxima en imágenes naturales, a lo largo de las direcciones del espacio YUV lineal. Estos valores $(\Delta T_{\max}^{(i)})_i$ son necesarios en la definición de contraste cromático dada en la ec. 13. Para ser consistentes con las restricciones experimentales que se aplican en la literatura psicofísica, cualquier elección particular de los valores de modulación máxima debe cumplir la condición siguiente: para una imagen natural típica, con color promedio T_0 , los colores extremos de las ecs. 12 deben estar dentro de la gama de colores generables por un display típico.

Experimento. Para encontrar estos valores, partimos de una estimación inicial arbitraria, p.e. $(\Delta T_{\max}^{(i)})_i = 256$, y comprobamos si la condición anterior se cumplía para $e_j^{(YB)}$ y $e_j^{(RG)}$ (calculados con ec. 12 y las proporciones experimentales 22) en un conjunto representativo de imágenes naturales en color (véase la sección siguiente para una descripción detallada de la base de datos). Si, para una imagen dada, alguno de los colores extremos se encontraba fuera de la gama de colores de un monitor CRT típico, se reducía en un 5% la modulación máxima correspondientes, $\Delta U_{\max}$ o $\Delta V_{\max}$. El proceso de búsqueda se detenía cuando todas las imágenes de la base de datos verificaban la condición.

Debido a las limitaciones de la base de datos (p.e. un rango limitado de luminancias promedio), del procedimiento descrito sólo resulta un conjunto de valores recomendados. Estos valores recomendados no garantizan un límite superior en el contraste cromático: podrían existir imágenes con contraste cromático superior a la unidad, particularmente en condiciones de baja luminancia y/o saturación elevada. Para tales imágenes, el uso de valores máximos de modulación más elevados podría dar lugar a valores de contraste más convenientes.

Base de datos. Realizamos el experimento numérico anterior sobre una base de datos consistente en 100 imágenes de la *McGill University color image database* [19] (fundamentalmente imágenes de flora y fauna) y 25 imágenes raw tomadas en nuestro laboratorio (fundamentalmente caras humanas). En el último caso, se utilizó una cámara Canon EOS 20 D para la captura de las imágenes calibrada con la librería COLORLAB y un espectroradiómetro PhotoResearch SpectraScan PR650.

Resultados. La Figura 3 muestra dos imágenes que limitan la máxima modulación en las direcciones V y U respectivamente. Los diagramas cromáticos CIE xy en la parte inferior de la figura muestran los colores de la imagen correspondiente (en gris), el color promedio (en negro), y los colores extremos (rojo y azul). El triángulo punteado en los diagramas cromáticos representa la gama de los colores disponibles en un monitor típico CRT (similar a NTSC o PAL RGB primarios). Como se puede ver en los diagramas cromáticos, la imagen de la izquierda limita la modulación en la dirección V (rojo-verde), mientras que la imagen de la derecha limita la modulación en la dirección U (amarillo-azul).

Las máximas modulaciones compatibles con las asunciones psicofísicas en la base de datos analizada fueron:

$$\Delta U_{\max} = 30 \pm 2 \quad (23)$$

$$\Delta V_{\max} = 43 \pm 2 \quad (24)$$

El error asociado, ± 2 , corresponde al paso del 5% en el procedimiento de búsqueda. Además, debido a las consideraciones anteriores sobre las limitaciones de la base de datos, estos valores son simplemente una recomendación que podría ser mejorada introduciendo algún tipo de dependencia con la luminancia media mediante una implementación más refinada.

2.4. Resumen: definición del contraste cromático y acromático

De acuerdo con el esquema general de la invención (véase Figura 1), el conjunto de transformaciones de la señal se lleva a cabo en los bloques 3, 4.X y 5.X, y se puede resumir así:

$$\mathbf{A}'(x) \xrightarrow{\text{Chrom.Transf.}} \mathbf{A}(x) \xrightarrow{\text{Space-Freq.Transf.}} \mathbf{a}(f) \xrightarrow{\text{Contrast.Transf.}} \mathbf{c}(f) \quad (25)$$

donde las componentes de los vectores, $\mathbf{A}'(x)$, son los valores triestímulo del color en el pixel x en el espacio de color inicial (no oponente) como por ejemplo NTSC RGB; los vectores, $\mathbf{A}(x)$, obtenidos a partir de la transformada cromática en el bloque 3, son los correspondientes colores en una representación oponente (como el espacio lineal YUV); los elementos de los vectores, $\mathbf{a}(f)$, son los coeficientes de las transformadas locales espacio-frecuenciales aplicadas a las imágenes $A_i(x)$, i.e., $a_1(f)$ representa los coeficientes de la transformada de la imagen acromática, $a_2(f)$ son los coeficientes de la transformada de la imagen Azul-Amarillo, y $a_3(f)$ son los coeficientes de la transformada de la imagen rojo-verde (obtenida usando los bloques 4.A, 4.B y 4.C, respectivamente). Finalmente, la transformada de contraste expresa cada amplitud, $a_i(f)$, en unidades de contraste cromática y acromática, dando los vectores $\mathbf{c}(f)$. Las componentes del vector de contrastes son: $c_1(f) = C_{\text{achrom}}(f)$, $c_2(f) = C_{\text{chrom}}^{(\text{YB})}(f)$, y $c_3(f) = C_{\text{chrom}}^{(\text{RG})}(f)$.

De acuerdo a la formulación y resultados experimentales presentados en esta sección, si la representación cromática es lineal YUV, y la representación espacial seleccionada es DCT a bloques con un tamaño de bloque $M \times N$, las transformaciones de contraste recomendadas son las listadas abajo:

- Para el canal de la luminancia (Y) (bloque 5.A):

$$C_{\text{achrom}}(p, o) = \sqrt{2} \frac{a_1(p, o)}{a_1(o, o)} \quad (26)$$

$$C_{achrom}(o, q) = \sqrt{2} \frac{a_1(o, q)}{a_1(o, o)} \quad (27)$$

$$C_{achrom}(p, q) = 2 \frac{a_1(p, q)}{a_1(o, o)} \quad (28)$$

• Para el canal azul-amarillo (U) (bloque 5.C):

$$C_{chrom}^{(2)}(p, o) = \frac{\sqrt{2}}{\sqrt{MN}} \frac{a_2(p, o)}{30} \quad (29)$$

$$C_{chrom}^{(2)}(o, q) = \frac{\sqrt{2}}{\sqrt{MN}} \frac{a_2(o, q)}{30} \quad (30)$$

$$C_{chrom}^{(2)}(p, q) = \frac{2}{\sqrt{MN}} \frac{a_2(p, q)}{30} \quad (31)$$

• Para el canal rojo-verde (V) (bloque 5.B):

$$C_{chrom}^{(3)}(p, o) = \frac{\sqrt{2}}{\sqrt{MN}} \frac{a_3(p, o)}{43} \quad (32)$$

$$C_{chrom}^{(3)}(o, q) = \frac{\sqrt{2}}{\sqrt{MN}} \frac{a_3(o, q)}{43} \quad (33)$$

$$C_{chrom}^{(3)}(p, q) = \frac{2}{\sqrt{MN}} \frac{a_3(p, q)}{43} \quad (34)$$

3. Transformaciones perceptuales no lineales

Esta sección se centra en la descripción de las transformaciones perceptuales no lineales aplicadas a las transformadas locales en frecuencia expresadas en unidades de contraste acromático y cromático.

3.1. Transformaciones de normalización divisiva (acromática, azul-amarillo y rojo-verde)

El segundo punto clave de la presente innovación se centra en en la aplicación de una transformación no lineal perceptual a los canales cromáticos y acromáticos. Los fundamentos de esta transformación vienen motivados por hechos y constataciones fundamentales tanto perceptuales (control de la ganancia perceptual y experimentos de enmascaramiento) como estadísticas (teoría de la codificación eficiente).

La transformación perceptual se modela en tres pasos: el primero consiste en un conjunto de filtros lineales aplicados a cada canal (este paso se corresponde con los bloques 4A, 4B y 4C en la Figura 1); a continuación, el contraste de cada coeficiente se obtiene usando el procedimiento descrito en la sección anterior (Bloques 5A, 5B y 5C en la figura 2); y, finalmente, se aplica una transformación no lineal a la salida de la segunda etapa (bloques 7A, 7B y 7C en la Figura 2). Esta última transformación puede ser una transformación no lineal punto-a-punto o una transformación no lineal preferiblemente [15]. En este último caso, la energía de cada coeficiente se normaliza por una combinación

de las energías de sus vecinos en frecuencia, y por tanto para cada canal la respuesta a una frecuencia particular se obtiene así:

$$r_i(f) = \frac{\text{sgn}(c_i(f)) \cdot |s_i(f) \cdot c_i(f)|^\gamma}{\theta(f) + \sum_{f'=1}^{N^2} h(f, f') |s_i(f') \cdot c_i(f')|^\gamma} \quad (35)$$

donde el subíndice $i = 1, 2, 3$ denota el canal; $c_i(f)$ son las salidas de los analizadores frecuenciales locales en unidades de contraste; $s_i(f)$ son las funciones tipo CSF; γ es un exponente; $\theta(f)$ es una función de regularización; u $h(f, f')$ determina la vecindad de interacción entre los coeficientes en la normalización no lineal de la energía.

$$h(f, f') \propto \exp \left(- \frac{\|f - f'\|^2}{\sigma_{|f|}^2} \right) \quad (36)$$

donde $\sigma_{|f|} = \frac{1}{6}|f| + 0.05$, $|f|$ viene dado en ciclos por grado (*cycles per degree*, cpd), y la Gaussiana se normaliza para tener volumen unidad.

El *kernel* de interacción anterior es sólo un ejemplo que incluye relaciones frecuenciales dentro de un mismo bloque pero podría generalizarse para que incluyera las interacciones espaciales y cromáticas.

Este modelo general de normalización divisiva está bien establecido para el caso acromático basado en umbrales incrementales de contrastes acromáticos a partir de sinusoides o bloques Gabor [15] y a partir de experimentos psicofísicos equivalentes. Este modelo acromático se ha empleado con éxito en diversas aplicaciones de procesamiento de imágenes, incluyendo compresión de imágenes acromáticas [11, 14] y restauración de imágenes acromáticas [13].

En el caso cromático, se ha obtenido un comportamiento no lineal similar tanto en el caso sinusoidal como Gabor [16, 21]. Esta es la razón por la que en la presente invención, los parámetros que controlan la no-linealidad en el caso cromático (γ, θ, h) tienen los mismos valores que en el caso acromático. Sin embargo, las sensibilidades frecuenciales totales se adaptan por canal cromático, basándose en trabajos previos con modelos lineales [17].

Dado que todos los resultados psicofísicos y psicológicos se basan en medidas usando redes sinusoidales o funciones de Gabor, los parámetros experimentales han de adaptarse al dominio particular de representación de la imagen (e.g. DCT, Wavelets, ICA, etc.). Si la representación de la imagen es en forma de block-DCT, recomendamos usar los valores particulares ilustrados en las figuras 4, 5 y 6 siguiendo así las recomendaciones de Dr. Uriegas [21], Dr. Watson [22] y Dr. Heeger [23].

La respuesta en el canal acromático se ha calculado bajo dos condiciones de entrada para así ilustrar cómo el modelo en la eq. 35 tiene en cuenta diferentes hechos perceptuales. Primero, la respuesta, $r_1(f)$, de un sensor sintonizado a una frecuencia f , se calcula cuando la entrada es $c_1 = [0, c_1(f), 0]$ (la Figura 7 muestra los resultados de dos sensores específicos: $r_1(f = 4)$ y $r_1(f = 10)$). Segundo, la respuesta de un sensor, $r_1(f)$, se ha obtenido en presencia de un patrón adicional con una frecuencia espacial diferente, es decir, cuando la entrada es $c_1 = [0, c_1(f), 0, c_1(f'), 0]$ (la Figura 9 muestra los resultado obtenidos por los sensores $r_1(f = 4)$ y $r_1(f = 10)$ cuando las entradas son $c_1 = [0, c_1(4), 0, c_1(6), 0]$ y $c_1 = [0, c_1(6), 0, c_1(10), 0]$ respectivamente). Los hechos perceptuales introducidos son:

- *Selectividad frecuencial*: la visibilidad de las distorsiones depende de la frecuencia espacial. Nótese que en la figura 7, la pendiente de la curva de la respuesta es mayor a 4 cpd que a 10 cpd, y por tanto, se necesita mayor distorsión a 10 cpd para obtener la misma distorsión percibida. En otras palabras, un ruido de 4 cpd es más visible que un ruido de la misma energía a 10 cpd. Este comportamiento general viene dado por la función pasa-banda $s_1(f)$ (véase la Figura 5).

- *Auto enmascaramiento*: la cantidad de distorsión necesaria para obtener una distorsión perceptual constante aumenta con el contraste de la entrada. Véase en la Figura 7 cómo Δc aumenta con el contraste del estímulo. Esto se debe al hecho de que la respuesta se atenúa cuando aumenta el contraste dado por el término de normalización en el denominador de Eq. (35).

- *Enmascaramiento cruzado*: La atenuación (y la correspondiente saturación de la respuesta y disminución de la sensibilidad) también ocurre cuando otros patrones $c_1(f')$ con $f' \neq f$ están presentes. Nótese en la Figura 8 que la cantidad de distorsión requerida aumenta con el contraste de la máscara a diferentes frecuencias aumenta. Es más, dada la forma Gaussiana de la vecindad de interacción, los patrones de frecuencias cercanas enmascaran la distorsión mas efectivamente que aquellos de frecuencia muy diferente. Así, la máscara a 6 cpd induce una mayor variación de ruido aceptable en 4 cpd que en 10 cpd.

Un comportamiento similar se produce en los canales cromáticos.

3.2 Experimento 3: inversión de la transformada no lineal perceptual Condición de invertibilidad

Las transformadas no lineales deben ser invertibles para poder reconstruir la imagen en el decodificador a partir de los datos transmitidos o almacenados. En [11] se propuso un procedimiento de inversión basado en una forma explícita, y se estudió la invertibilidad para el caso acromático. Este procedimiento requiere la inversión de la matriz $(I - D_{ri} \cdot h)$, donde I es la matriz identidad, D_{ri} es la matriz diagonal conteniendo los valores absolutos de los elementos de $r_i(f)$, y h es la matriz que modela la relación entre coeficientes. La condición de inversión asegura que todos los autovalores de $D_{ri} \cdot h$ deben ser menores que uno. No existe una garantía teórica de que esta condición se cumpla para todas las imágenes, tanto en el caso acromático propuesto como en el cromático presentado aquí. Por lo tanto, la invertibilidad de las transformadas no lineales cromáticas (Bloques 7.B y 7.C) sólo se pueden comprobar empíricamente.

Comprobación de la condición de invertibilidad en los casos cromáticos. La condición de invertibilidad fue empíricamente comprobada calculando el mayor autovalor de las matrices $D_{r2} \cdot h$ y $D_{r3} \cdot h$ para 25600 bloques de 16x16 de un conjunto de imágenes en color [19]. La condición de invertibilidad también fue comprobada para la señal reconstruida $\hat{r}_i$ a partir de los pesos cuantizados por SVM a distintas tasas de compresión en el rango [0.2, 2.2] bits/pix. En estos experimentos, el mayor autovalor para cada bloque fue, en todos los casos, menor que 1.

4. Support Vector Regression (SVR) con perfil adaptativo

El tercer bloque del esquema de codificación presentado es la inclusión de un método de aprendizaje máquina ("machine learning") de forma que éste sea el responsable de la selección de los coeficientes más relevante en el dominio concreto de representación. El método debe considerar las características particulares del dominio, y por tanto la distinta importancia perceptual de cada coeficiente en el dominio. En este apartado ilustramos el bloque de aprendizaje-máquina mediante un método específico, la máquina de vectores soporte para regresión ("support vector regression, SVR), que ha demostrado buenos resultados en un gran número de problemas de regresión y aproximación de funciones, empleando para ello pocos parámetros (muestras o vectores soporte) [24]. En la siguiente sección revisamos la SVR estándar. Dado que esta aproximación define un parámetro de insensibilidad constante e independiente de la relevancia perceptual de cada muestra, a continuación presentamos una nueva formulación de la SVR para trabajar en dominios no-Euclideos de representación. Finalmente, resumimos algunas puntualizaciones a este importante característica a tener en cuenta.

4.1. SVR estándar

En el esquema de compresión propuesto, se deben aplicar tres modelos SVR independientemente sobre los coeficientes de cada representación no lineal $r(f) = [r_1(f), r_2(f), r_3(f)]^T$. A continuación, y por claridad, presentamos la notación de la SVR sólo para el caso genérico de una representación no lineal de un canal en concreto, r . En esencia, para el desarrollo del modelo, tenemos un conjunto de parejas de posiciones y coeficientes (f, r) , donde $f = 1, \dots, N$ representa el índice spacio-frecuencial de los coeficientes $r \in \mathbb{R}$ en el canal transformado perceptual no lineal correspondiente ($r_1(f)$, $r_2(f)$ o $r_3(f)$).

La formulación estándar de la SVR mapea los datos de entrada f a un espacio de mayor dimensionalidad donde se aplica una regresión lineal, que está relacionada no linealmente con el espacio de representación de entrada. Por lo tanto, se puede definir el modelo de regresión como:

$$\hat{r}_f = \mathcal{F}(f) = \langle \mathbf{w}, \phi(f) \rangle + b \quad (37)$$

donde $\hat{r}_f$ son las estimaciones de los coeficientes r , $(\cdot, \cdot)$ representa la operación de producto escalar, ϕ es un mapeo no lineal a un espacio de Hilbert de mayor dimensión $\phi: \mathbb{R} \rightarrow \mathcal{H}$; $\mathbf{w}$ es un vector de pesos en el espacio de características de mayor dimensión; y b es un término de sesgo del modelo.

La SVR consiste en resolver el siguiente problema de regularización con restricciones lineales:

$$\min_{\mathbf{w}, \xi_f, \xi_f^*, b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + \lambda \sum_{f=1}^N (\xi_f + \xi_f^*) \right\} \quad (38)$$

sujeto a:

$$r_f - (\langle \mathbf{w}, \phi(f) \rangle + b) \leq \varepsilon + \xi_f \quad \forall f = 1, \dots, N \quad (39)$$

$$(\langle \mathbf{w}, \phi(f) \rangle + b) - r_f \leq \varepsilon + \xi_f^* \quad \forall f = 1, \dots, N \quad (40)$$

$$\xi_f, \xi_f^* \geq 0 \quad \forall f = 1, \dots, N \quad (41)$$

El parámetro libre λ se debe ajustar y controla la importancia del ajuste a los datos (minimización de los errores ε_f y ε_f^*) frente a mantener un modelo con pesos pequeños $\|\mathbf{w}\|$ (forzar suavidad en el espacio de características). El método emplea la denominada función de coste ε -insensitiva [24], que penaliza errores mayores que ε de forma lineal. Nótese que el parámetro ε representa el error o distorsión permitida.

El procedimiento habitual para resolver este problema de optimización consiste en introducir las restricciones lineales (39)-(41) en la Eq. (38) mediante multiplicadores de Lagrange asociados $\alpha_f^{(*)}$, después se calculan las condiciones de Karush-Kuhn-Tucker, con las que resuelve el problema dual de Wolfe [25]. Esto lleva a tener que resolver un problema de programación cuadrática (QP) en un espacio de parámetros o pesos duales $\alpha_f^{(*)}$ en lugar de tener que hacerlo en el espacio de los parámetros del modelo $\mathbf{w}$, de los que se desconoce la dimensión. Un resultado muy importante de esta optimización resulta en que, al hacer nula la derivada del funcional dual, se obtiene una relación del vector de pesos en el espacio de características expresado como combinación lineal de las muestras mapeadas, esto es:

$$\mathbf{w} = \sum_{f=1}^N (\alpha_f - \alpha_f^*) \phi(f). \quad (42)$$

Ahora, introduciendo (42) en (37), se puede obtener la solución para una entrada particular f' de esta forma:

$$\hat{r}_{f'} = \mathcal{F}(f') = \sum_{f=1}^N (\alpha_f - \alpha_f^*) \langle \phi(f), \phi(f') \rangle + b, \quad (43)$$

que depende explícitamente del producto escalar de las muestras mapeadas y de los pesos duales obtenidos α_f y $\alpha_f^{(*)}$. Este resultado nos permite trabajar en un espacio de dimensión superior sin tener que conocer las coordenadas de las muestras en este espacio, sólo los productos escalares entre ellas. Estos productos se denominan funciones *kernel*, $K(f, f')$, y definen la función de regresión:

$$\hat{r}_{f'} = \mathcal{F}(f') = \sum_{f=1}^N (\alpha_f - \alpha_f^*) K(f, f') + b, \quad (44)$$

donde el productos interior (escalar) $\langle \phi(f), \phi(f') \rangle$ viene representado por una matriz *kernel* $K(f, f')$. Nótese que sólo las muestras con multiplicadores de Lagrange $\alpha_f^{(*)}$ distintos de cero intervienen en la solución y se llaman vectores soporte ("support vectors", SVs). La ventaja inmediata del método es que se pueden aproximar funciones de forma muy flexible con relativamente pocos parámetros, llevando este concepto de *sparsity* a la idea inherente de compresión.

La matriz de Gram (o kernel) $K(f, f') = \langle \phi(f), \phi(f') \rangle$ se puede ver como una matriz de similitud entre las muestras, y una definición apropiada es la clave para el correcto funcionamiento de un método SVR. En todos nuestros experimentos empleado una función de base radial (RBF):

$$K(f, f') = \exp \left(- \frac{\|f - f'\|^2}{2\sigma^2} \right) \quad (45)$$

Este método introduce un tercer parámetro libre a ser optimizado, la anchura del kernel o longitud de escala σ .

4.2. SVR con insensibilidad adaptativa

El principal problema cuando se considera la solución anterior es que se asume una relevancia igual para cada muestra, lo cual no es cierto en general. Esto se puede resolver empleando un factor de penalización diferente para cada una las muestras f de acuerdo a una determinada función de confianza k_f . Esta idea se puede también extender al

diseño de una sensibilidad ε por muestra. La SVR con insensibilidad adaptativa [26] aumenta o disminuye la región de insensibilidad ε en función de la muestra de entrenamiento. Así, la función a minimizar es:

$$\min_{\mathbf{w}, \xi_f, \xi_f^*, b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + \lambda \sum_{f=1}^N k_f (\xi_f + \xi_f^*) \right\} \quad (46)$$

y las restricciones sobre las variables auxiliares dependen ahora de la muestra en particular:

$$r_f - (\langle \mathbf{w}, \phi(f) \rangle + b) \leq \frac{\varepsilon}{k_f} + \xi_f \quad \forall f = 1, \dots, N \quad (47)$$

$$(\langle \mathbf{w}, \phi(f) \rangle + b) - r_f \leq \frac{\varepsilon}{k_f} + \xi_f^* \quad \forall f = 1, \dots, N \quad (48)$$

$$\xi_f, \xi_f^* \geq 0 \quad \forall f = 1, \dots, N \quad (49)$$

Por tanto, ahora cada muestra tiene una sensibilidad propia al error $\varepsilon_f = \varepsilon / k_f$, que significa intuitivamente que se permite un error diferente a cada una de ellas en función de su relevancia (en nuestro caso, perceptual). Incluyendo las restricciones (47)-(49) en el correspondiente funcional (46), se puede desarrollar la solución de la misma manera que en el caso estándar, ya que de nuevo se obtiene un problema QP análogo.

4.3 Puntualizaciones: insensibilidad adaptativa y el dominio de trabajo

En los esquemas de codificación basados en SVM como los presentados en [5, 12, 14], la señal viene descrita por los multiplicadores de Lagrange asociados a los vectores soporte necesarios para mantener el error de regresión bajo los umbrales ε_f . Aumentar estos umbrales reduce el número de vectores soporte, reduciendo así la entropía de la imagen codificada pero aumentando la distorsión. El punto clave es por tanto la elección de valores adecuados para ε_f de acuerdo con algún criterio apropiado al dominio de trabajo. Por ejemplo, en un dominio Euclideo, se puede trabajar con valores constantes de ε_f para todas las f . Esta condición se cumple en la transformada perceptual no lineal incluida en el esquema de la presente invención, pero no en dominios lineales DCT o wavelet como los empleados en [12], donde cada coeficiente tiene una relevancia diferente en la representación de la señal. Véase [5, 13, 14] para una discusión más detallada.

5. Resultados experimentales para evaluar el comportamiento del método

Esta sección ilustra el comportamiento de una implementación del método para la compresión de un conjunto de imágenes en color en comparación con una implementación del estándar JPEG.

5.1. Algoritmos, conjunto de comparación y experimentos

La implementación del método propuesto, identificado por las siglas C-NL-SVR (Color-Non-Linear-SVR), incluye la representación del color en el espacio lineal YUV, la representación de la información especial en el dominio DCT-local con bloques 16X16, transformaciones de contraste como las referidas en la sección 2, transformadas perceptuales no-lineales de cada canal cromático con los parámetros dados en la sección 3, y un proceso de aprendizaje de esta última representación no-lineal mediante SVR tal como se describe en la sección 4. Finalmente, la descripción seleccionada por el algoritmo de regresión (los pesos) son uniformemente cuantizados y codificados de acuerdo con el esquema descrito en la sección 1.

La implementación de JPEG empleada aquí utiliza la misma representación del color y la misma DCT local con bloques 16X16. En el caso JPEG se usaron matrices de cuantización basadas en las CSFs acromática y cromáticas de Mullen [17].

Los experimentos se efectuaron sobre un conjunto de imágenes en color formado por 25 imágenes naturales representativas (véase la figura 9). Todos las imágenes de este conjunto se comprimieron a diferentes tasas en el rango [0.1, 2.2] bits/pix. Debe resaltarse que el volumen usual en electrónica de consumo (teléfonos móviles, cámaras digitales de gama baja, cámaras semi-profesionales y cámaras digitales profesionales de gama alta) oscila en el rango entre 1 bit/pix para ficheros JPEG de baja calidad y 2.2 bits/pix para ficheros JPEG de alta calidad.

En todos los experimentos, se entrenaron tres modelos SVR (uno para cada canal cromático y bloque transformado), y modelizamos el valor (con signo) de los coeficientes de la respuesta, $r_i(f)$, $i = 1, 2, 3$. En nuestro dominio de representación no-lineal, r , utilizamos SVRs con kernel RBF, parámetro de insensibilidad ε constante, y fijamos un valor de λ arbitrariamente alto. En todos los casos, ε y el tamaño de los pasos de cuantización fueron conjuntamente modificados para satisfacer las especificaciones de distorsión y/o entropía deseada (véase la Figura 1).

Después del entrenamiento, la señal se describe por los multiplicadores de Lagrange de los vectores soporte necesarios para mantener el error por debajo de los umbrales ε , siendo estos multiplicadores cuantizados uniformemente. El último paso es la codificación entrópica de los pesos cuantizados. La tasa de compresión (o la distorsión introducida) está esencialmente controlada por un factor aplicado a los umbrales ε .

5.2. Comparación numérica entre C-NL-SVR y JPEG: curvas tasa-distorsión y ganancia de compresión

Con objeto de evaluar la calidad de las imágenes codificadas a la misma tasa de compresión, se utilizaron tres medidas de distorsión: el error cuadrático medio estándar (RMSE), y dos medidas perceptualmente significativas; el Structural SIMilarity (SSIM) index [27] y la métrica S-CIELab [28].

Los promedios de las curvas de tasa-distorsión para todas las imágenes del conjunto de entrenamiento de la figura 9 se muestran en la figura 10. En las gráficas de tasa-distorsión las barras indican la desviación estándar del error para cada tasa de compresión. Estas gráficas muestran que el C-NL-SVR propuesto es mayor que JPEG (la línea continua está por debajo de la línea discontinua para las distorsiones y supera a la otra en el caso de la similaridad en el rango de entropías de interés comercial). Sin embargo, una interpretación naïf de la superposición de las barras de desviación estándar puede llevar a cuestionar la significancia de esta ganancia eventual.

Nótese sin embargo, que en las gráficas de tasa-distorsión, la superposición entre estas barras no necesariamente significa que el comportamiento sea igual o que la ganancia sea estadísticamente despreciable: en este caso, el solapamiento proviene del hecho de que diferentes imágenes en la base de datos tienen una complejidad intrínsecamente diferente dando lugar a distorsiones muy diferentes cuando son codificadas a una cierta tasa. La Figura 11(a) muestra un ejemplo de lo anterior: en este caso, se muestra el comportamiento tasa-distorsión de los dos métodos (líneas continua y discontinua) para dos imágenes diferentes de la base de datos (líneas negra y azul). Naturalmente, si se promedia sobre tal clase de imágenes la desviación estándar va a ser muy grande, pero el hecho relevante es que la ganancia del método propuesto es consistente en cada imagen: nótese por ejemplo que, tomando algún valor de entropía particular para el resultado JPEG, en ambas imágenes (por ejemplo los puntos rojo o púrpura en la Figura 11(a)), la entropía del resultado C-NL-SVR con la misma distorsión es consistentemente mucho menor.

Por tanto, tiene más sentido definir algún tipo de medida de ganancia de compresión para cada imagen y tasa, y promediar estas ganancias para el conjunto de imágenes en la base de datos.

Aquí, definimos la *ganancia de compresión* (*compression gain*) de un método sobre otro de referencia para una entropía dada, $G(H)$, en términos del cociente de las tasas para el mismo nivel de distorsión:

$$\text{Compression Gain : } G(H) = \frac{H_{\text{JPEG}}(D(H))}{H_{\text{C-NL-SVR}}(D(H))} \quad (50)$$

Fig. 11(b) muestra un esquema ilustrativo para determinar la *ganancia de compresión* para una medida de distorsión sobre una imagen particular (zoom de la curva tasa-distorsión RMSE para las líneas negras del gráfico de la izquierda). En el ejemplo que se ilustra en la Figura 12 (b) consideramos los dos casos posibles (ganancia inferior a 1 a baja entropía, y ganancia superior a 1 a alta entropía). Nótese que las ganancias superiores a 1 implican que el método propuesto utiliza menos bits que el método de referencia para representar la imagen a la misma distorsión. Esta ganancia de compresión también puede expresarse en términos porcentuales mediante:

$$\text{Compression Gain(en \%): } PG(H) = 100 \cdot (G(H) - 1) \quad (51)$$

De forma práctica, para una entropía dada, $G = 2$ o $PG = 100\%$ significa que 20 imágenes C-NL-SVR ocupan el mismo volumen (en bits) que 10 imágenes JPEG (para la misma calidad de reproducción).

Por supuesto, los valores particulares de ganancia de compresión dependerán de la medida de distorsión utilizada y del conjunto de imágenes de calibración.

La Figura 12 muestra las ganancias de compresión obtenidas por el método C-NL-SVR sobre JPEG par las tres medidas de distorsión consideradas (RMSE, SSIM y S-CIELab) para cada entropía considerada en el rango de tasas analizado. En estas gráficas hemos incluido la desviación estándar de las ganancias para valorar la fiabilidad de la aproximación.

Nótese que en el rango comercialmente significativo [1, 2.2] bits/pix, el método propuesto sobrepasa ampliamente a JPEG y permite mejoras en la compresión del orden de [60, 180]% para RMSE, [25, 60]% para SSIM, y de [80, 275]% para S-CIELAB. Como las barras de desviación estándar no cruzan por debajo de la línea $G = 1$, la ganancia efectiva (superior a 1) es consistente sobre un amplio rango de imágenes, por tanto, el método propuesto supera claramente a JPEG.

5.3. Comparación visual

Las Figuras 13-17 muestran resultados representativos de los métodos considerados sobre 5 imágenes (“Parrot”, “Lena”, “Roof”, “Flower3”, “Face1”) a diferentes tasas en el rango [1.0, 2.2] bits/pix. Los resultados visuales confirman que las ganancias *numéricas* mostradas en la figura 13 también redundan en una ganancia *perceptual*. En general, JPEG da lugar a peores resultados (con mas efecto de bloques). Además, merece la pena señalar que los detalles de alta frecuencia se suavizan más en JPEG que el método propuesto. Estos efectos negativos se reducen mediante la introducción de técnicas de aprendizaje de máquina como las SVRs en el dominio perceptual no-lineal. Véanse por ejemplo los ojos y las mejillas de Lena (Fig.14), las plumas del sombrero (Fig.14), la mejor reproducción del patrón de alta frecuencia en el ojo del loro (Fig.13), el listón vertical bajo el tejado la casa (Fig.15), o el patente efecto de bloques en la flor (Fig.16) y en la cara de la figura 17.

6. Implementación del método

Un posible aparato para implementar el método anteriormente descrito es un sistema de computación que comprende:

- Un interfaz de entrada y salida para recibir datos de imágenes en color que puede ser un simple canal de comunicación o múltiples canales de comunicación.

- Un procesador comunicado con el interfaz de entrada/salida para realizar los pasos del método de acuerdo con código almacenado en sistemas de almacenamiento adecuadamente conectados con el procesador.

- Sistemas de visualización (monitores) y tarjetas gráficas para la realización (off-line) de los experimentos psicofísicos anteriormente descritos.

- Sistemas de memoria para alojar la base de datos utilizada (off-line) en los experimentos anteriormente descritos.

Se pueden introducir modificaciones en el aparato que se acaba de describir, que estarían en todo caso incluidas en el ámbito recogido por las siguientes reivindicaciones.

Bibliografía

[1] G.K. **Wallace**. The JPEG still picture compression standard. *Communications of the ACM*, 34(4):31-43, 1991.

[2] M. **Naillon** and J-B. **Theeten**. Method of and arrangement for image data compression by means of a neural network. US Patent No. 5,005,206, 1991.

[3] R. A. **Nanni** and G. **Abraham**. Neural-network-based method of image compression. US Patent No. 6798914, 2004.

[4] V. **Kecman** and J. **Robinson**. Method, apparatus and software for lossy data compression and function approximation. Patent Appl. No. WO 03/050959, 2003.

[5] G. **Gómez**, G. **Camps-Valls**, J. **Gutiérrez**, and J. **Malo**. Perceptual adaptive insensitivity for support vector machine image coding. *IEEE Transactions on Neural Networks*, 16(6):1574-1581, Jun 2005.

[6] A. B. **Watson**. Image data compression having minimum perceptual error. US Patent No. 5,426,512, 1995.

[7] J. **Malo**, J. **Gutiérrez**, I. **Epifanio**, F. **Ferri**, and J. M. **Artigas**. Perceptual feed-back in multigrid motion estimation using an improved DCT quantization. *IEEE Transactions on Image Processing*, 10(10):1411-1427, October 2001.

[8] J. **Malo**. Tecnología del color, chapter Almacenamiento y transmisión de imágenes en color, pages 117-164. Col.lecció Materials. *Servei de Publicacions de la Universitat de Valencia*, Valencia, 2002.

[9] W. **Zeng**, S. **Daly**, and S. **Lei**. An overview of the visual optimization tools in JPEG2000. *Signal Processing: Image Communication*, 17(1):85-104, 2002.

[10] Y. **Navarro**, J. **Rovira**, J. **Gutiérrez**, and J. **Malo**. Gain control for the chromatic channels in JPEG2000. *Proc. of the 10th Intl. Conf. AIC.*, 1:539-542, 2005.

[11] J. **Malo**, I. **Epifanio**, R. **Navarro**, and R. **Simoncelli**. Non-linear image representation for efficient perceptual coding. *IEEE Transactions on Image Processing*, 15(1):68-80, 2006.

[12] J. **Robinson** and V. **Kecman**. Combining Support Vector Machine learning with the discrete cosine transform in image compression. *IEEE Transactions on Neural Networks*, 14(4):950-958, July 2003.

[13] J. **Gutiérrez**, G. **Gómez-Pérez**, J. **Malo**, and G. **Camps-Valls**. Perceptual image representations for support vector machine image coding. In G. Camps-Valls, J. L. Rojo-Alvarez, and M. Martínez-Ramón, editors, *Kernel Methods in Bioengineering, Signal and Image Processing*. Idea Group Publishing, Hershey, PA (USA), Jan 2007.

[14] G. **Camps-Valls**, J. **Gutiérrez**, G. **Gómez**, and J. **Malo**. On the suitable domain for SVM training in image coding. *Journal of Machine Learning Research*, 9(1):49-66, 2008.

[15] A. B. **Watson** and J. A. **Solomon**. A model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America A*, 14(9):2379-2391, September 1997.

[16] E. **Martínez-Uriegas**. Color detection and color contrast discrimination thresholds. In *Proceedings of the OSA Annual Meeting ILS-XIII*, page 81, Los Angeles, 1997.

[17] K. T. **Mullen**. The contrast sensitivity of human colour vision to red- green and yellow-blue chromatic gratings. *Journal of Physiology*, 359:381-400, 1985.

[18] J. S. **Lim**. Two-dimensional signal and image processing. *Prentice-Hall, Inc., Upper Saddle River, NJ, USA*, 1990.

[19] A. **Parraga**. McGill Calibrated Colour Image Database. *Dept. of Vision Science, McGill University*, 2003. Available at <http://tabby.vision.mcgill.ca>.

[20] T. T. **Norton**, D. A. **Corliss**, and J. E. **Bailey**. The Psychophysical Measurement of Visual Function. *Butterworth Heinemann*, 2002.

[21] E. M. **Uriegas**. Personal communication on contrast incremental thresholds for chromatic gratings measured at Stanford research international. *Unpublished results*, May 1998.

[22] A. B. **Watson**. Personal communication on plausible parameters for divisive normalization in DCT basis. *Unpublished results*, March 2001.

[23] D. **Heeger**. Personal communication on plausible parameters for divisive normalization in DCT basis. *Unpublished results*, March 2001.

[24] A. J. **Smola** and B. **Schölkopf**. A tutorial on support vector regression. *Statistics and Computing*, 14:199-222, 2004.

[25] B. **Schölkopf** and A. **Smola**. Learning with Kernels - Support Vector Machines, Regularization, Optimization and Beyond. *MIT Press Series*, 2002.

[26] G. **Camps-Valls**, E. **Soria-Olivas**, J. **Pérez-Ruixo**, A. **Artés-Rodríguez**, F. **Pérez-Cruz**, and A. **Figueiras-Vidal**. A profile-dependent kernel-based regression for cyclosporine concentration prediction. In *Neural Information Processing Systems (NIPS) - Workshop on New Directions in Kernel-Based Learning Methods*, Vancouver, Canada, December 2001. No published proceedings. Available at <http://www.uv.es/~gcamps>.

[27] Z. **Wang**, A. C. **Bovik**, H. R. **Sheikh**, and E. P. **Simoncelli**. *Image quality assessment: From error visibility to structural similarity*. *IEEE Transactions on Image Processing*, 13(4):600-612, 2004.

[28] X. **Zhang** and B. **Wandell**. Color image fidelity metrics evaluated using image distortion maps. *Signal Processing*, 70(3):201-214, 1998.

REIVINDICACIONES

1. Un método de compresión de imágenes en color que comprende los siguientes pasos:

a) recepción de la imagen en color de entrada expresada en canales RGB;

b) transformación de los canales RGB a un espacio oponente de representación del color para obtener un canal acromático (A), un canal rojo-verde (RG), y un canal azul-amarillo (YB);

c) transformación de dichos canales A, RG e YB a un dominio espacio-frecuencial local, como suma de un conjunto de funciones base multiplicadas por el conjunto correspondiente de coeficientes y la codificación de los residuos de baja frecuencia;

d) expresión de los coeficientes de dichas representaciones espacio- frecuencias locales de los canales acromáticos y cromáticos oponentes en unidades de contraste perceptualmente significativas;

e) aplicación de transformaciones no-lineales de tipo normalización divisiva con los parámetros ajustados al uso de contrastes en la expresión de las representaciones de partida, a cada uno de los canales A, RG, YB;

f) estimación de los coeficientes de dichas representaciones no lineales perceptuales utilizando un procedimiento de aprendizaje de máquina;

g) realización de una cuantización de la descripción obtenida, seguida de la codificación entrópica de los resultados posteriores a la cuantización.

2. Un método de compresión de imágenes en color de acuerdo con la reivindicación 1, **caracterizado** por el hecho de que la transformada espacio frecuencial local del paso c) se selecciona entre: la DCT o cualquier tipo de transformada basada en funciones de Gabor o wavelets.

3. Un método de compresión de imágenes en color de acuerdo con cualquiera de las reivindicaciones 1-2, **caracterizado** por el hecho de que en el paso d) el contraste del canal acromático se calcula como función de un número de parámetros relevantes obtenidos mediante experimentos psicofísicos y numéricos y los contrastes de los canales cromáticos se calculan a partir de un número de parámetros relevantes obtenidos a partir de experimentos psicofísicos y numéricos.

4. Un método de compresión de imágenes en color de acuerdo con cualquiera de las reivindicaciones 1-3, **caracterizado** por el hecho de que los parámetros de la transformación de contraste obtenidos por medios psicofísicos incluyen las proporciones de la luminancia media de las componentes de las funciones base cromáticas y los parámetros obtenidos por medio de experimentos numéricos incluyen valores razonables de la máxima modulación de color existente en una base de datos representativa de imágenes en color.

5. Un método de compresión de imágenes en color de acuerdo con cualquiera de las reivindicaciones 1-4, **caracterizado** por el hecho de que en el paso e) la transformación perceptual no-lineal aplicada al canal acromático es una normalización divisiva con un kernel de interacción totalmente general, y las transformaciones perceptuales no-lineales aplicadas a los canales cromáticos son normalizaciones divisivas con los parámetros de sensibilidad frecuencial específicamente adaptados y escalados respecto del canal acromático.

6. Un método de compresión de imágenes en color de acuerdo con cualquiera de las reivindicaciones 1-5, **caracterizado** por el hecho de que en el paso f) el proceso de aprendizaje de máquina se entrena para identificar una secuencia de funciones base y sus pesos correspondientes, guiado con criterios adecuados para optimizar la compacidad de la secuencia de datos codificados resultante y/o la calidad de la imagen codificada.

7. Un método de compresión de imágenes en color de acuerdo con cualquiera de las reivindicaciones 1-6, **caracterizado** por el hecho de que en el paso f) el procedimiento de aprendizaje de máquina es un procedimiento SVR.

8. Un método de compresión de imágenes en color de acuerdo con cualquiera de las reivindicaciones 1-7, **caracterizado** por el hecho de que en el paso f) el procedimiento de aprendizaje de máquina es un procedimiento implementado mediante alguno de los siguientes métodos: redes neuronales, técnicas basadas en splines, métodos kernel, redes Bayesianas, procesos Gaussianos o lógica difusa.

9. Un método de compresión de imágenes en color de acuerdo con cualquiera de las reivindicaciones 1-8, **caracterizado** por el hecho de que en el paso g) el sub paso de codificación de los resultados se realiza mediante la codificación entrópica del flujo de datos.

10. Un método de compresión de imágenes en color de acuerdo con cualquiera de las reivindicaciones 1-9, **caracterizado** por el hecho de que en los pasos e) y g) se incluyen sub pasos adicionales para el procesamiento de la información de fase de forma separada de la información de amplitud.

ES 2 310 157 B1

11. Sistema de almacenamiento de datos que contenga un programa tal que, cuando se ejecute en un procesador, haga que éste realice el método descrito en alguna de las reivindicaciones 1-10.

12. Aparato para producir datos comprimidos de imágenes en color que comprende métodos de recepción de las imágenes en color de entrada y métodos de procesamiento para:

a) transformar dichos datos de entrada a un espacio de color oponente consistente en un canal acromático, A, y dos canales cromáticos, rojo-verde, RG, y azul-amarillo YB;

b) transformar dichos canales A, RG e YB a dominios espacio-frecuencia locales como una suma de funciones base multiplicadas por el conjunto correspondiente de coeficientes y la codificación de los residuos pasa baja;

c) expresar los coeficientes de dichas representaciones espacio-frecuencia local de A, RG e YB, en unidades de contraste perceptualmente significativas;

d) aplicar una transformación no-lineal de tipo normalización divisiva a los canales de contraste A, RG e YB;

e) estimar los coeficientes de dicha representación perceptual no lineal de la imagen utilizando un procedimiento de aprendizaje de máquina;

f) realizar una cuantización de la descripción obtenida y la codificación de los resultados.

13. Aparato para producir datos comprimidos de imágenes en color de acuerdo con la reivindicación 12 **caracterizado** por el hecho de que también comprende sistemas de visualización y tarjetas gráficas para la realización (off-line) de experimentos psicofísicos para la obtención de las proporciones de luminancia media de las componentes de las funciones base cromáticas.

14. Aparato para producir datos comprimidos de imágenes en color de acuerdo con cualquiera de las reivindicaciones 12-13 **caracterizado** por el hecho de que también comprende una base de datos de imágenes en color para realizar experimentos numéricos para estimar los valores máximos de la modulación del color en cada dirección cromática cardinal.

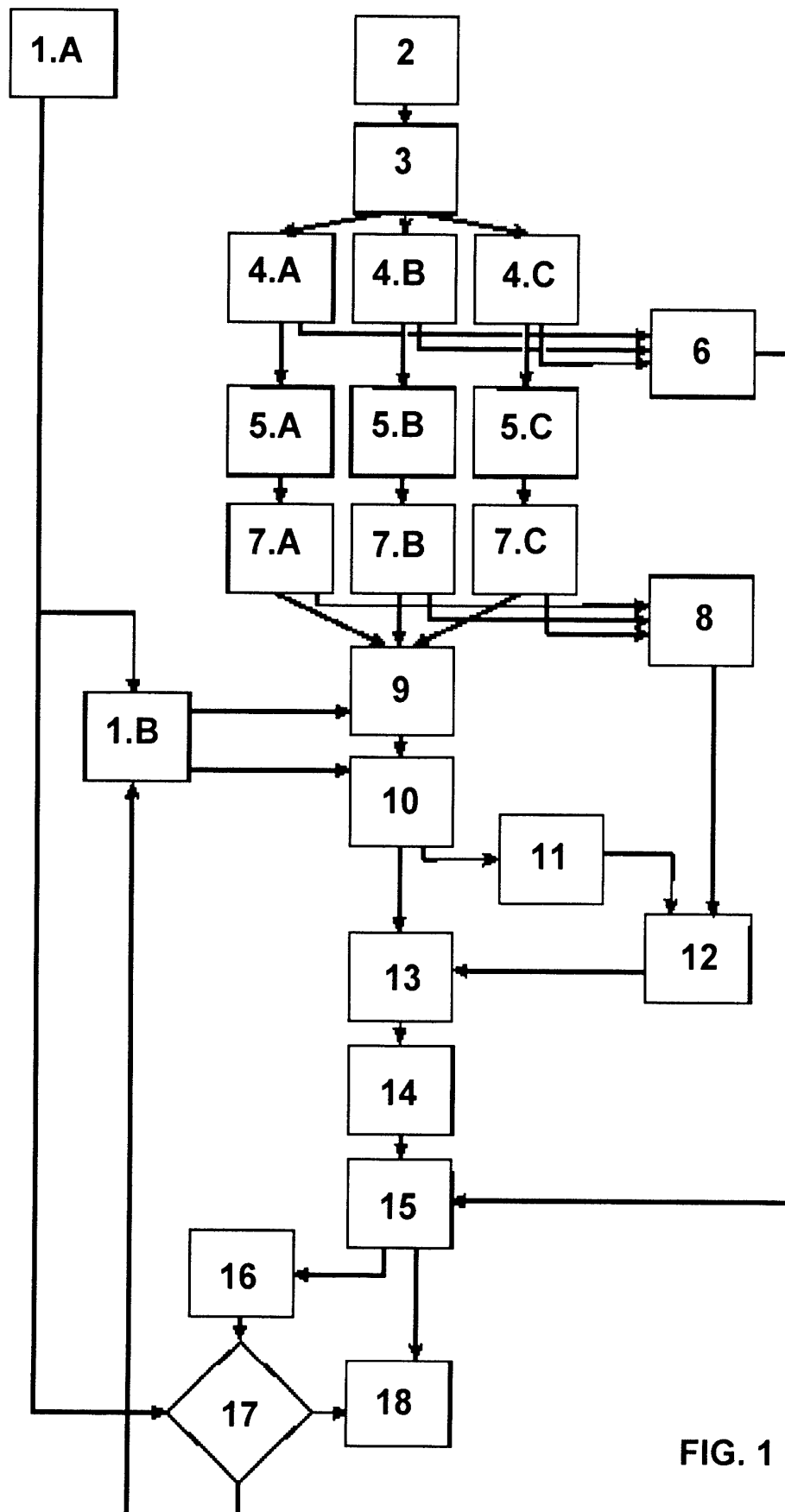


FIG. 1

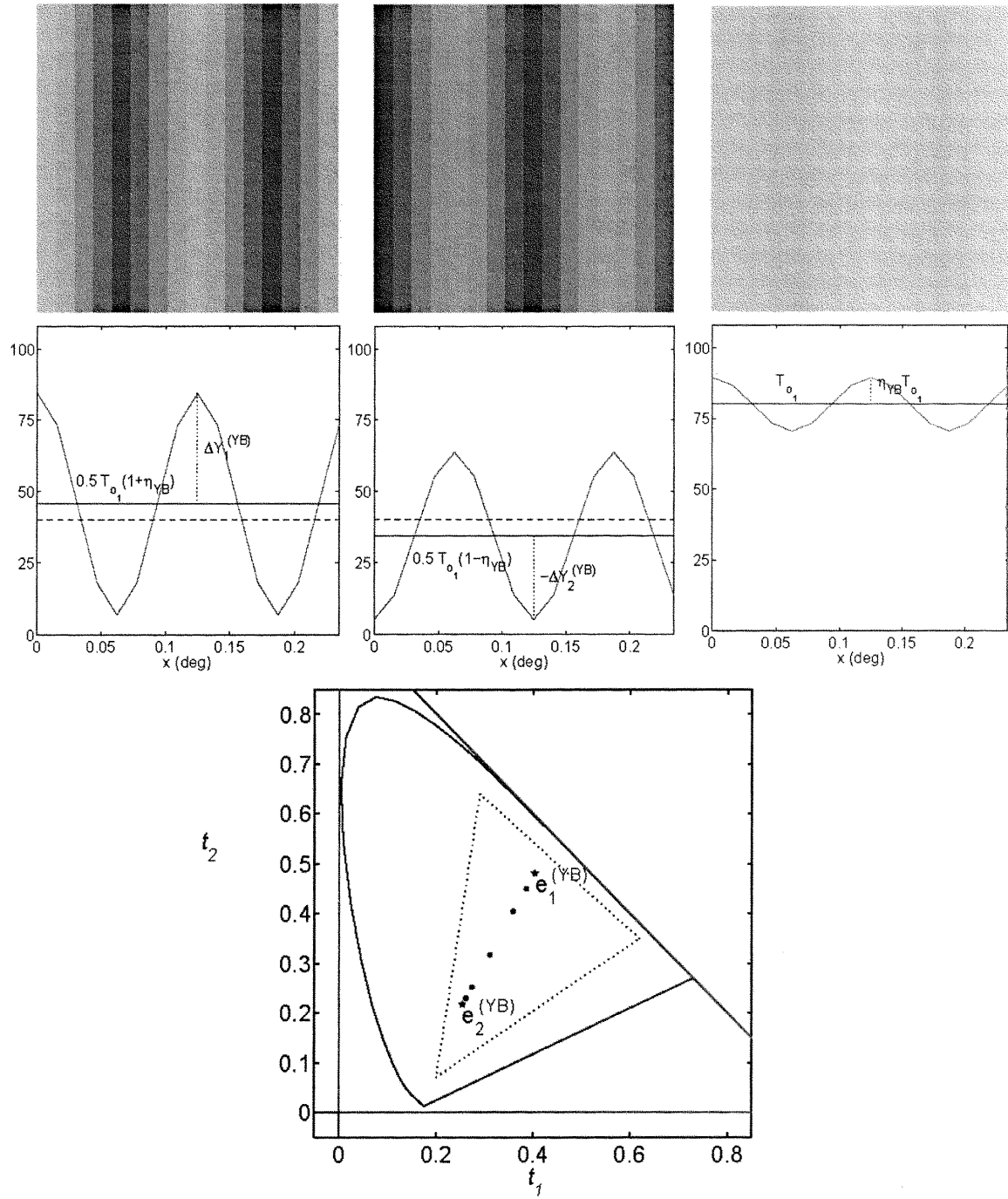


FIG. 2

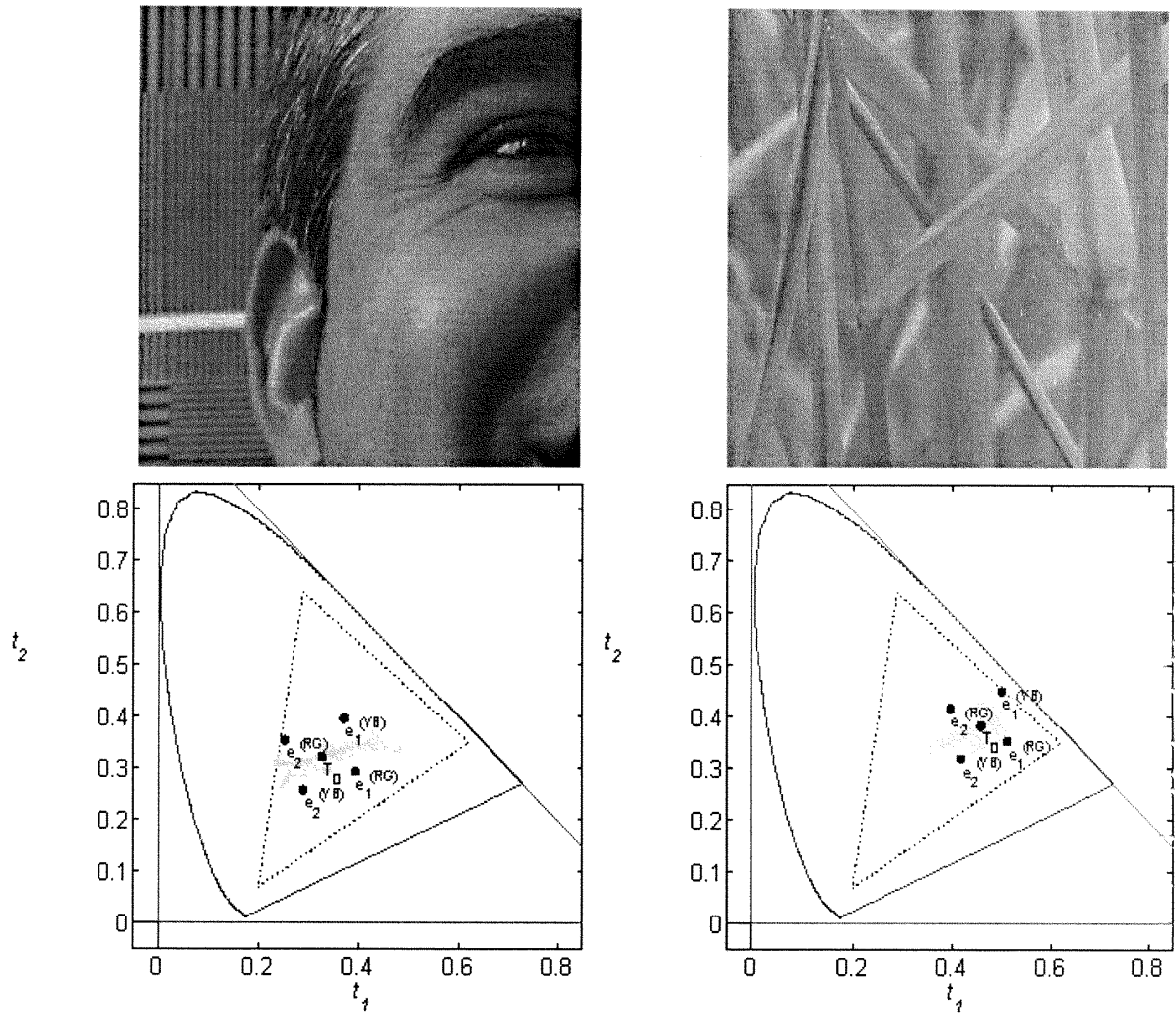


FIG. 3

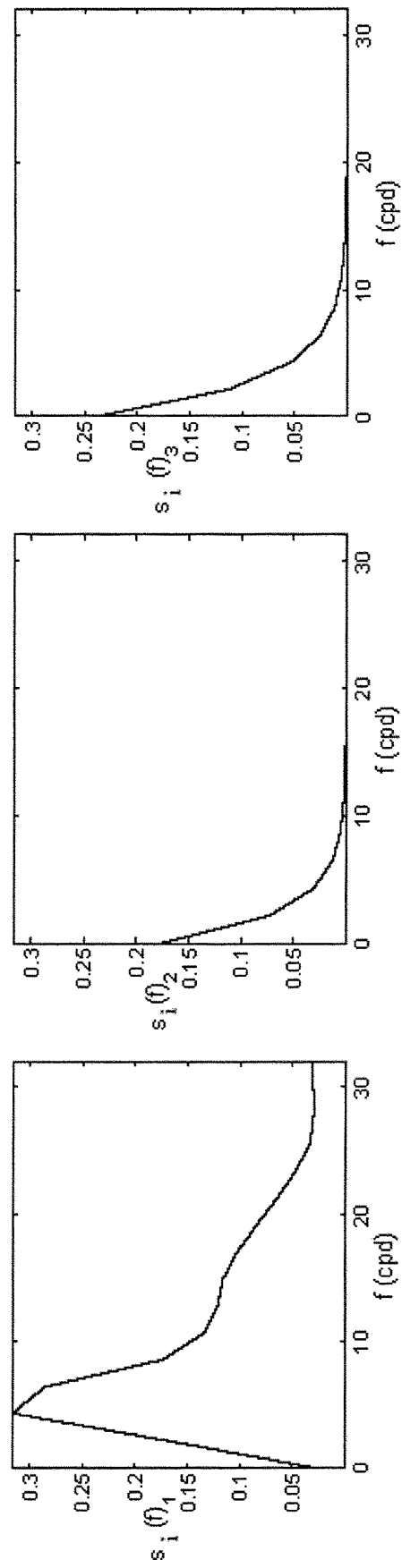


FIG. 4

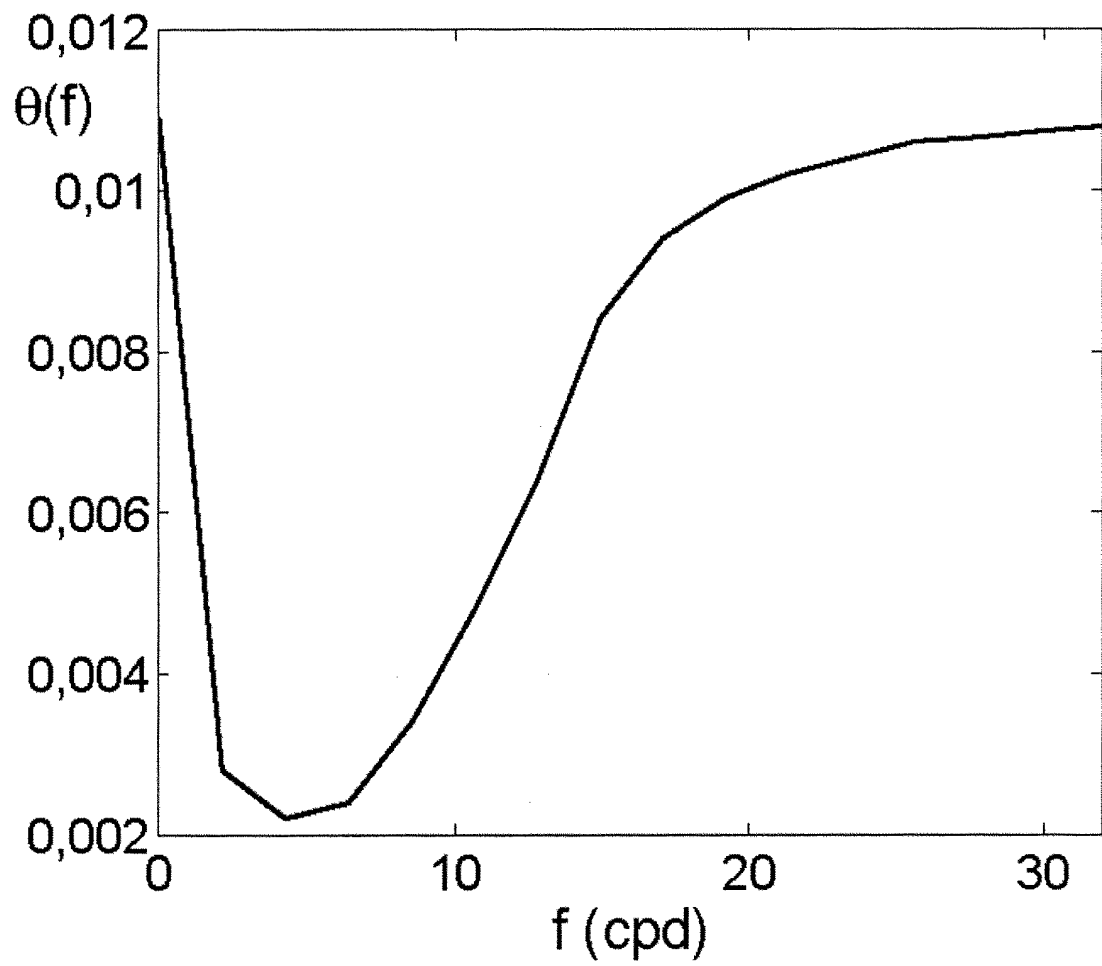


FIG. 5

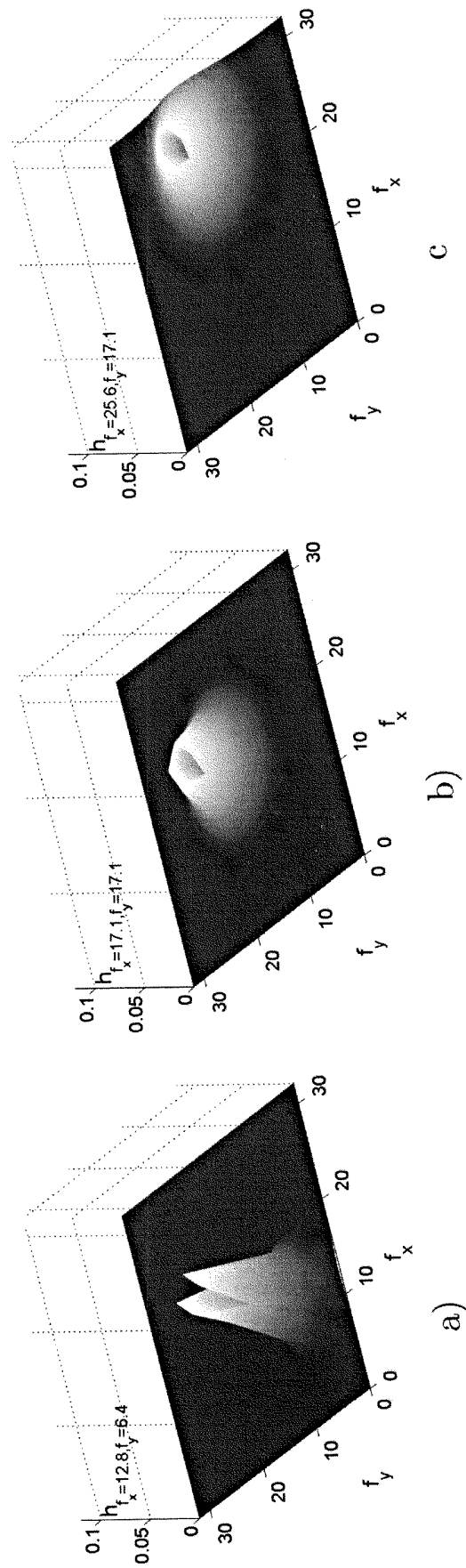


FIG. 6

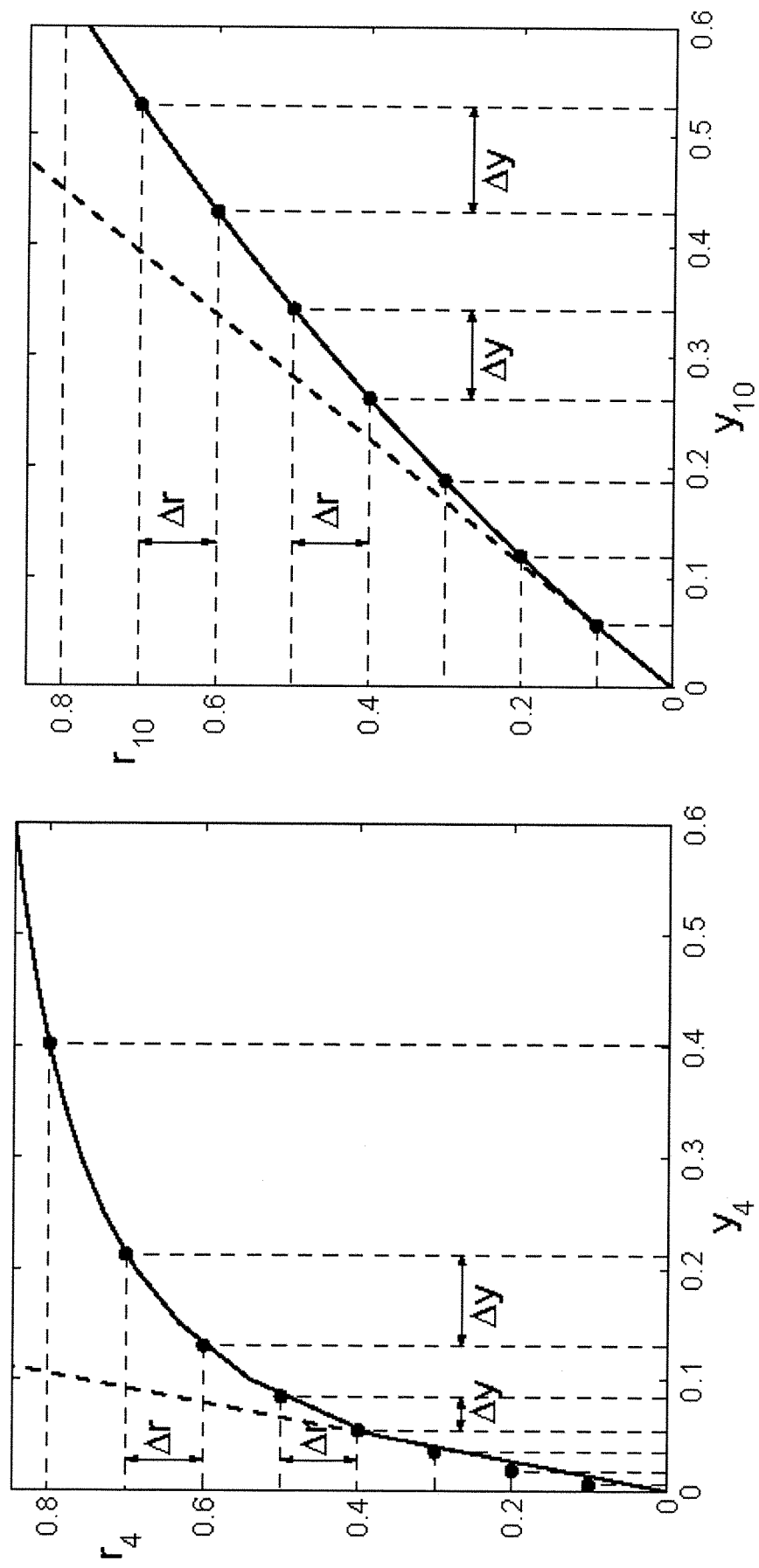


FIG. 7

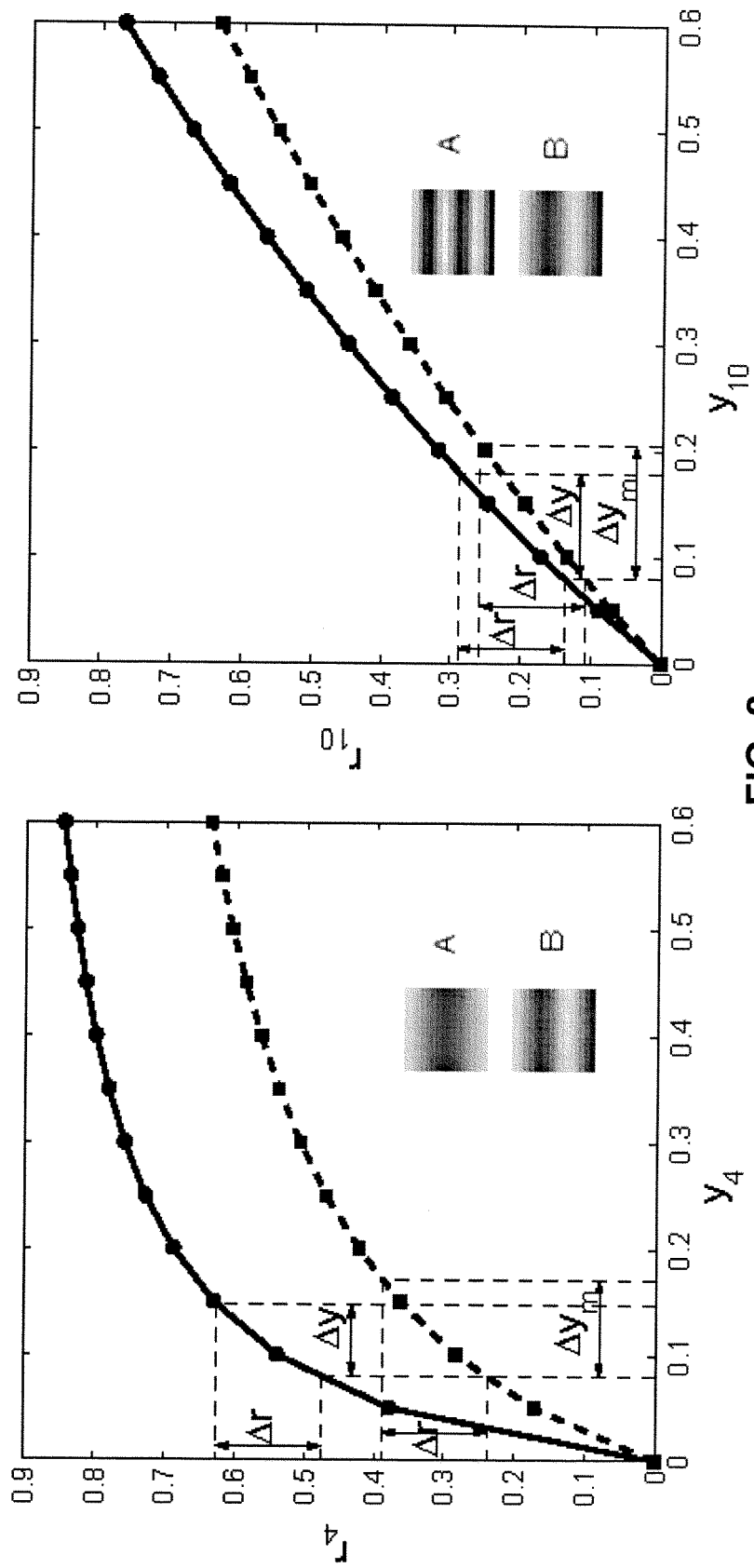


FIG. 8

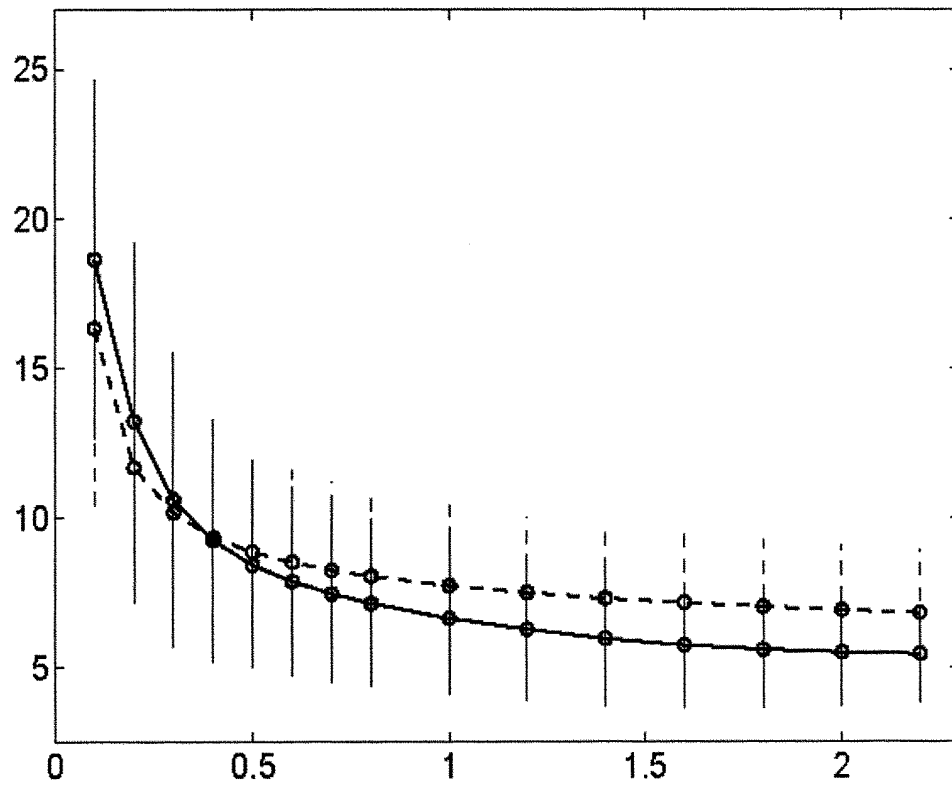


FIG. 10a

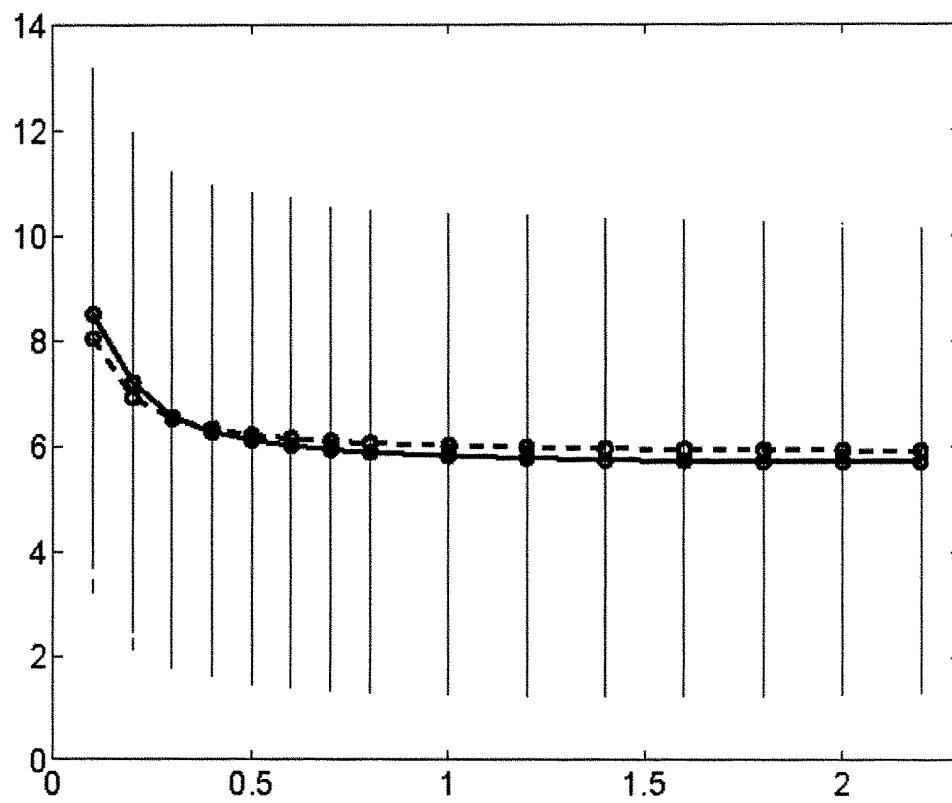


FIG. 10b

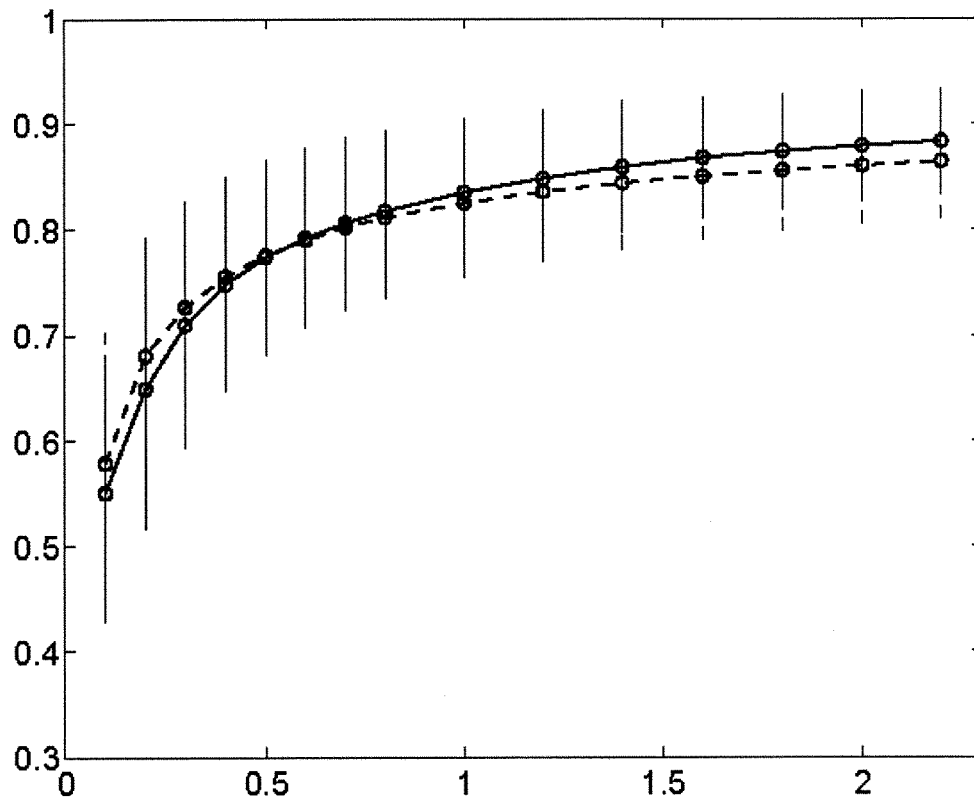


FIG. 10c

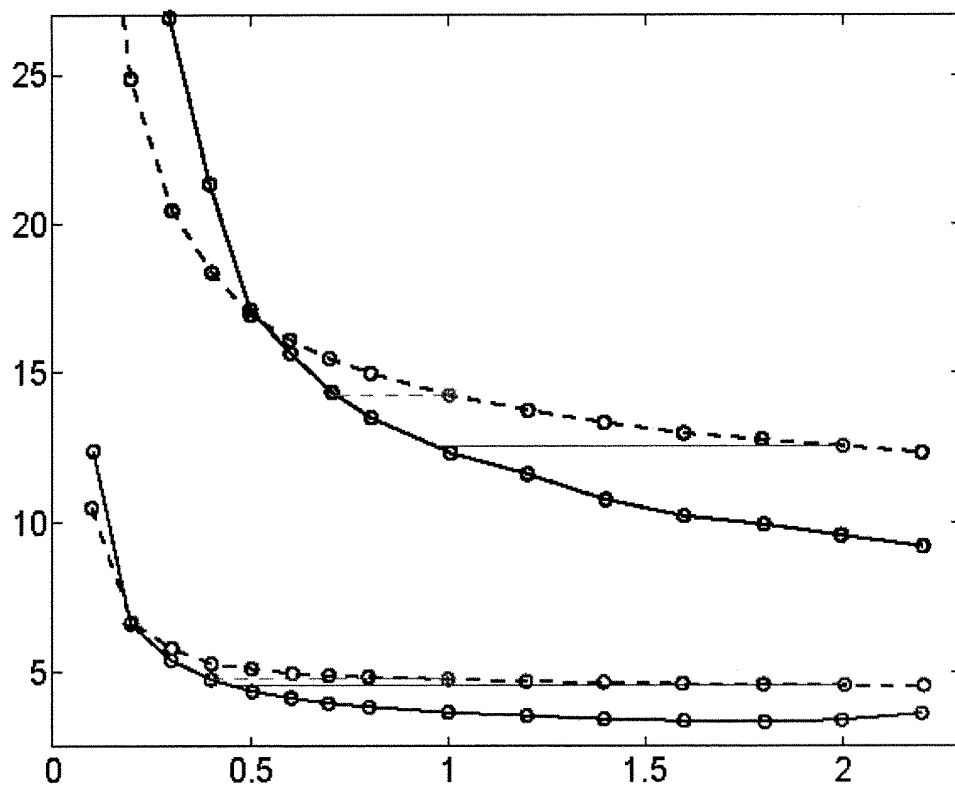


FIG. 11a

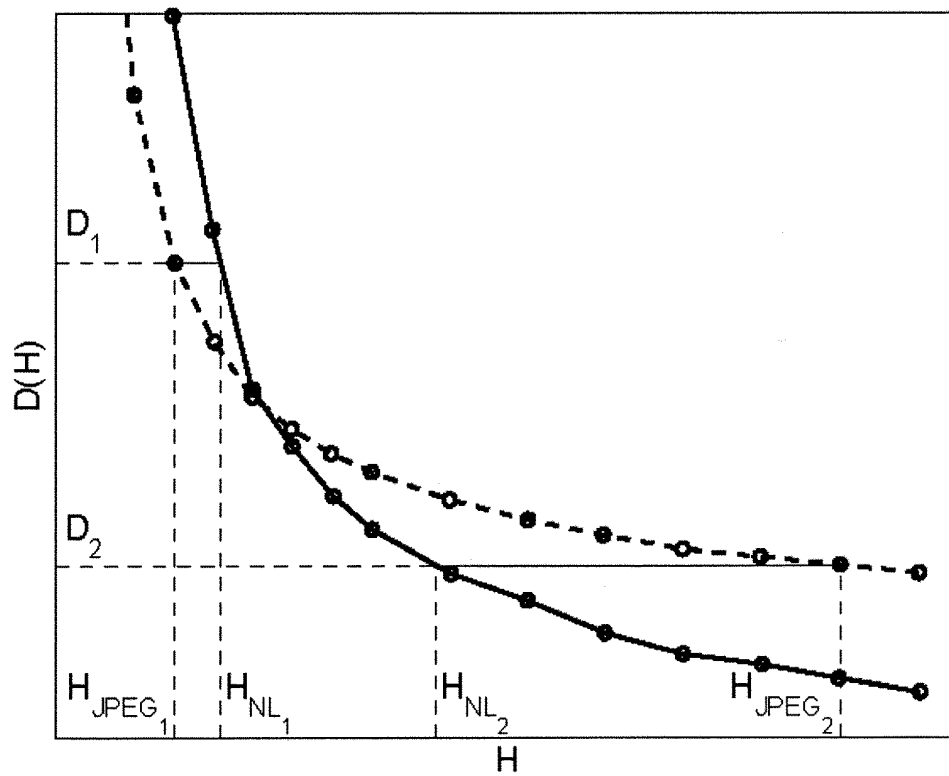


FIG. 11b

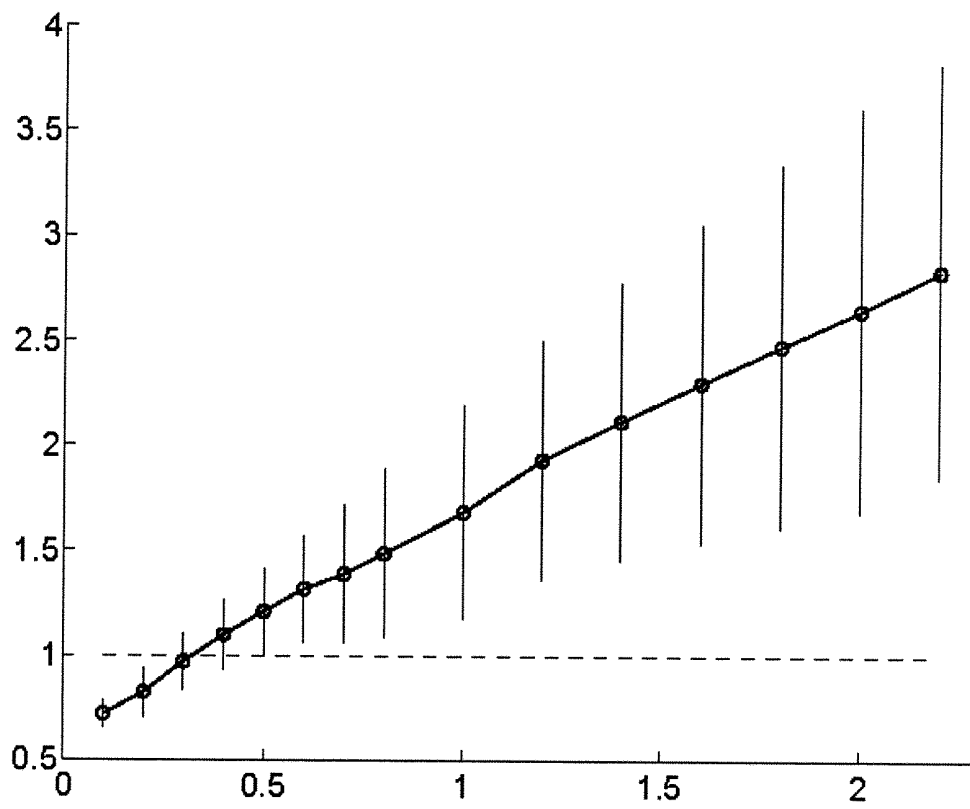


FIG. 12a

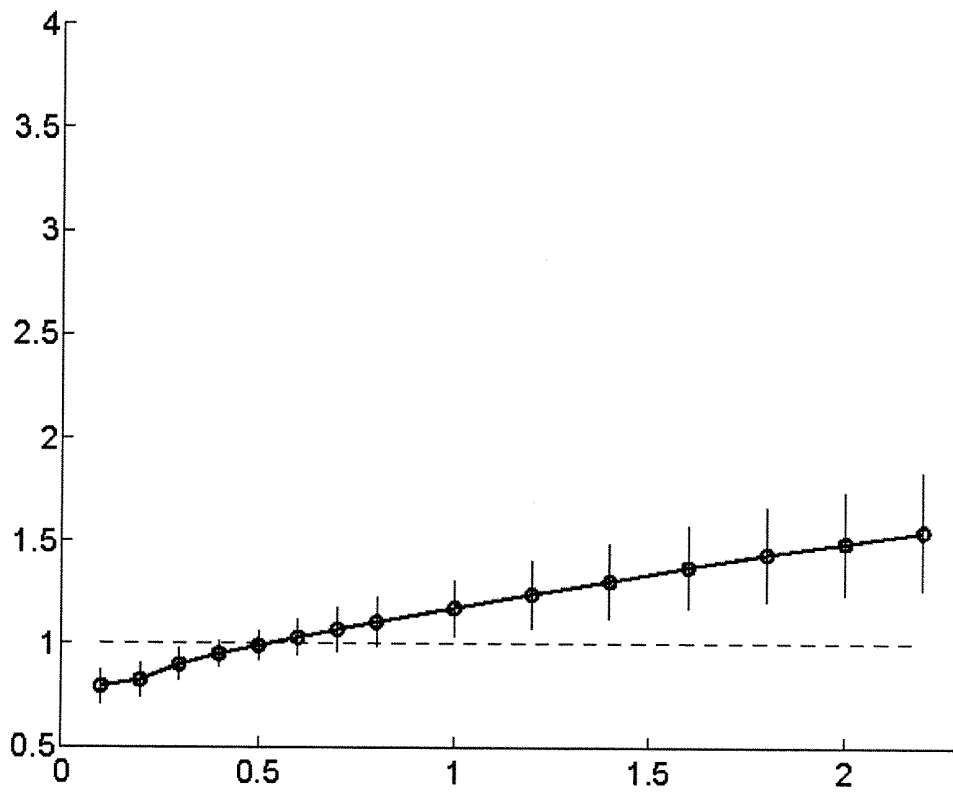


FIG. 12b

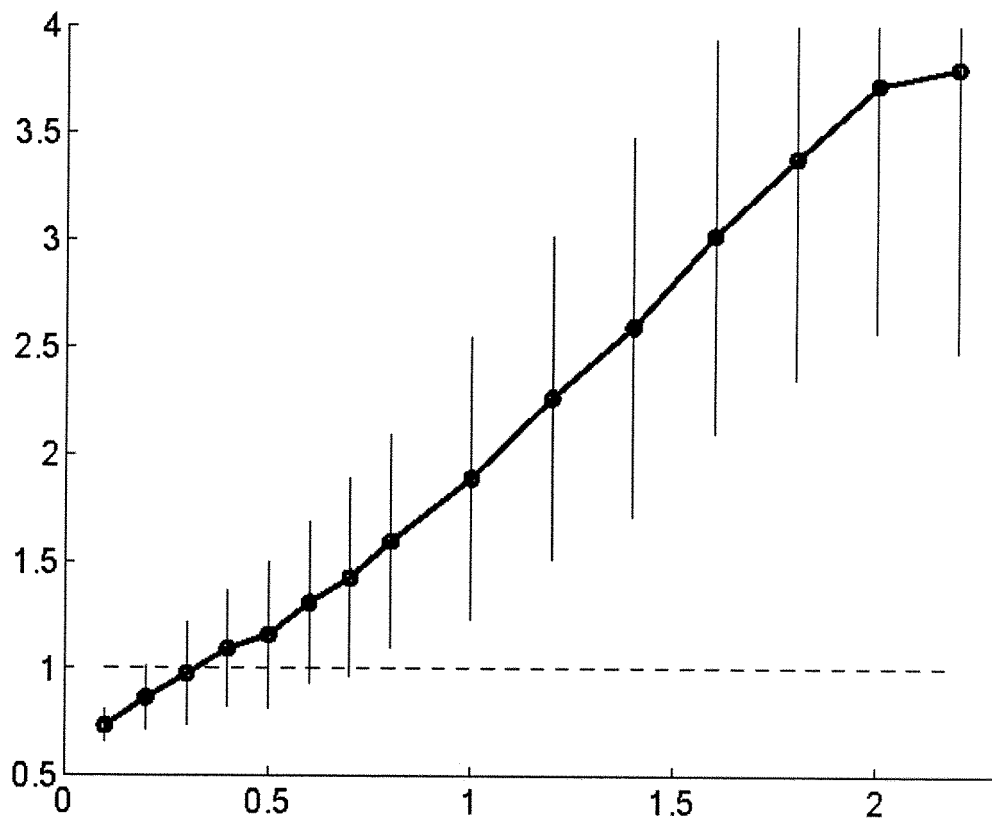


FIG. 12c

ES 2 310 157 B1

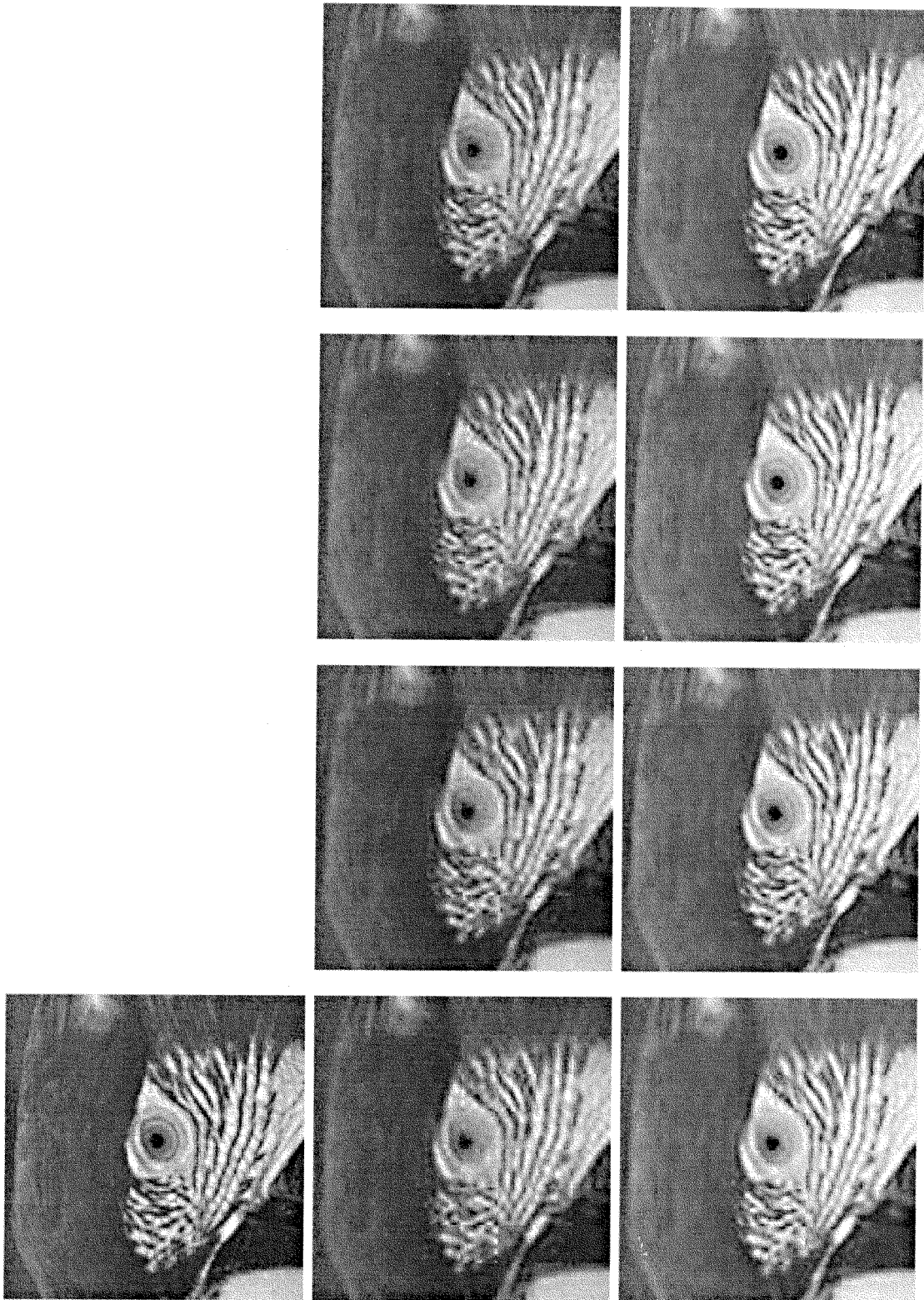


FIG. 13

ES 2 310 157 B1

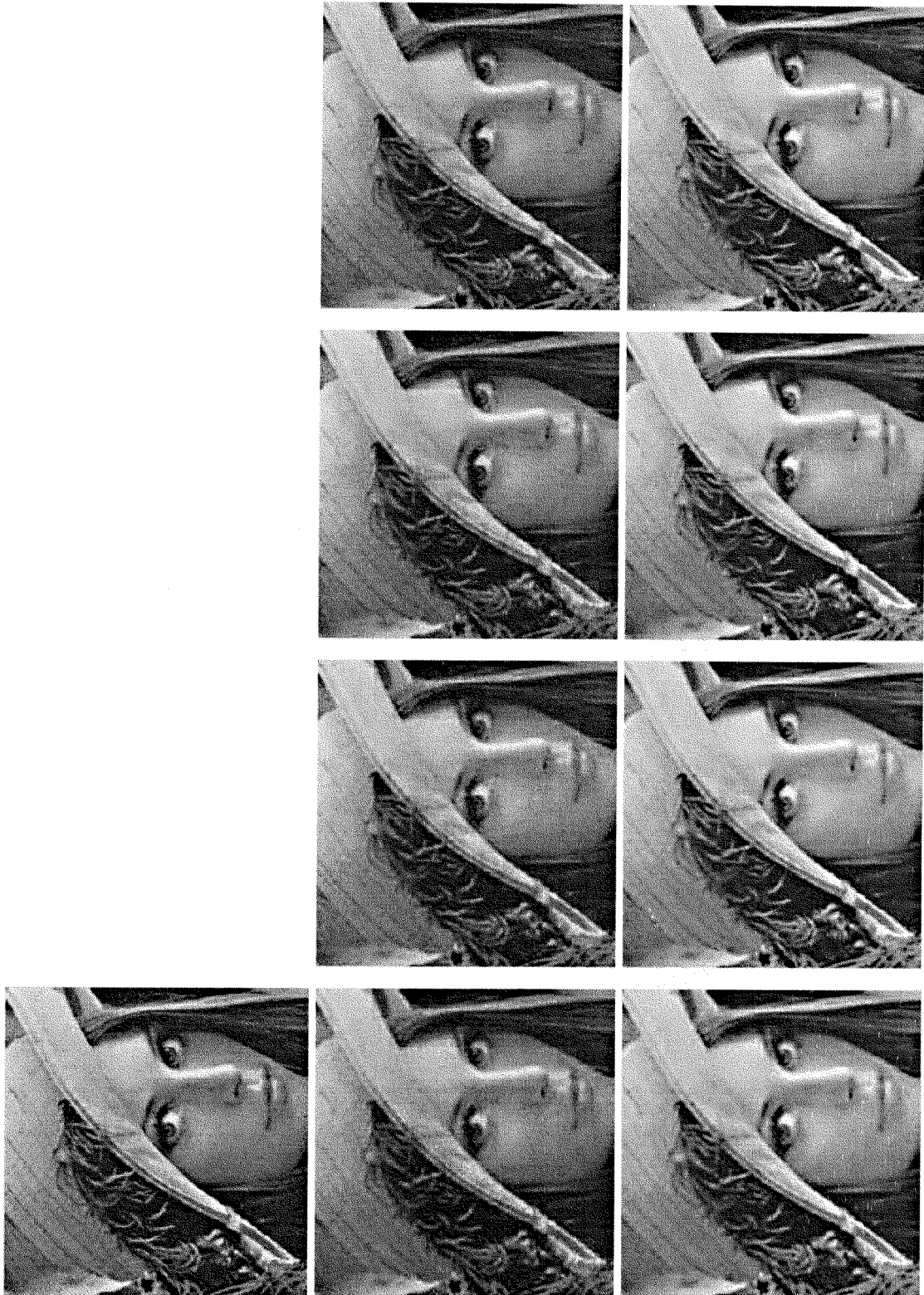


FIG. 14

ES 2 310 157 B1

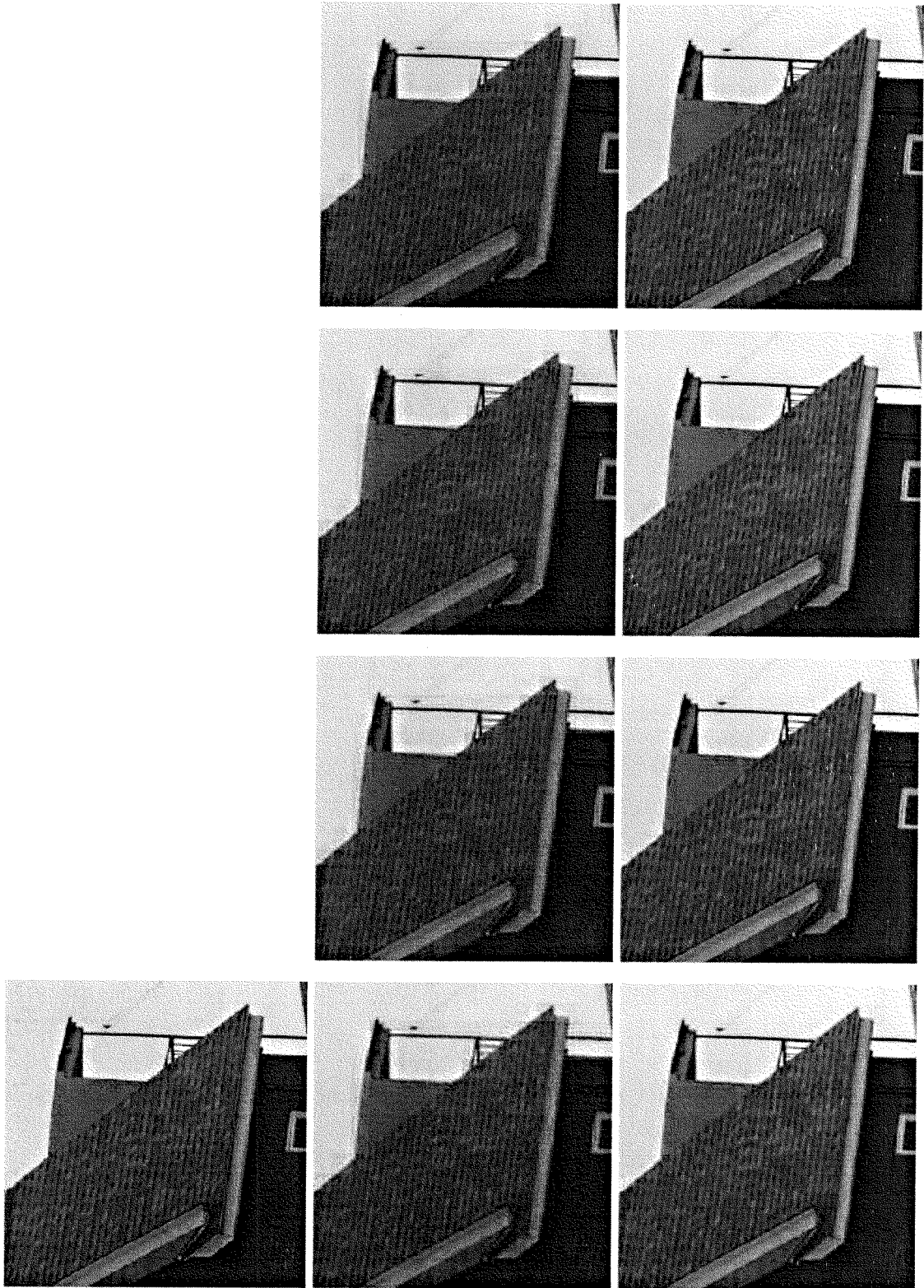


FIG. 15

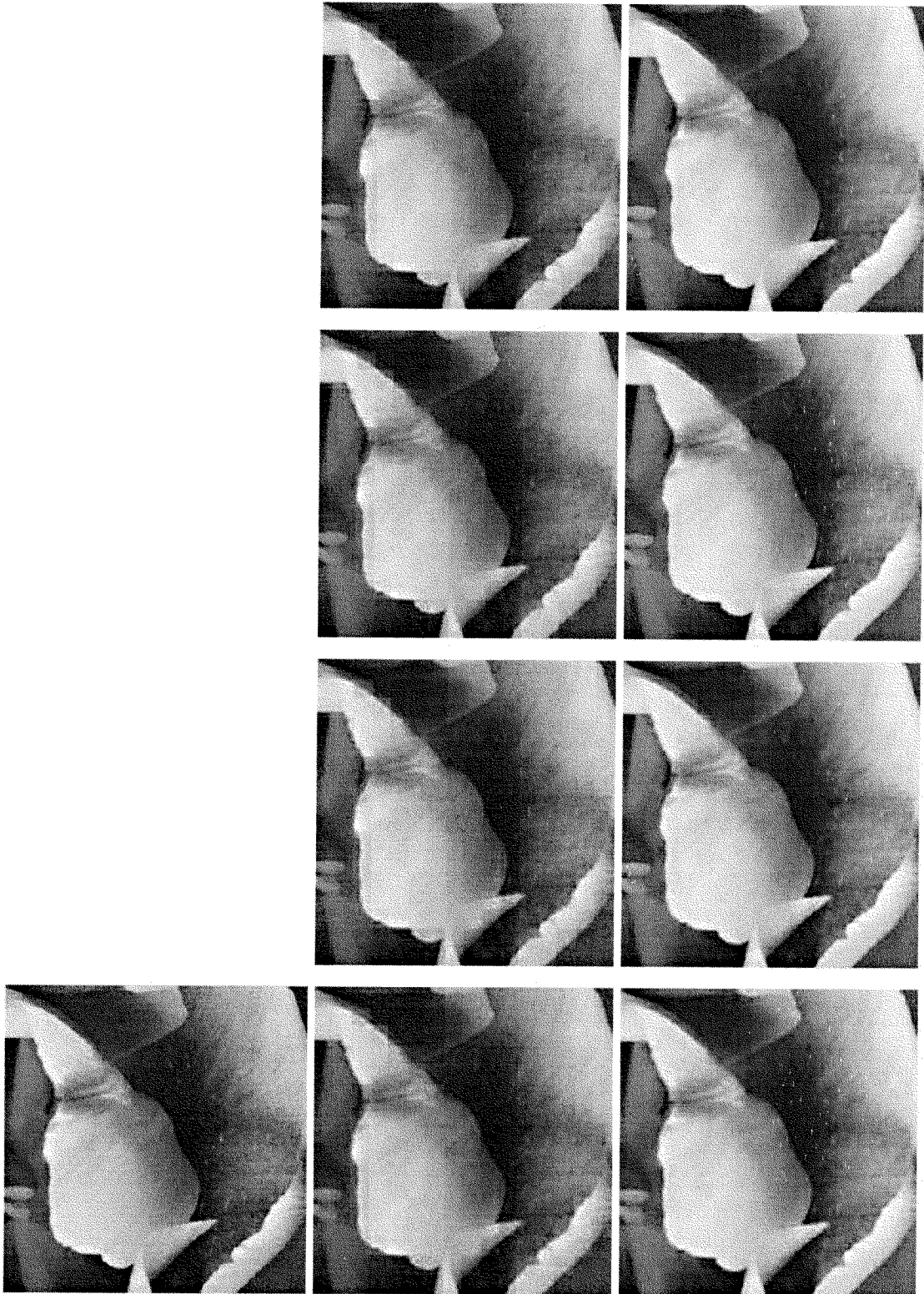


FIG. 16

ES 2 310 157 B1

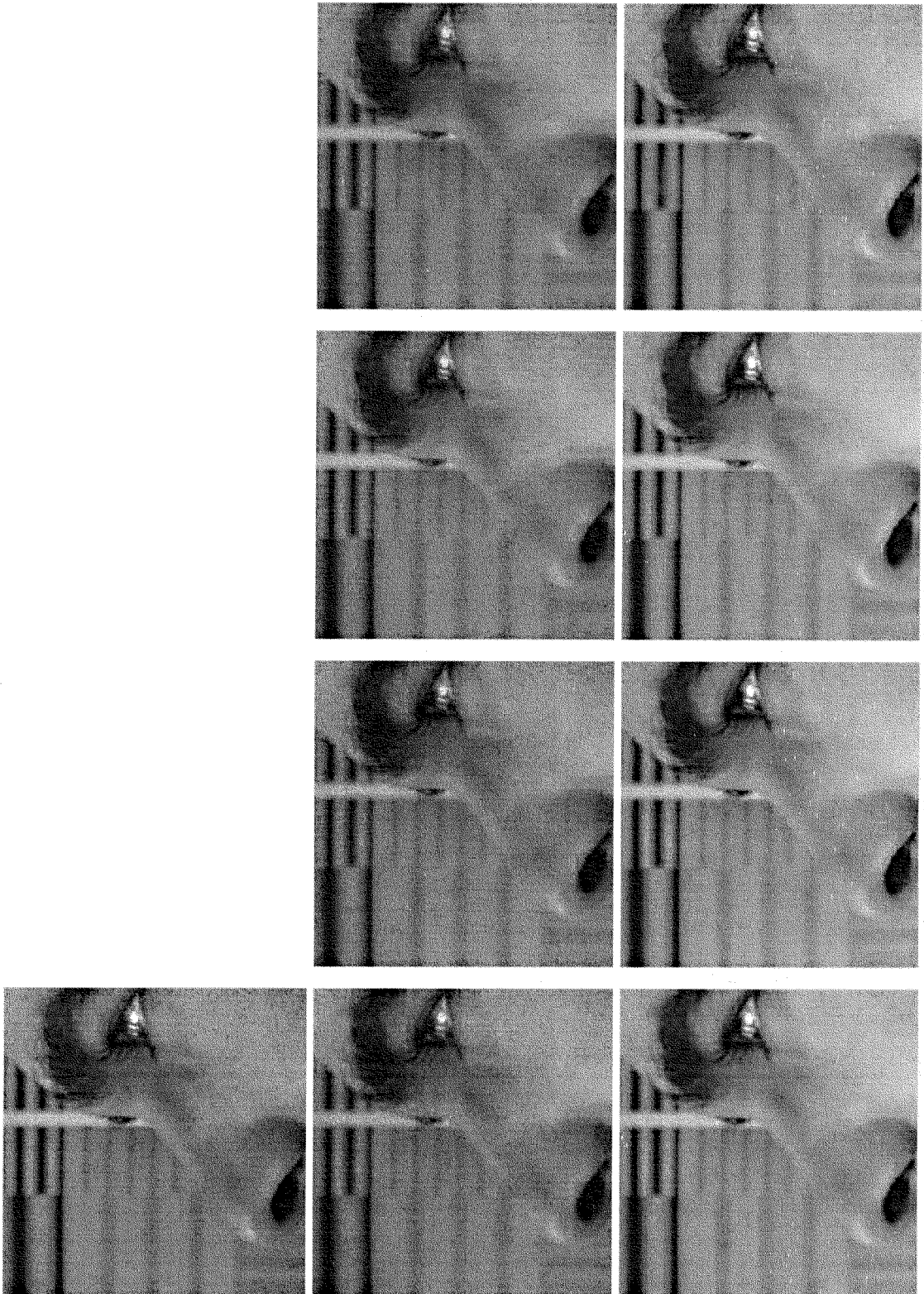


FIG. 17



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

⑪ ES 2 310 157

⑫ Nº de solicitud: 200801943

⑬ Fecha de presentación de la solicitud: 20.06.2008

⑭ Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TÉCNICA

⑮ Int. Cl.: Ver hoja adicional

DOCUMENTOS RELEVANTES

Categoría	⑯ Documentos citados	Reivindicaciones afectadas
A	WO 03050959 A1 (AUCKLAND UNISERVICES LTD; KECMAN VOJISLAV; ROBINSON JONATHAN) 19.06.2003	1,12
A	US 5426512 A (WATSON et al.) 20.06.1995	1,12
A	US 5883979 A (BERETTA et al.) 16.03.1999	1,12
A	US 2003152165 A1 (KONDO et al.) 14.08.2003	1,12

Categoría de los documentos citados

X: de particular relevancia

Y: de particular relevancia combinado con otro/s de la misma categoría

A: refleja el estado de la técnica

O: referido a divulgación no escrita

P: publicado entre la fecha de prioridad y la de presentación de la solicitud

E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado

☒ para todas las reivindicaciones

☐ para las reivindicaciones nº:

Fecha de realización del informe

28.11.2008

Examinador

Mª C. González Vasserot

Página

1/2

CLASIFICACIÓN DEL OBJETO DE LA SOLICITUD

H04N 7/30 (2006.01)

H03M 7/30 (2006.01)

G06T 9/00 (2006.01)

H04N 7/32 (2006.01)

H04N 7/34 (2006.01)