Autenticación de imágenes digitales

Dra. Mariko Nakano Miyatake

ESIME Culhuacan

IPN

Importancia

Actualmente todas las imágenes son digitales capturadas por cámaras digitales o escáneres. Estas imágenes son usadas para evidencias de hechos, tales como accidentes de automóviles, escándalos políticos entre otros.

Imágenes digitales se modifican fácilmente usando herramientas computacionales (Photoshop, CorelDRAW)

Necesidad de desarrollar métodos eficiente para autenticar imágenes digitales

Métodos para autenticar imágenes digitales

1. **Métodos basados en forense digital**
2. **Métodos basados en firma digital**

   (Hash criptográfico)

3. **Métodos basados en la técnica de hash perceptual**

   (Image Hashing)

4. **Métodos basados en la técnica marca de agua digital**
   - **Marca de agua frágil**
   - **Marca de agua semi-frágil**
     - **Propuesta 1**
     - **Propuesta 2**

Métodos basados en forense digital

- Cuando ocurre incidente, se analiza algunos artefactos en la imagen para determinar existencia de alteración.
  - Dirección de Luz
  - Dirección de sombra
  - Continuidad de los objetos

- Se puede localizar región donde ocurrió alteración.
- Los métodos no es robusto a compresión de imágenes

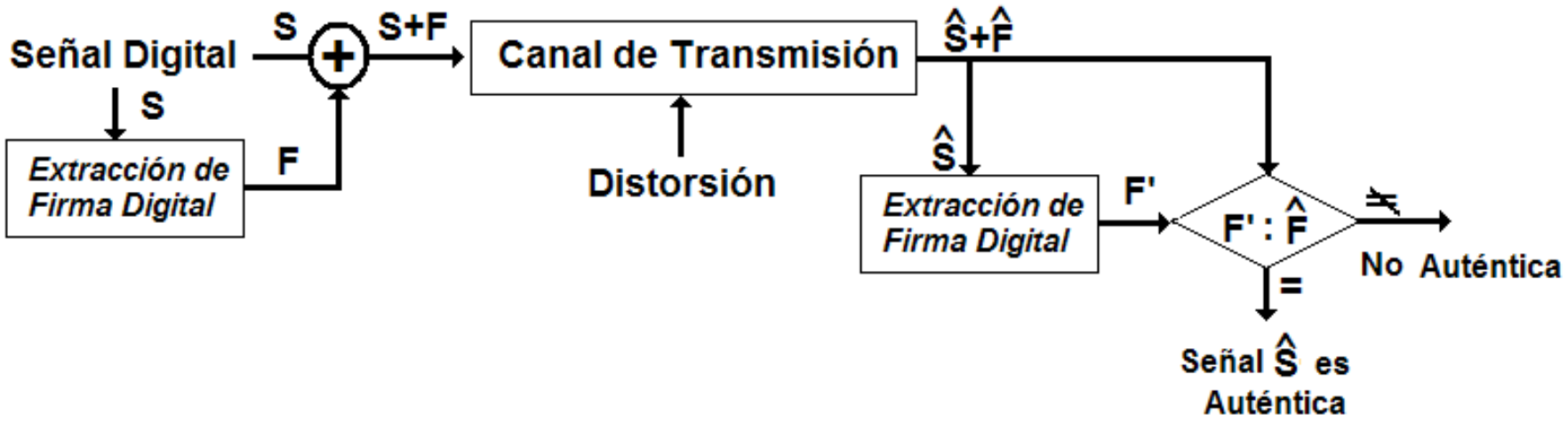
Métodos basados en forense digital Ejemplo

Discontinuidad de región donde aparece el rostro

Tasa de compresión de JPEG es continuo ó no?

Métodos basados en firma digital (Hash Criptográfico)

- Uso de función hash criptográfico
  - MD5, SHA-1, SHA-256
- Sensible a cualquier cambio



Métodos basados en firma digital (Hash Criptográfico)

Imagen en formato bit map

Cambio de formato

Imagen en formato jpeg

MD5 =7b40a69c0d1fa518ef4aa ……..

MD5 =c47da18f4ee320dsa8d……..

Características de imágenes digitales

- Existen varios formatos (JPEG, TIFF, PNG, etc.) y la conversión del formato es muy común.
- Generalmente transmite o almacena en la versión comprimida. (JPEG ó JPEG2000)
- La menor distorsión no se percibe por sistema visual humano.

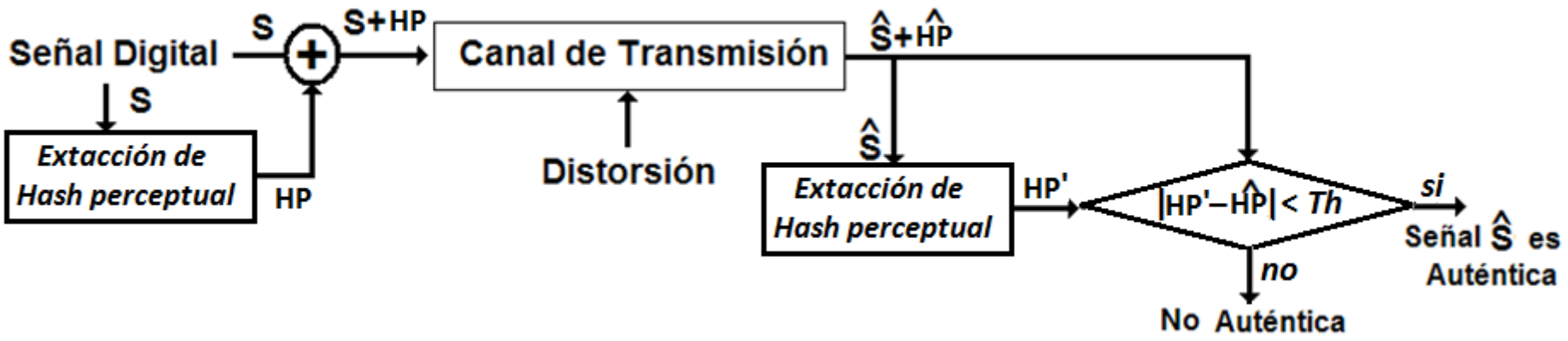
FC=100 8bits/pixel

FC=80 2.5bits/pixel

FC=60 0.8 bits/pixel

Métodos basados en Hash Perceptual

- Esquema es mismo que Hash Criptográfico
- Extrae una secuencia (el código Hash) de la imagen, la cual tiene que ser robusto a compresión y contaminación por ruido.
- El código Hash se transmite junto con la imagen original



Características de Hash Perceptual

- Las secuencias extraídas (código hash) de dos imágenes similares son muy parecidas.
- Las secuencias extraídas (código hash) de dos imágenes diferentes son muy diferentes.

$$HP(A) \approx HP(A_{comp})$$

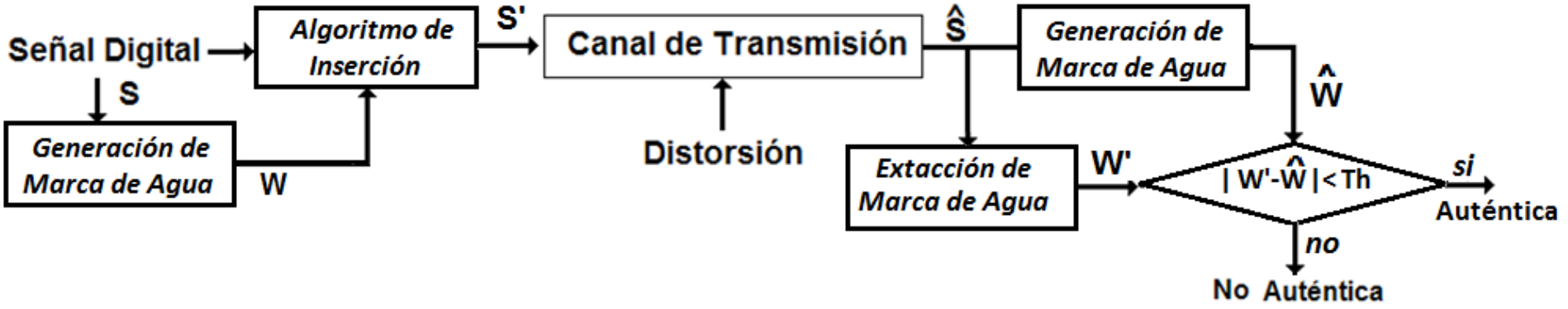
$$HP(A) \neq HP(B)$$

$$Si \;\; dist\left(HP(A), HP(A_{recibido})\right) < Th \;\; \text{entonces}$$

$$A_{recibido} \;\; \text{es} \;\; \text{Auténtica}$$

Métodos basado en marca de agua

- Extrae una secuencia robusta desde la imagen que quiere autenticar
- La secuencia extraída se inserta dentro de la misma imagen para generar imagen marcada
- La imagen marcada se circula en la red pública



Marca de Agua Frágil y Semi-Frágil

- Marca de Agua Frágil

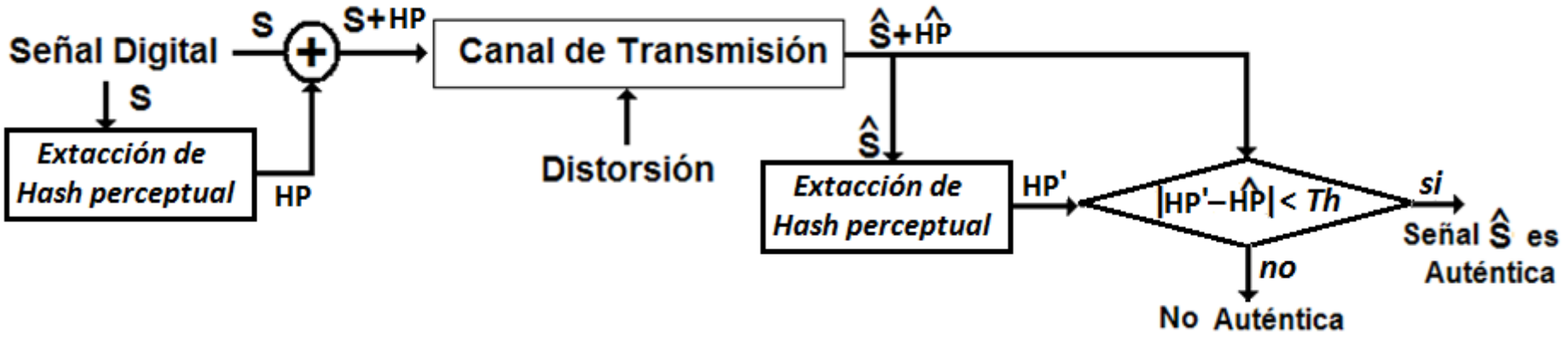
  Autenticación completa – Detecta cualquier cambio o modificación.

- Marca de Agua Semi-Frágil

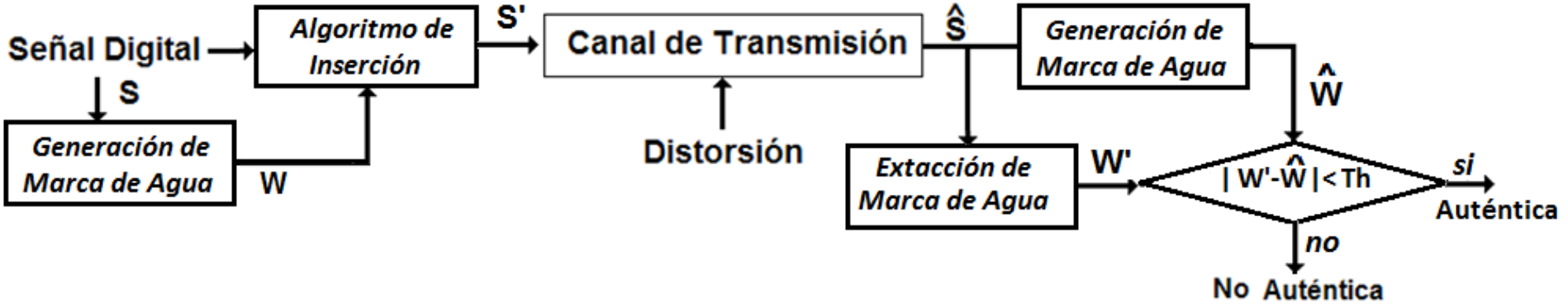
  Autenticación de contenido – Detecta cambio de contenido. Robusto a procesamiento que no altera el contenido de la imagen, tales como compresión, contaminación por ruido

Hash Perceptual : Marca de agua

Hash Perceptual (Image Hashing)



Marca de Agua (Watermarking)



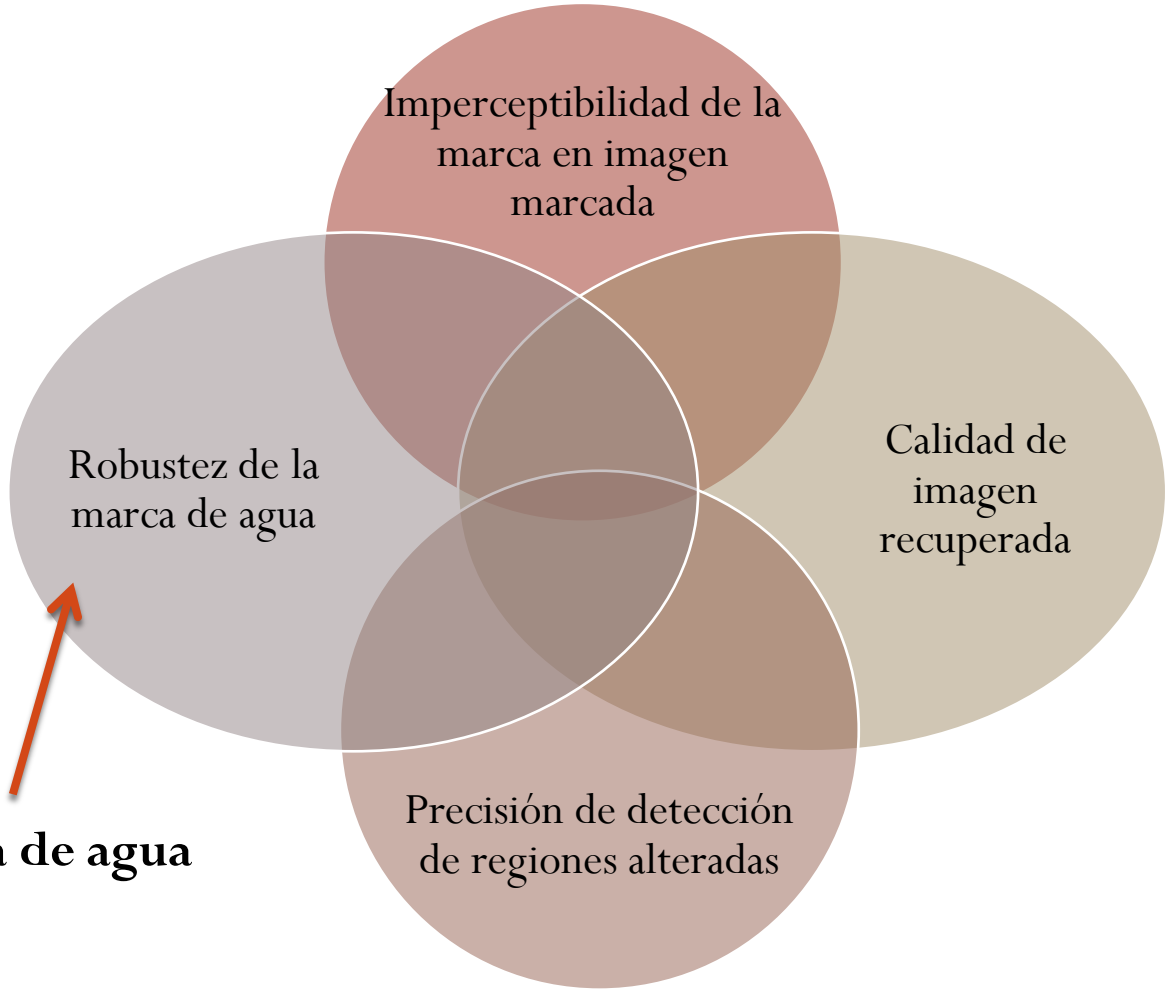
Hash Perceptual : Marca de agua

	Hash Perceptual	Marca de agua
Ventajas	• No causa distorsión a la imagen que se transmite	• No aumenta el ancho de banda en la transmisión • Puede recuperar versión original de región alterada
Desventajas	• Aumenta el ancho de banda en la transmisión • No se puede recuperar el contenido alterado	• Causa distorsión aunque no se percibe por el sistema visual humano

Comparación de los métodos de autenticación

Métodos	Autenticación (si ó no)	Localización de regiones alteradas	Recuperación de regiones alteradas
Forense digital	✓	✓	X
Hash criptográfico	✓	X	X
Hash Perceptual (Image Hashing)	✓	✓	X
Marca de Agua Frágil	✓	✓	✓
Marca de Agua Semi-Fragil	✓	✓	✓

Requerimientos de técnica de marca de agua



Caso de marca de agua Semi-Frágil

Propuesta 1 :Marca de agua semi-frágil usando la técnica de tono- medio (Halftoning)

Características

1. **Detección precisa de regiones alteradas**
2. **Uso de técnica de tono-medio (halftoning)**
3. **Recuperación de regiones alteradas con mayor calidad**
4. **Robusto a los procesos que conservan el contenido. (compresión, ruido aditivo)**
5. **Inserción y extracción de marca de agua se realzan en el dominio de frecuencia.**

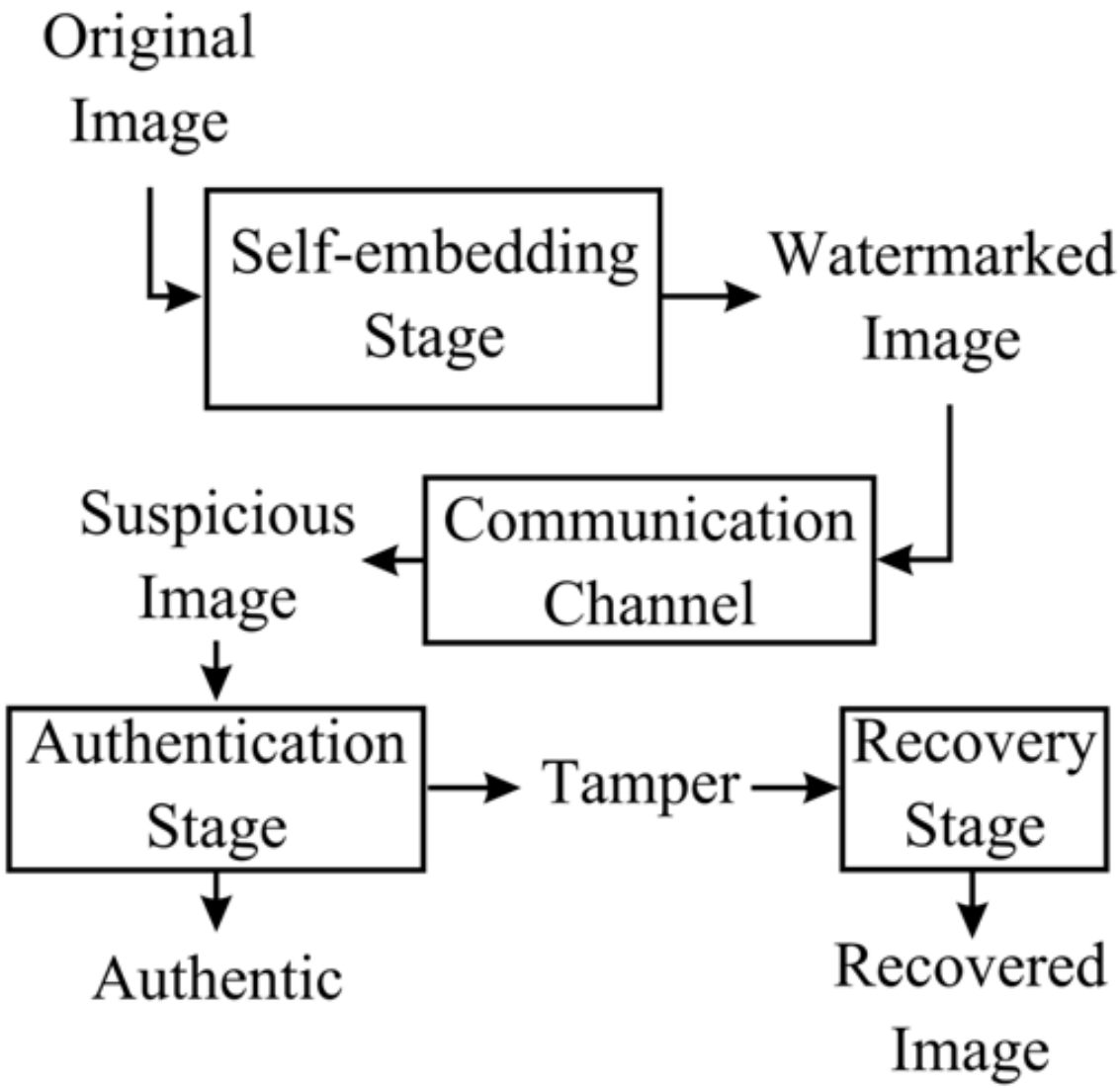
Ejemplos

Imagen original

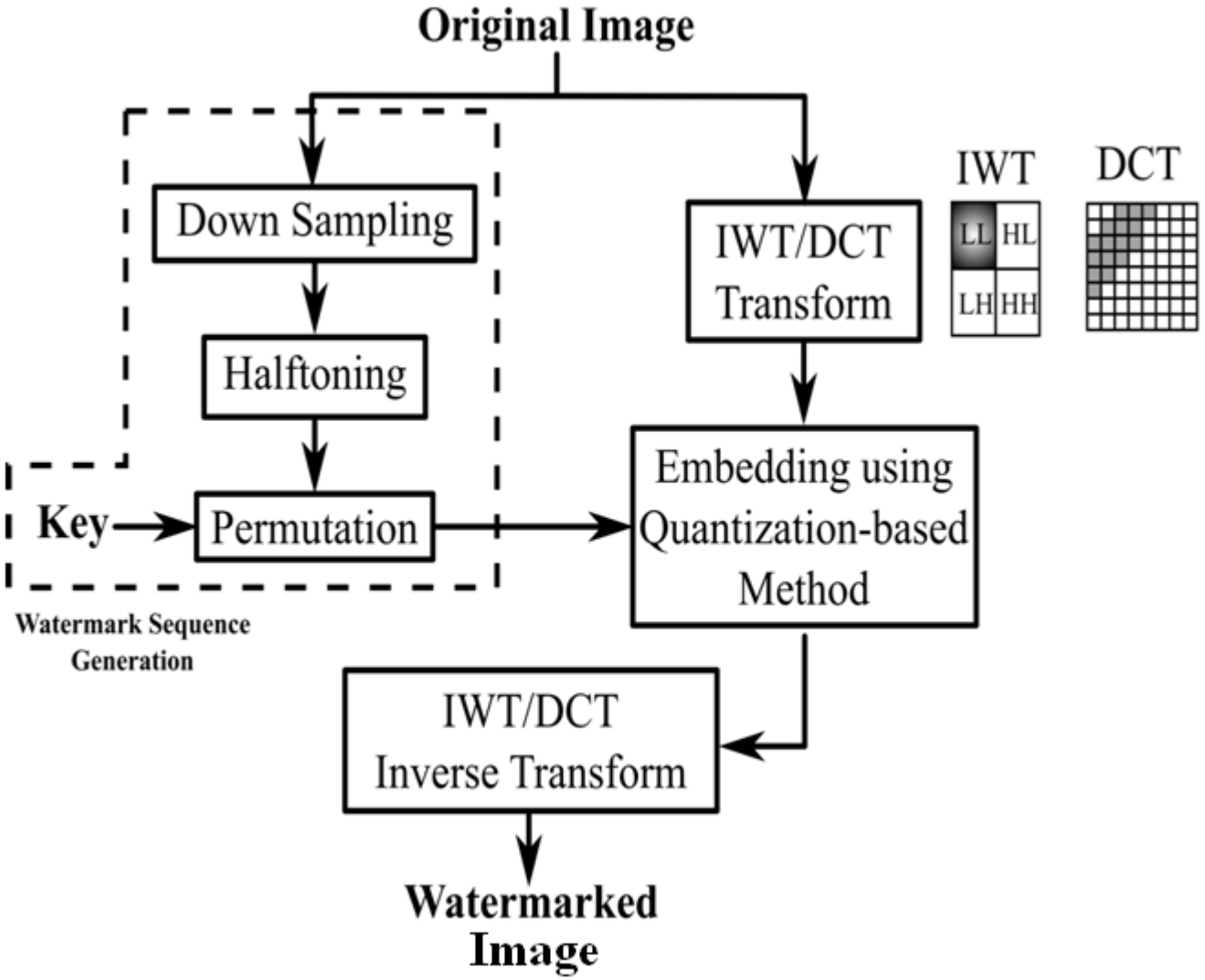
Imagen alterada

Imagen recuperada

Esquema global

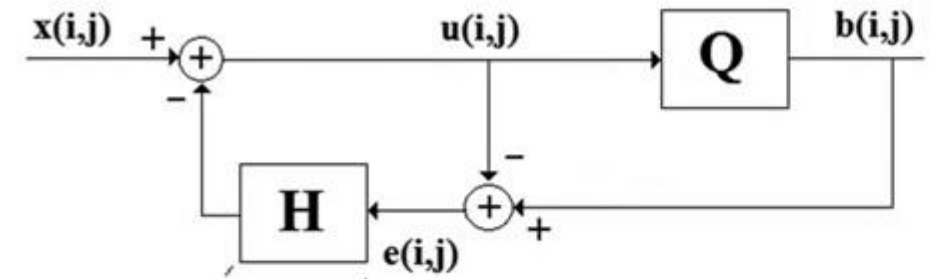


Inserción de marca de agua

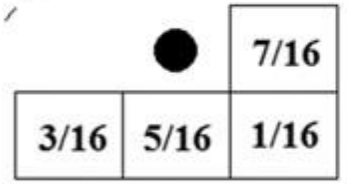


Generación de la marca de agua

Marca de agua es una versión *halfton* de la imagen original



$$b(i,j)=\begin{cases}0 & \text{if } u(i,j)<T_b\\ 1 & \text{otherwise}\end{cases}$$



Filtro de Floyd-Steinburg

Imagen original (escala de grises) 8 bits

Imagen halftone (binaria) 1 bit

Generación de marca de agua

Imagen halftone

Imagen binaria, pero tiene una apariencia de imagen con escala de grises.

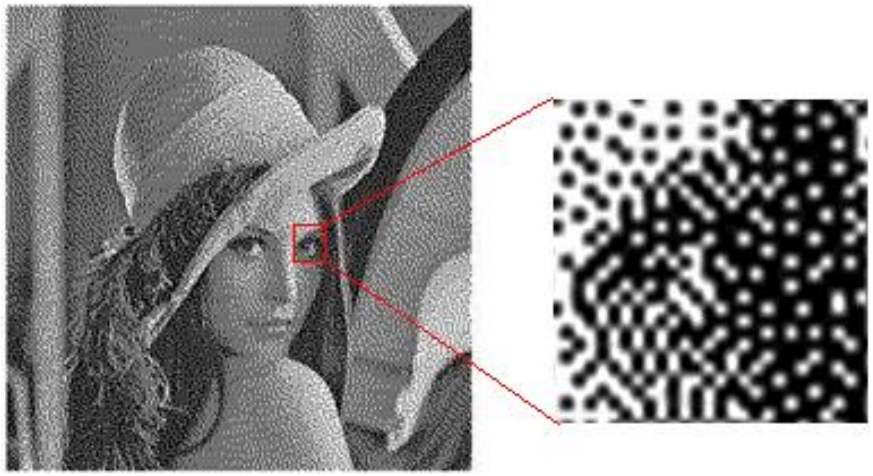
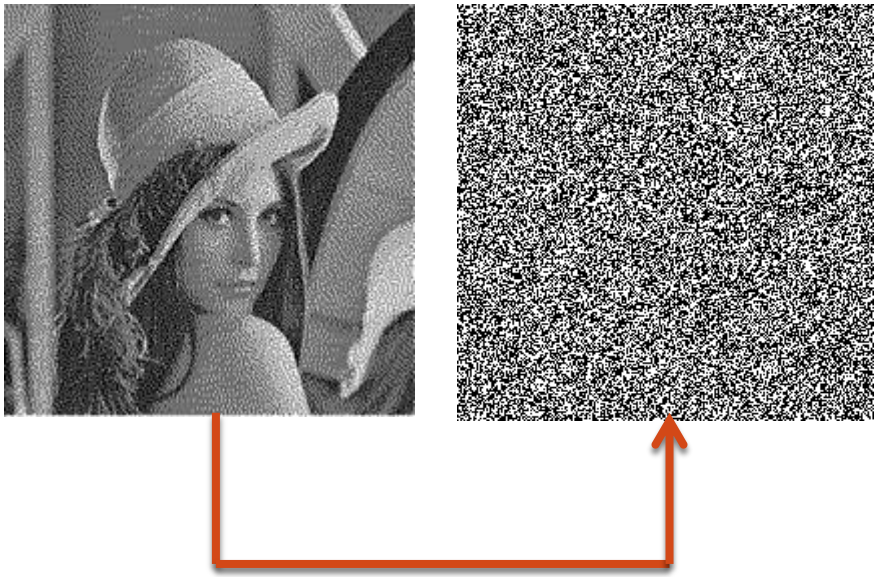


Imagen Halftone

Marca de agua (binaria)



Permutación

(llave secreta)

Algoritmo de inserción

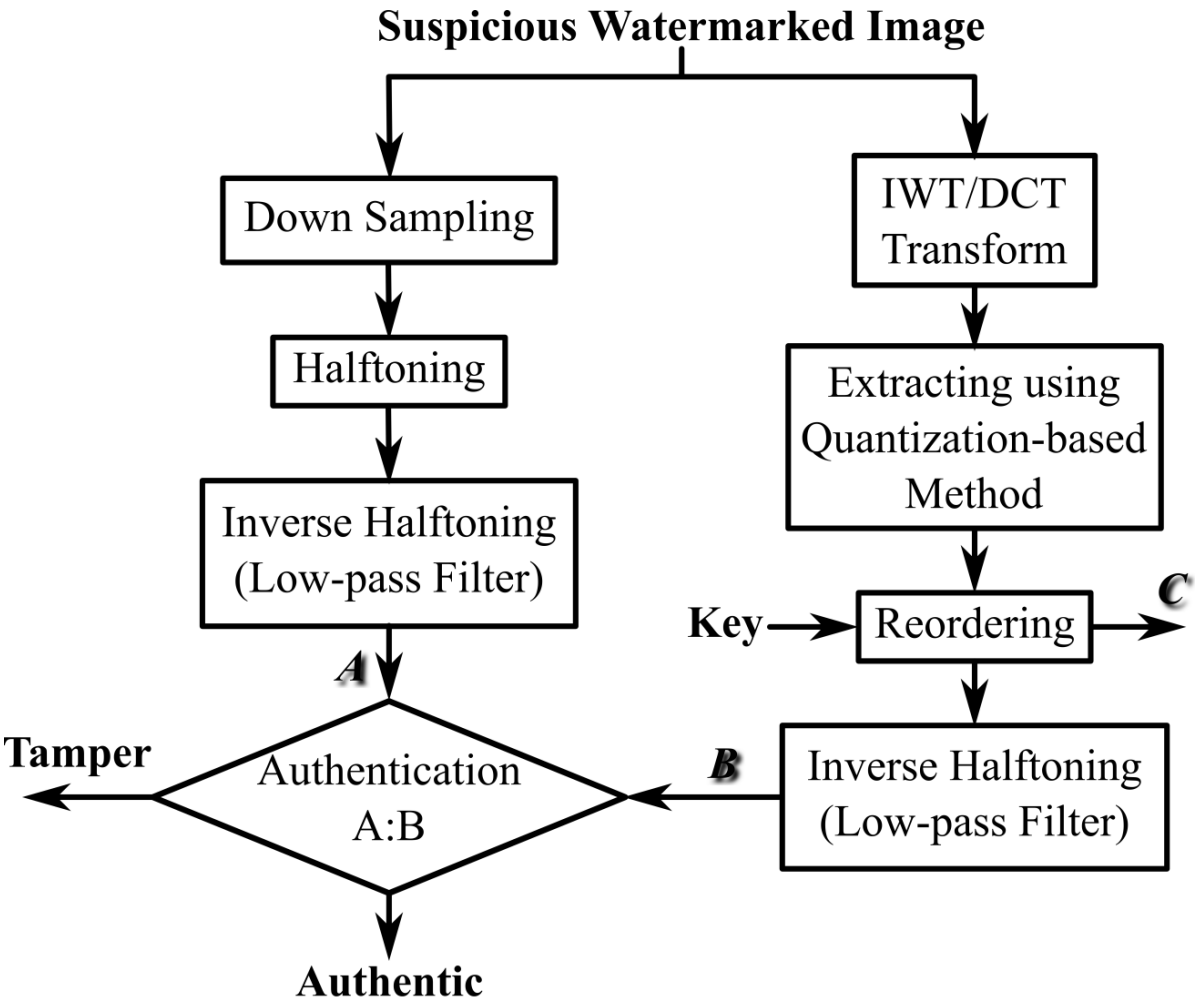
- El algoritmo de inserción de marca de agua está basado en el método de Cuantificación

$$\tilde{c}_n = \begin{cases} v_1 & \text{if } |c_n - v_1| \le |c_n - v_2| \\ v_2 & \text{otherwise} \end{cases}$$

$$v_1 = \begin{cases} \operatorname{sgn}(c_n) \times \left\lfloor \frac{|c_n|}{2S} \right\rfloor \times 2S, & \text{if } w_n = 0 \\ \operatorname{sgn}(c_n) \times \left(\left\lfloor \frac{|c_n|}{2S} \right\rfloor \times 2S + S \right), & \text{if } w_n = 1 \end{cases}, \quad n = 1..N$$

$$v_2 = v_1 + \operatorname{sgn}(c_n) \times 2S$$

Detección de regiones alteradas



Detección de regiones alteradas

Extracción de marca de agua

$$\tilde{w}_n = \begin{cases} 0 & \text{if} \quad round\left(\hat{c}_n / S\right) = even \\ 1 & \text{if} \quad round\left(\hat{c}_n / S\right) = odd \end{cases} , \; n = 1..N$$

Halftoning Inversa --- Filtro pasa baja

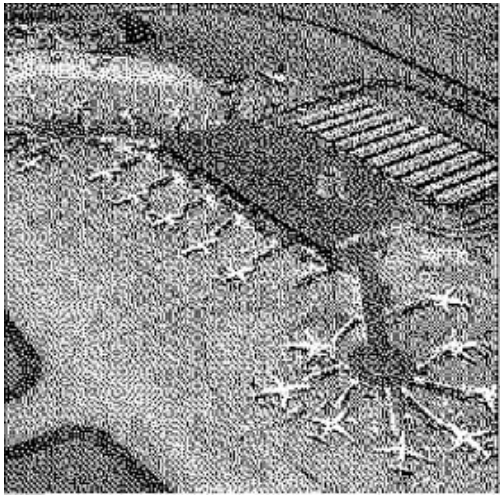
$$F_G = \frac{1}{11.566}\begin{bmatrix} 0.1628 & 0.3215 & 0.4035 & 0.3215 & 0.1628 \\ 0.3215 & 0.6352 & 0.7970 & 0.6352 & 0.3215 \\ 0.4035 & 0.7970 & 1 & 0.7970 & 0.4035 \\ 0.3215 & 0.6352 & 0.7970 & 0.6352 & 0.3215 \\ 0.1628 & 0.3215 & 0.4035 & 0.3215 & 0.1628 \end{bmatrix}$$

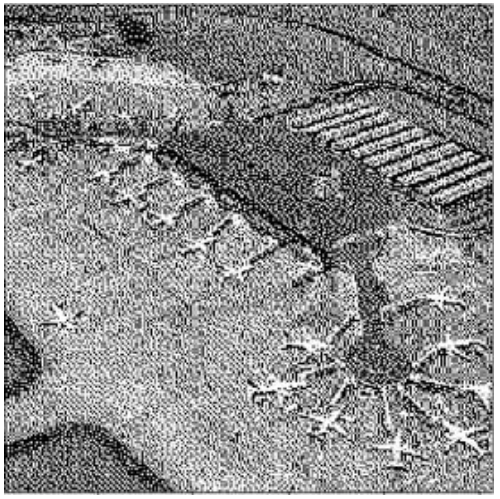
Imagen de entrada

Imagen original

Halftoning

Extracción de marca de agua



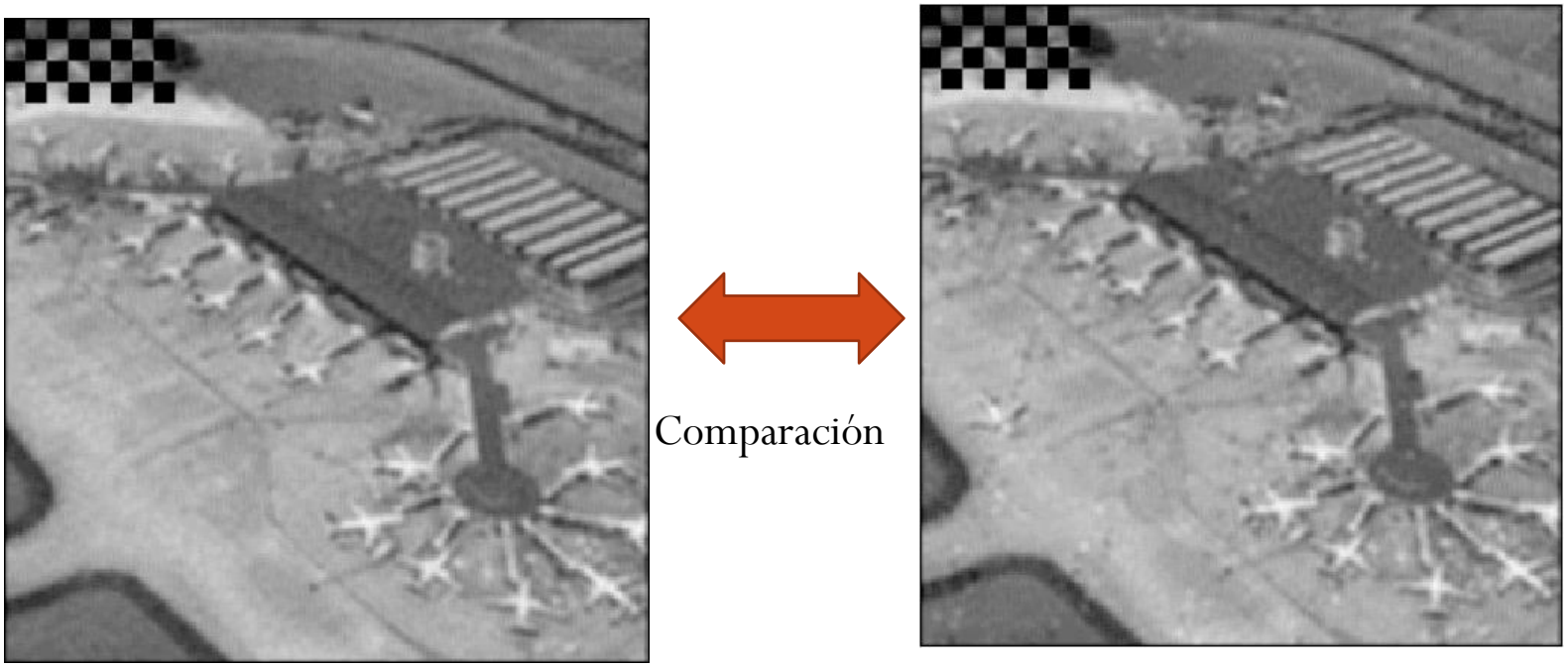


Esta comparación produce grandes errores debido a que las dos imágenes son binarias

Detección de regiones alteradas

- Aplicar halftoning inversa a ambas imágenes para la comparación perceptiva.
- Comparación por bloques (64 x 64)

A **B**

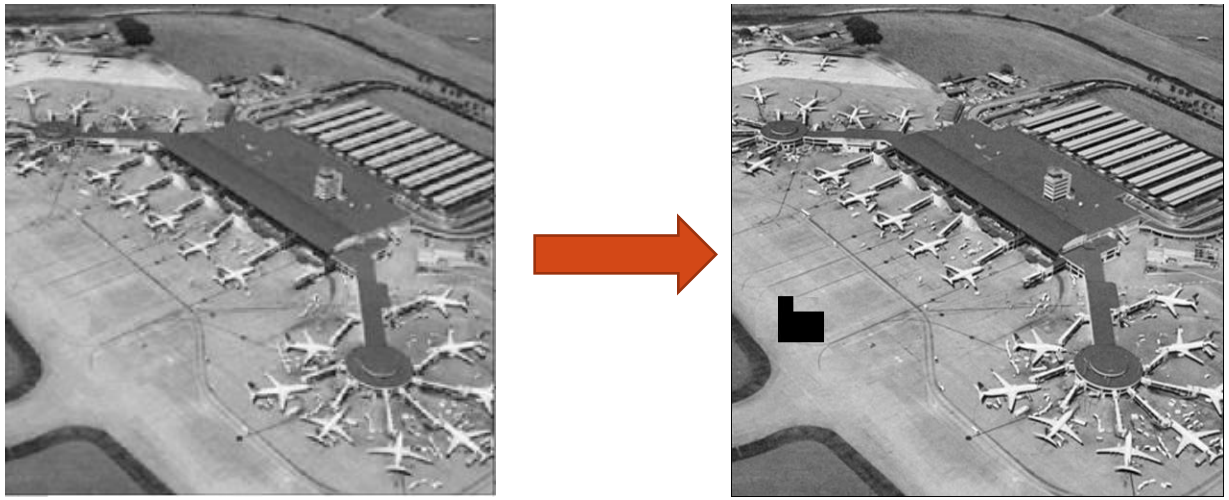


Detección de regiones alteradas

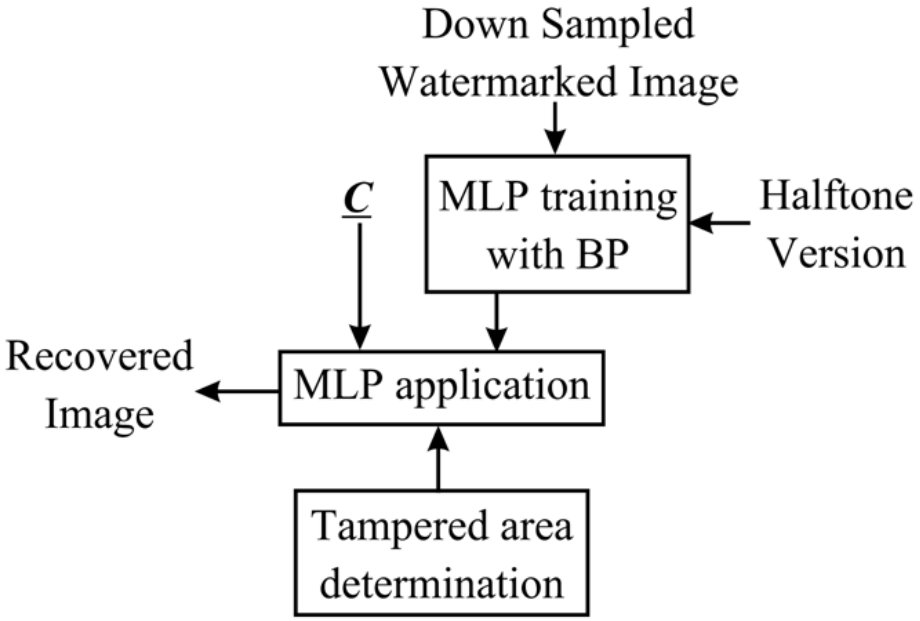
Aplicando siguiente criterio de diferencia a cada bloque, obtiene bloques alterados.

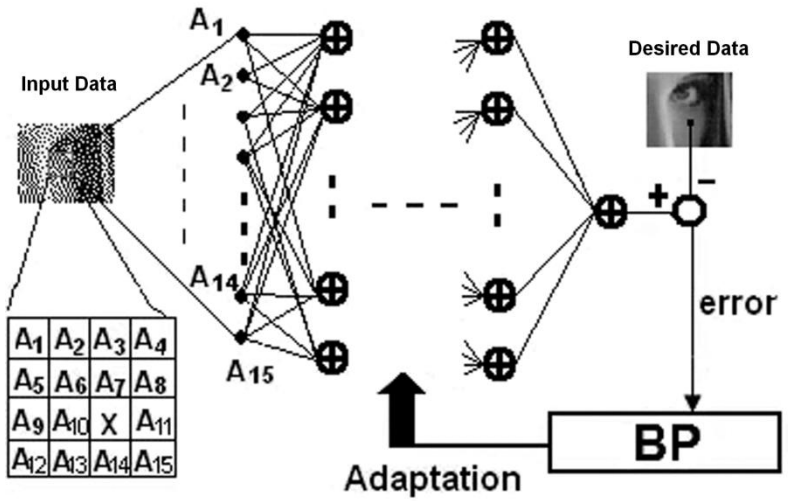
$$\begin{cases} SSIM_k(A_k, B_k) \le Th & k-th \text{ block is tampered} \\ SSIM_k(A_k, B_k) > Th & k-th \text{ block is authentic} \end{cases}$$

$$SSIM_k(A_k, B_k) = [l(A_k, B_k)]^{\alpha}[c(A_k, B_k)]^{\beta}[s(A_k, B_k)]^{\gamma} \quad k = 1...K$$



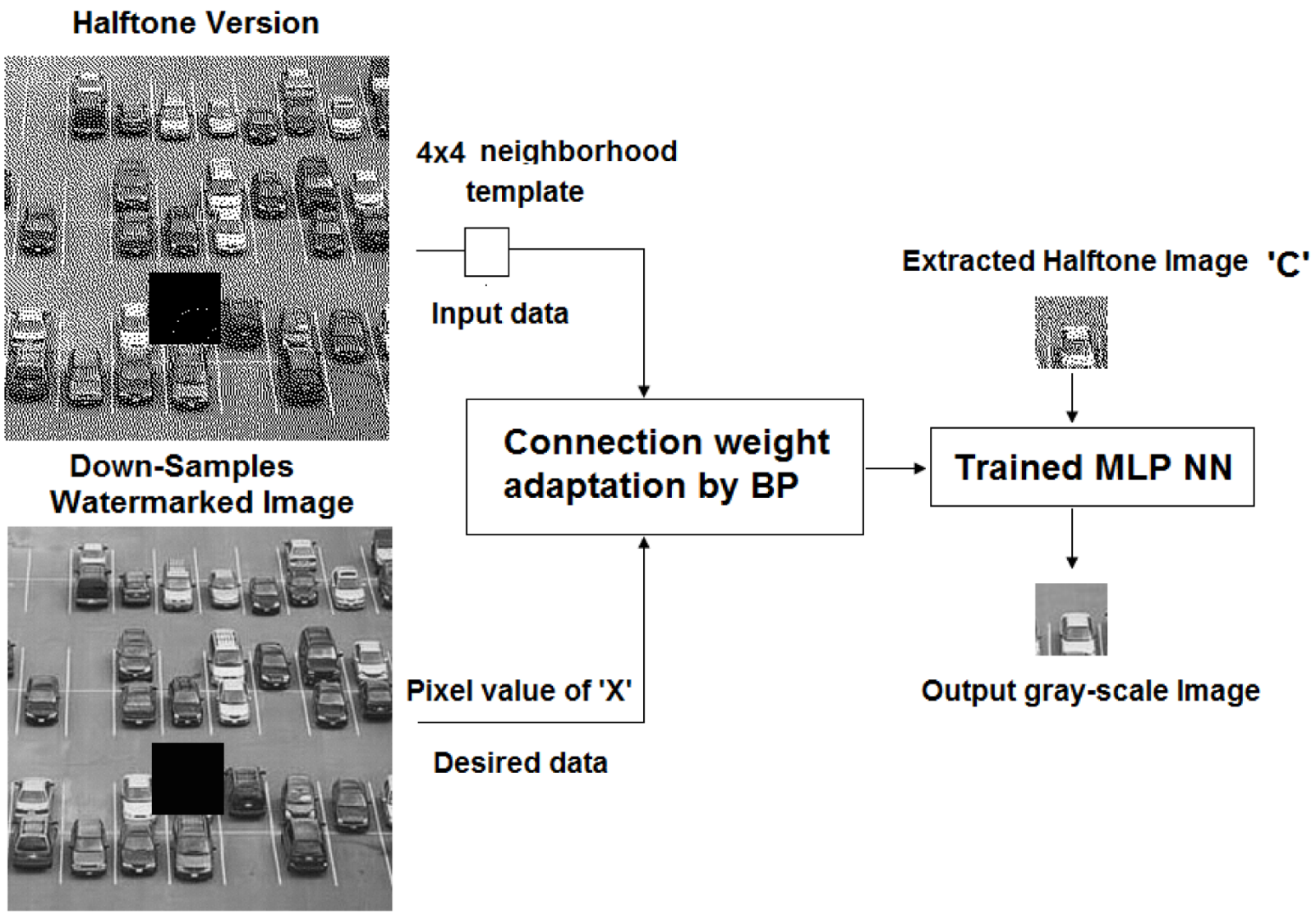
Recuperación





Perceptron Multicapas (MLP)

Recuperación



Resultados: imperceptibilidad

Algorithms / *S*	6	7	8	9	10	11	12	13	14
WIA-IWT	39.58	38.25	37.10	36.13	35.18	34.40	33.60	32.98	32.23
WIA-DCT	41.54	40.12	38.87	37.80	36.73	35.81	35.03	34.36	33.74

Original

Imagen marcada PSNR=37.1dB

f_a

Robustez

Compresión JPEG error de falsa alarma

	without attack			80			70			65			60		
Th	0.3	0.4	0.5	0.3	0.4	0.5	0.3	0.4	0.5	0.3	0.4	0.5	0.3	0.4	0.5
(%)	0	0	0	0	0	0	0	0	0	0	0.02	0.09	0.03	0.4	4.6

Error de falsa negativa

Th	**0.3**	**0.4**	**0.5**
(%)	0.0026	4.76×10^{-5}	9.52×10^{-6}

Resultados

(1) **Imagen Original**

(2) **Imagen Alterada**

(3) **Imagen con bloques alterados detectados**

(4) **Imagen Recuperada**

(1)

(2)

(3)

(4)

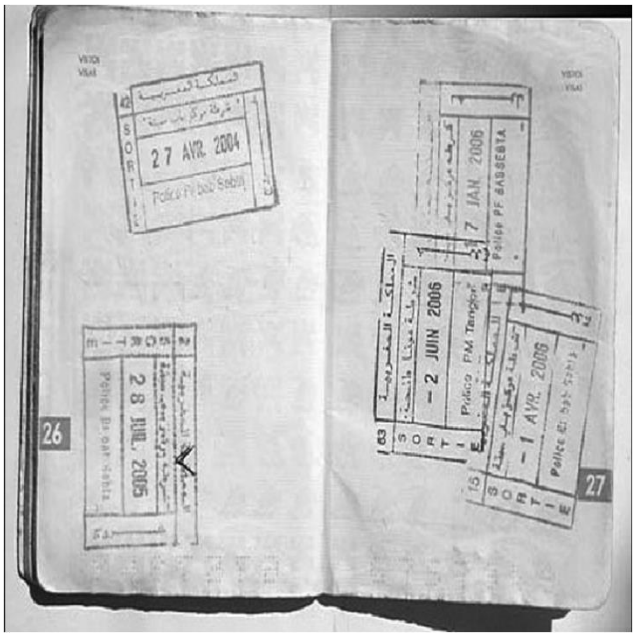
(1)

(2)

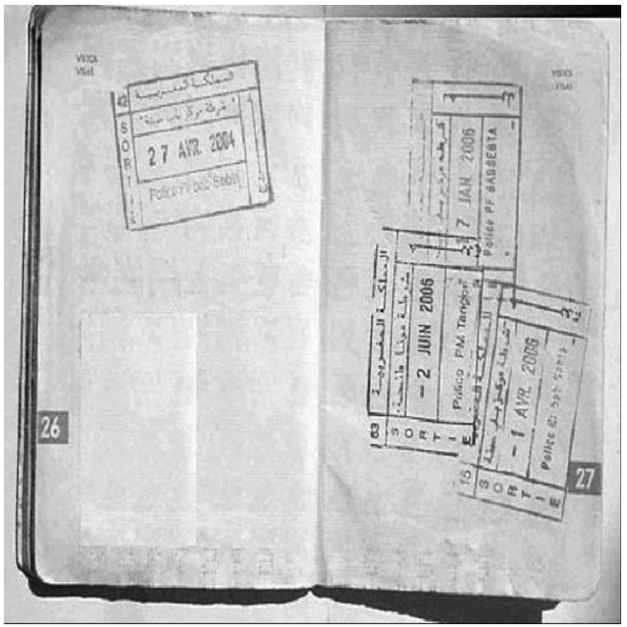
(3)

(4)

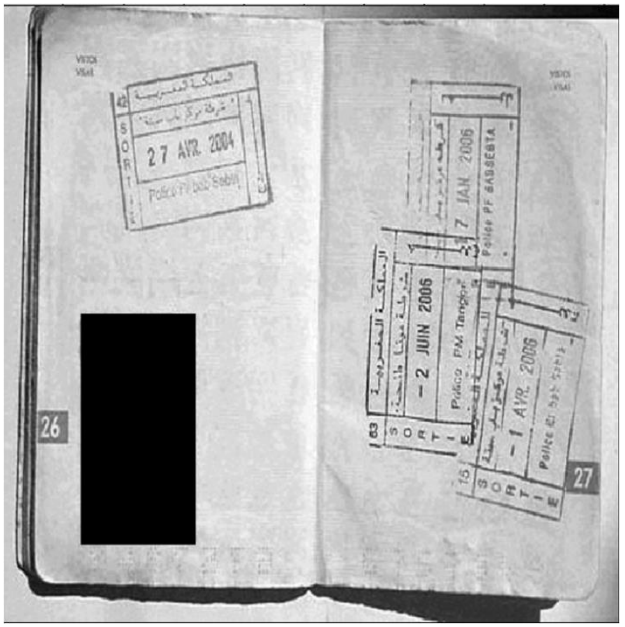
(1)


27 AVR. 2004
Police PJ Bab Sebta
28 JUIL. 2005
2 JUIN 2006
Police PM Tanger
7 JAN. 2006
Police PF SEBTA
1 AVR. 2006
26
27

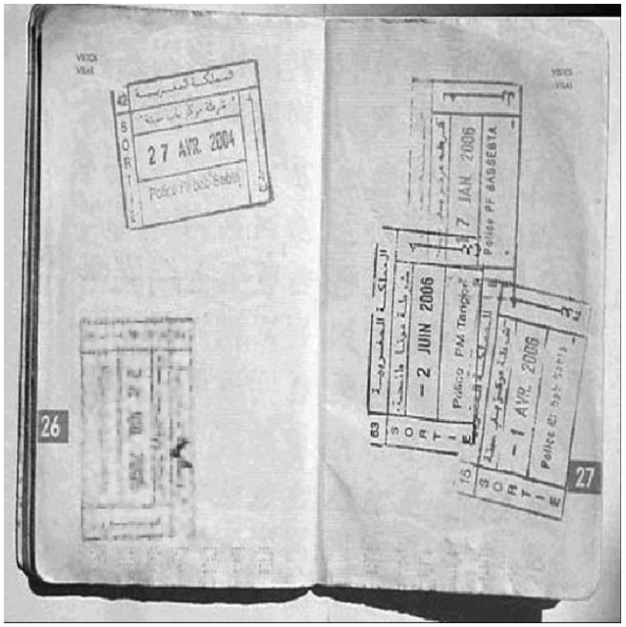
(2)


27 AVR. 2004
Police PJ Bab Sebta
2 JUIN 2006
Police PM Tanger
7 JAN. 2006
Police PF SEBTA
1 AVR. 2006
26
27

(3)


27 AVR. 2004
Police PJ Bab Sebta
2 JUIN 2006
Police PM Tanger
7 JAN. 2006
Police PF SEBTA
1 AVR. 2006
26
27

(4)


27 AVR. 2004
Police PJ Bab Sebta
2 JUIN 2006
Police PM Tanger
7 JAN. 2006
Police PF SEBTA
1 AVR. 2006
26
27

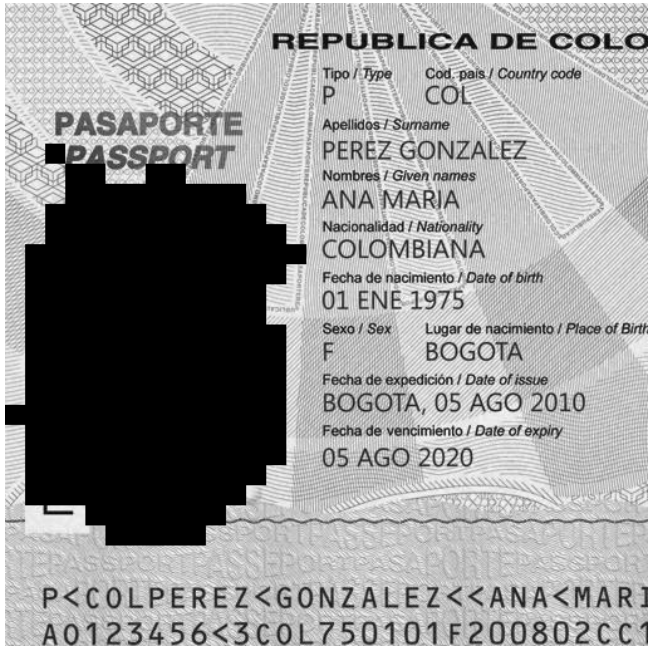
(1)


REPUBLICA DE COLO
PASAPORTE
PASSPORT
Tipo / Type
P
Cod. pais / Country code
COL
Apellidos / Surname
PEREZ GONZALEZ
Nombres / Given names
ANA MARIA
Nacionalidad / Nationality
COLOMBIANA
Fecha de nacimiento / Date of birth
01 ENE 1975
Sexo / Sex
F
Lugar de nacimiento / Place of Birth
BOGOTA
Fecha de expedición / Date of issue
BOGOTA, 05 AGO 2010
Fecha de vencimiento / Date of expiry
05 AGO 2020
P<COLPEREZ<GONZALEZ<<ANA<MARI
A0123456<3COL750101F200802CC1

(2)


REPUBLICA DE COLO
PASAPORTE
PASSPORT
Tipo / Type
P
Cod. pais / Country code
COL
Apellidos / Surname
PEREZ GONZALEZ
Nombres / Given names
ANA MARIA
Nacionalidad / Nationality
COLOMBIANA
Fecha de nacimiento / Date of birth
01 ENE 1975
Sexo / Sex
F
Lugar de nacimiento / Place of Birth
BOGOTA
Fecha de expedición / Date of issue
BOGOTA, 05 AGO 2010
Fecha de vencimiento / Date of expiry
05 AGO 2020
P<COLPEREZ<GONZALEZ<<ANA<MARI
A0123456<3COL750101F200802CC1

(3)


REPUBLICA DE COLO
PASAPORTE
PASSPORT
Tipo / Type
P
Cod. pais / Country code
COL
Apellidos / Surname
PEREZ GONZALEZ
Nombres / Given names
ANA MARIA
Nacionalidad / Nationality
COLOMBIANA
Fecha de nacimiento / Date of birth
01 ENE 1975
Sexo / Sex
F
Lugar de nacimiento / Place of Birth
BOGOTA
Fecha de expedición / Date of issue
BOGOTA, 05 AGO 2010
Fecha de vencimiento / Date of expiry
05 AGO 2020
P<COLPEREZ<GONZALEZ<<ANA<MARI
A0123456<3COL750101F200802CC1

(4)


REPUBLICA DE COLO
PASAPORTE
PASSPORT
Tipo / Type
P
Cod. pais / Country code
COL
Apellidos / Surname
PEREZ GONZALEZ
Nombres / Given names
ANA MARIA
Nacionalidad / Nationality
COLOMBIANA
Fecha de nacimiento / Date of birth
01 ENE 1975
Sexo / Sex
F
Lugar de nacimiento / Place of Birth
BOGOTA
Fecha de expedición / Date of issue
BOGOTA, 05 AGO 2010
Fecha de vencimiento / Date of expiry
05 AGO 2020
P<COLPEREZ<GONZALEZ<<ANA<MARI
A0123456<3COL750101F200802CC1

(1)

(2)

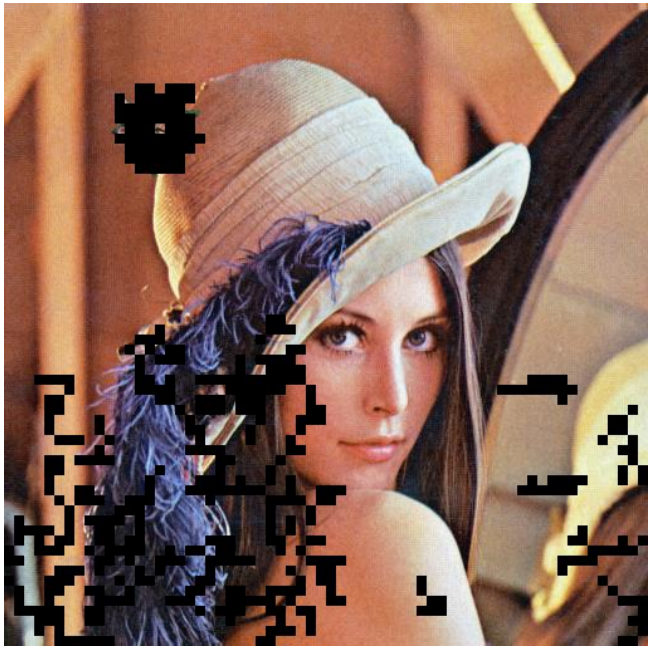
(3)

(4)

(1)

(2)

(3)



(4)

(1)

(2)

(3)

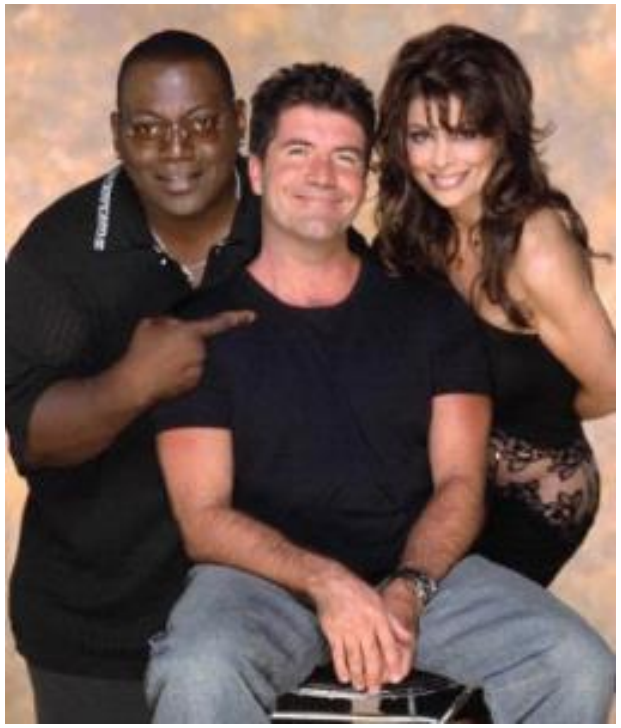
(4)

Propuesta 2 :Marca de agua semi-frágil usando compresión SPIHT y detección de rostro

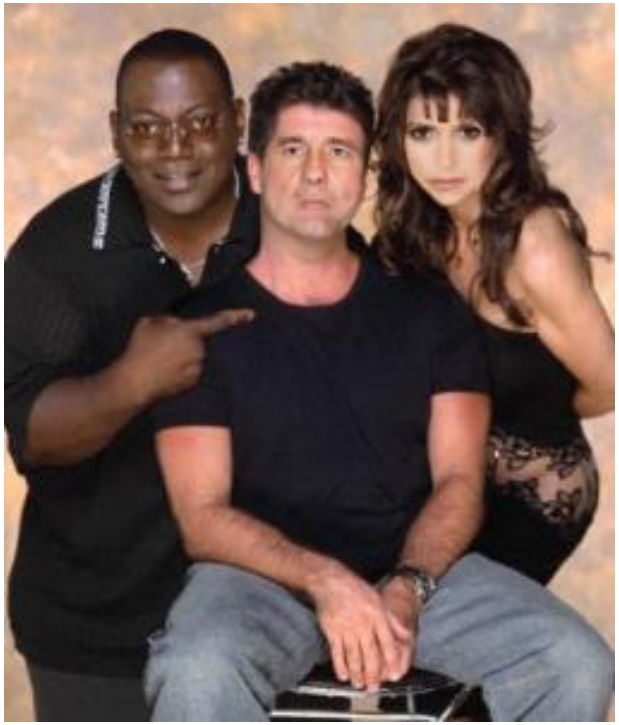
Características

1. **Introducción de concepto ROI (región de Interés)**
2. **Regiones de rostros se consideran como ROI**
3. **Uso de método de compresión SPIHT**
4. **Recuperación de regiones alteradas con mayor calidad**
5. **Robusto a los procesos que conservan el contenido. (compresión, ruido aditivo)**
6. **Inserción y extracción de marca de agua se realzan en el dominio de frecuencia.**

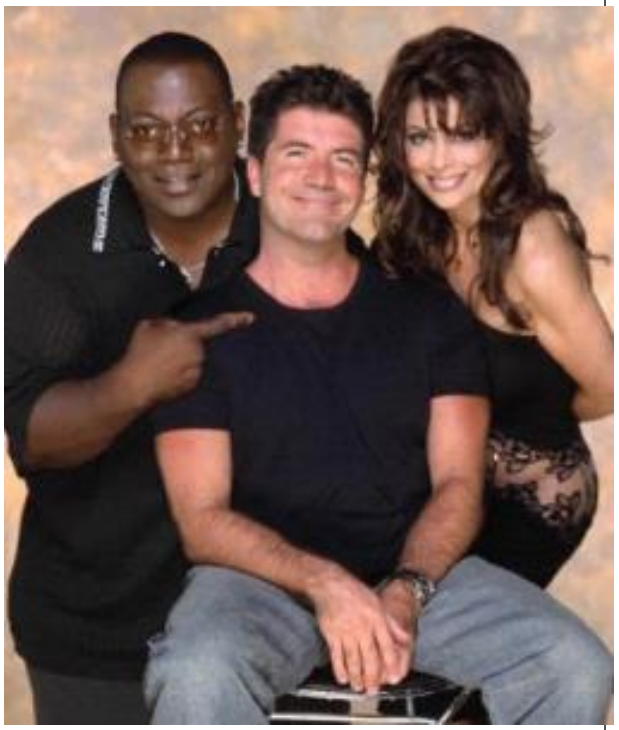
Ejemplos



Original



Alterada



Recuperada

Ideas principales

- Regiones de rostros de fotografías digitales son más alteradas.

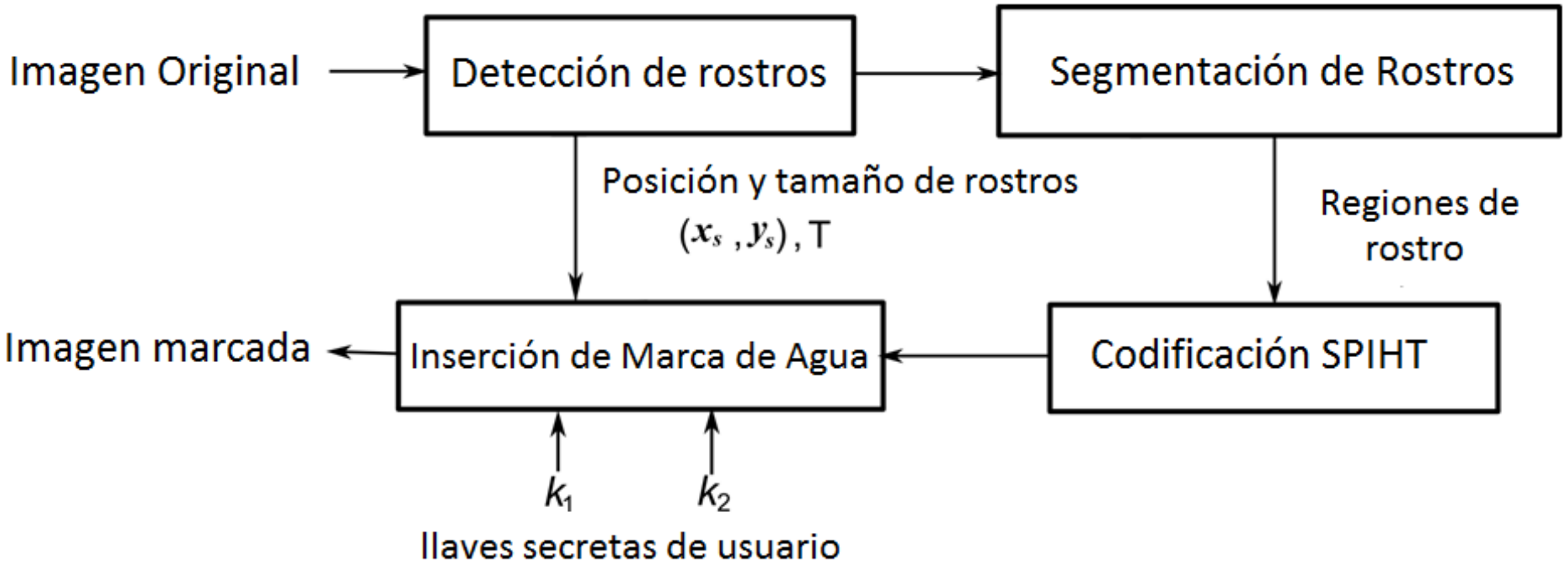
Regiones de Interés (ROI) = Regiones de rostros

- Detectar automáticamente regiones de rostro usando algoritmo de Viola & Jones.
- Codificar regiones de rostro detectadas usando algoritmo de compresión de imágenes SPHIT (Set Pertitioning in Hierarchical Trees)
- Datos comprimidos por el SPHIT se inserta dentro de Regiones de fondo (ROB).

Etapa de protección

Etapa de autenticación & recuperación

Etapa de protección

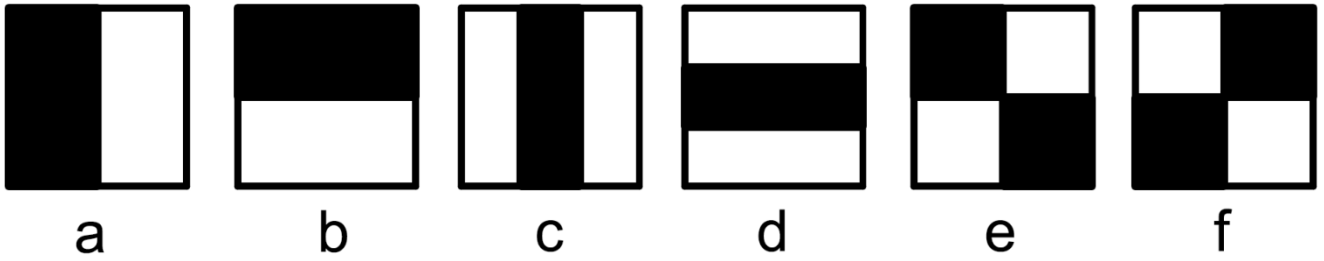


Detección de Rostros
Algoritmo de Viola & Jones

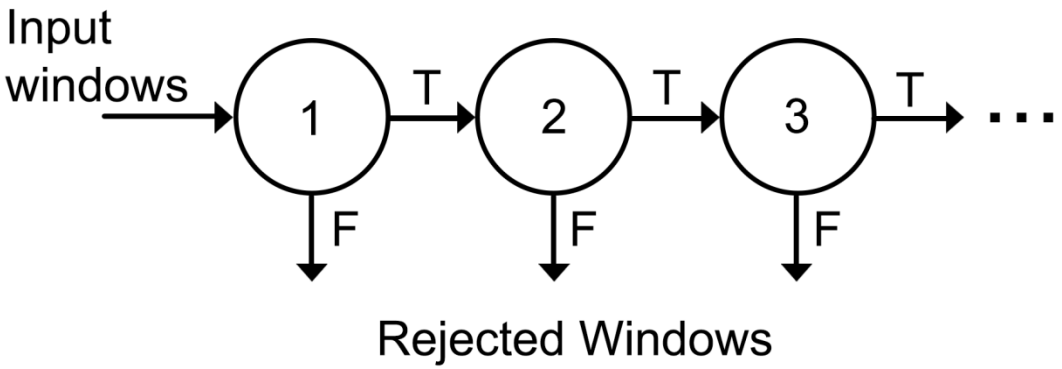
- **Bajo costo computacional**
- **Alta tasa de detección de rostro (aprox. 70%)**
- Solo detecta rostro frontal. No detecta rostro lateral
- Sensible a cambio de iluminación
- No detecta rostros con siguientes condiciones
  - Rostros con baja contraste
  - Rostros rotados

Algoritmo de Viola & Jones

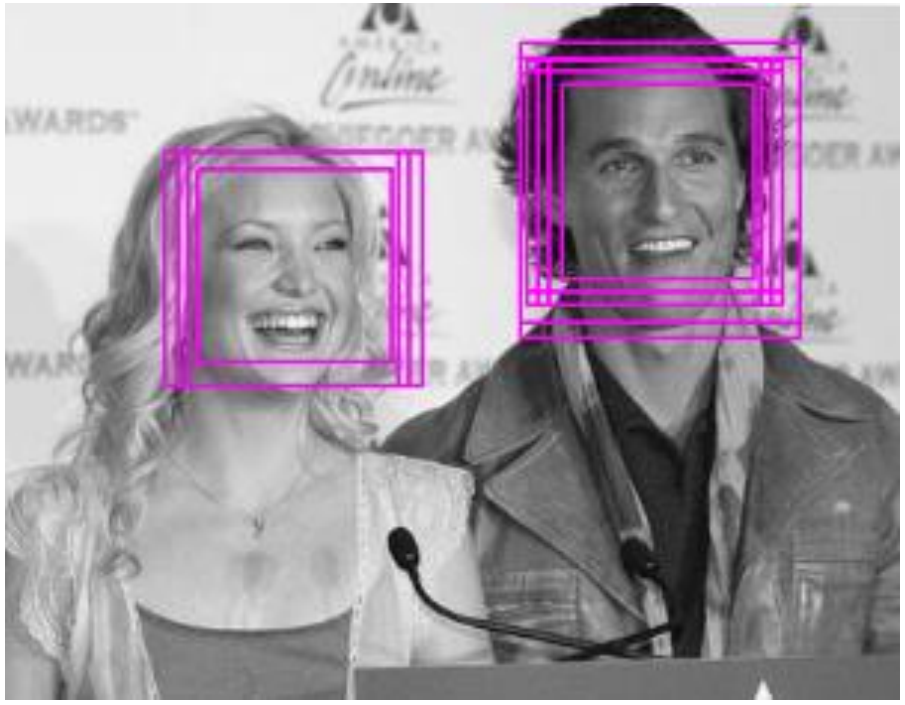
- Usando patrones sencillos, se realiza clasificación



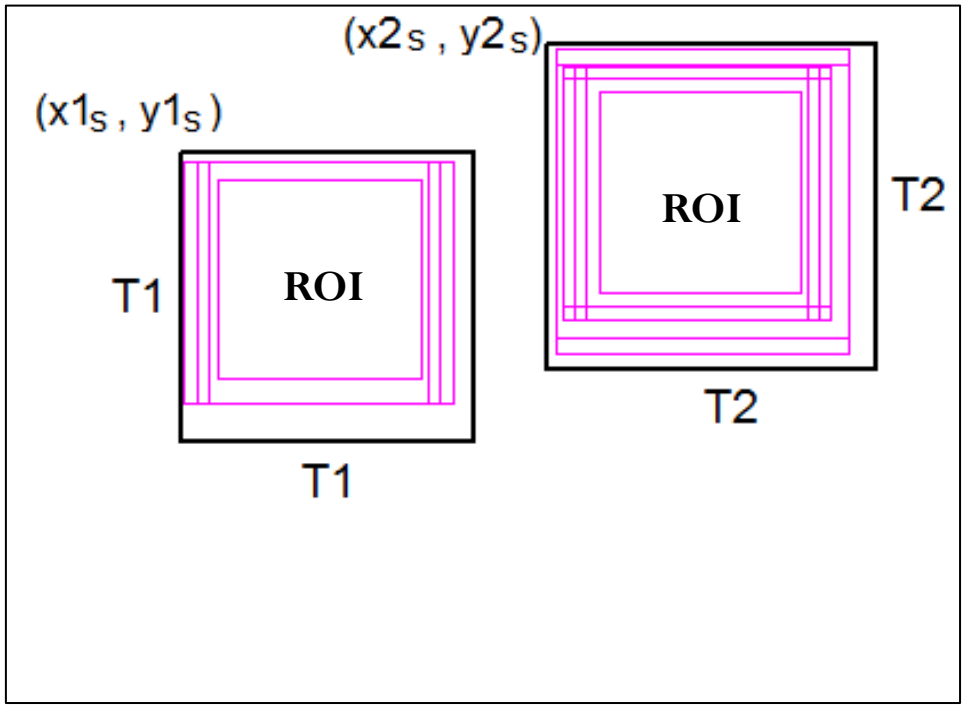
- Toma de decisión en forma cascada



Ajuste de ROI



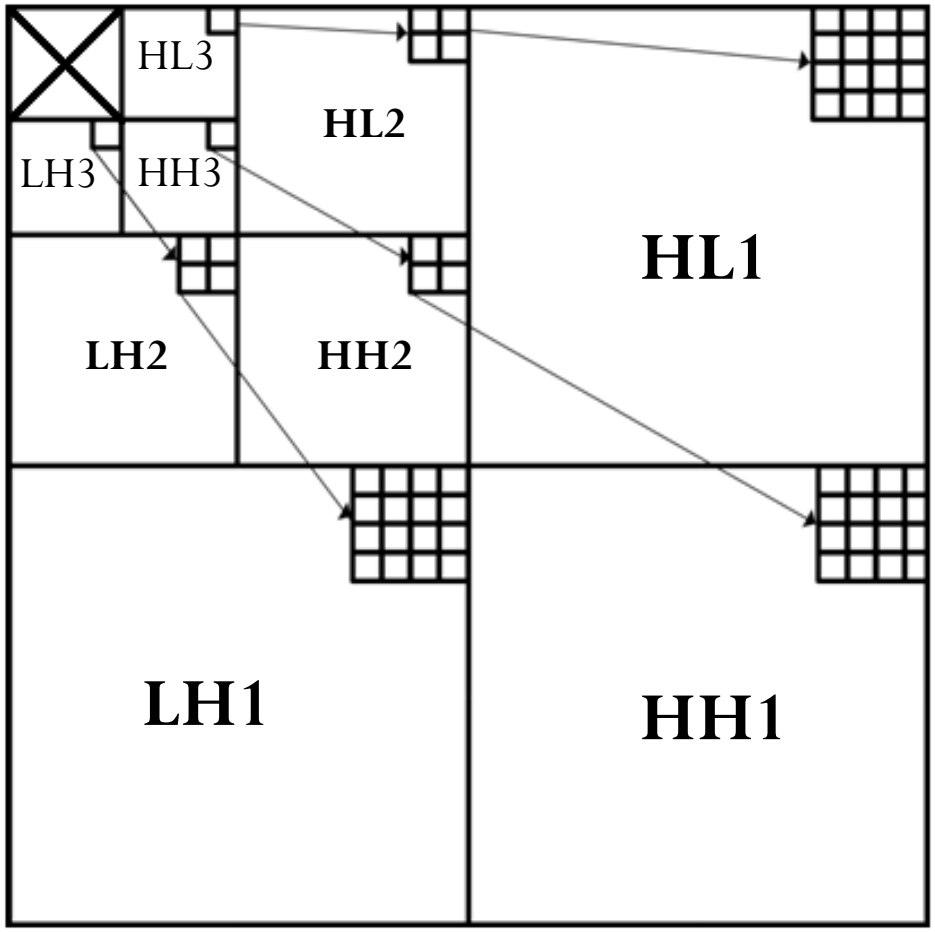
Detección de regiones de rostro
Varios candidatos



Ajuste de ROIs

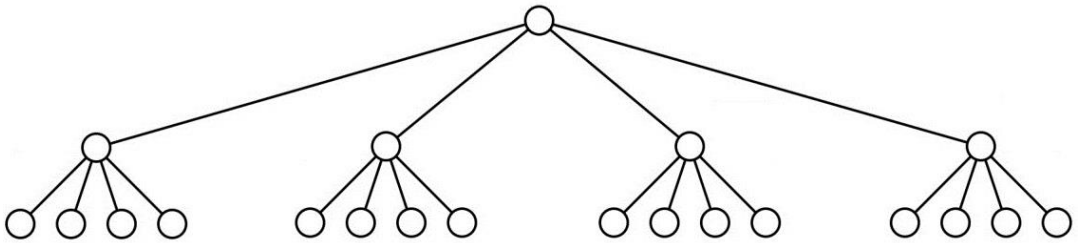
Codificación de ROIs usando SPIHT

- Set Partitioing in Hierarchical Trees (SPIHT) es un algoritmo de compresión de imágenes basado en Transformada Discreta de Wavelets (DWT)
- SPIHT aprovecha fuerte relación entre sub-bandas de diferentes niveles de DWT.
- Mejor calidad de imagen decodificada comparando con el algoritmo de JPEG.
- Puede controlar el número de bits de secuencia codificada.

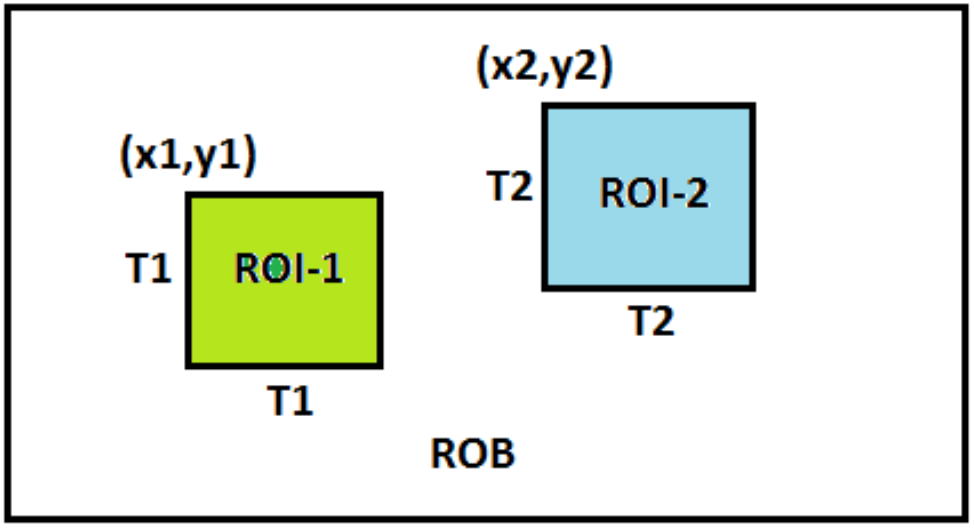


Descomposición de Wavelet

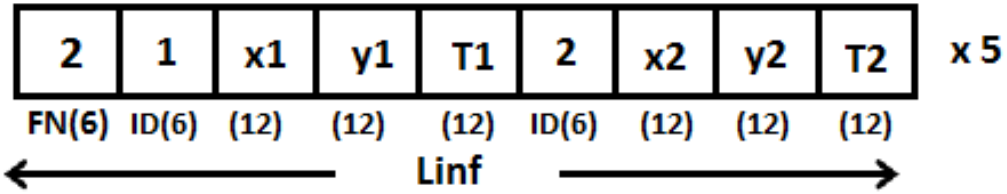
Quadotree



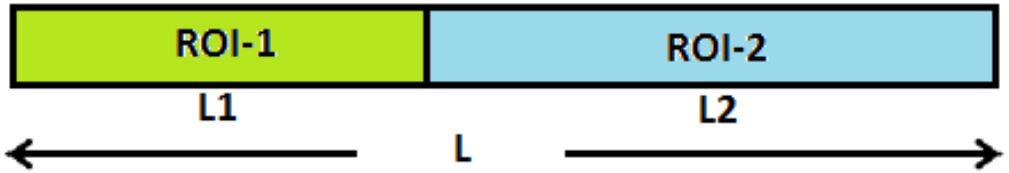
Secuencia de marca de agua



Secuencia de bits (información de ROI)



Secuencia de bits (ROIs codificados)



$$L = N_b(ROB) - Linf \times 5$$

$$L = L1 + L2$$

$$L1 : L2 = T1 : T2$$

$$L1 = \frac{T1 \times L}{T1 + T2} \qquad L2 = \frac{T2 \times \boldsymbol{L}}{T1 + T2}$$

Inserción de marca de agua

1. Dividir ROB en bloques de 8x8 pixeles .
2. Aplica DCT-2D a cada bloque ROB.
3. Dos secuencias de marca de agua se inserta en los coeficientes de baja frecuencias.

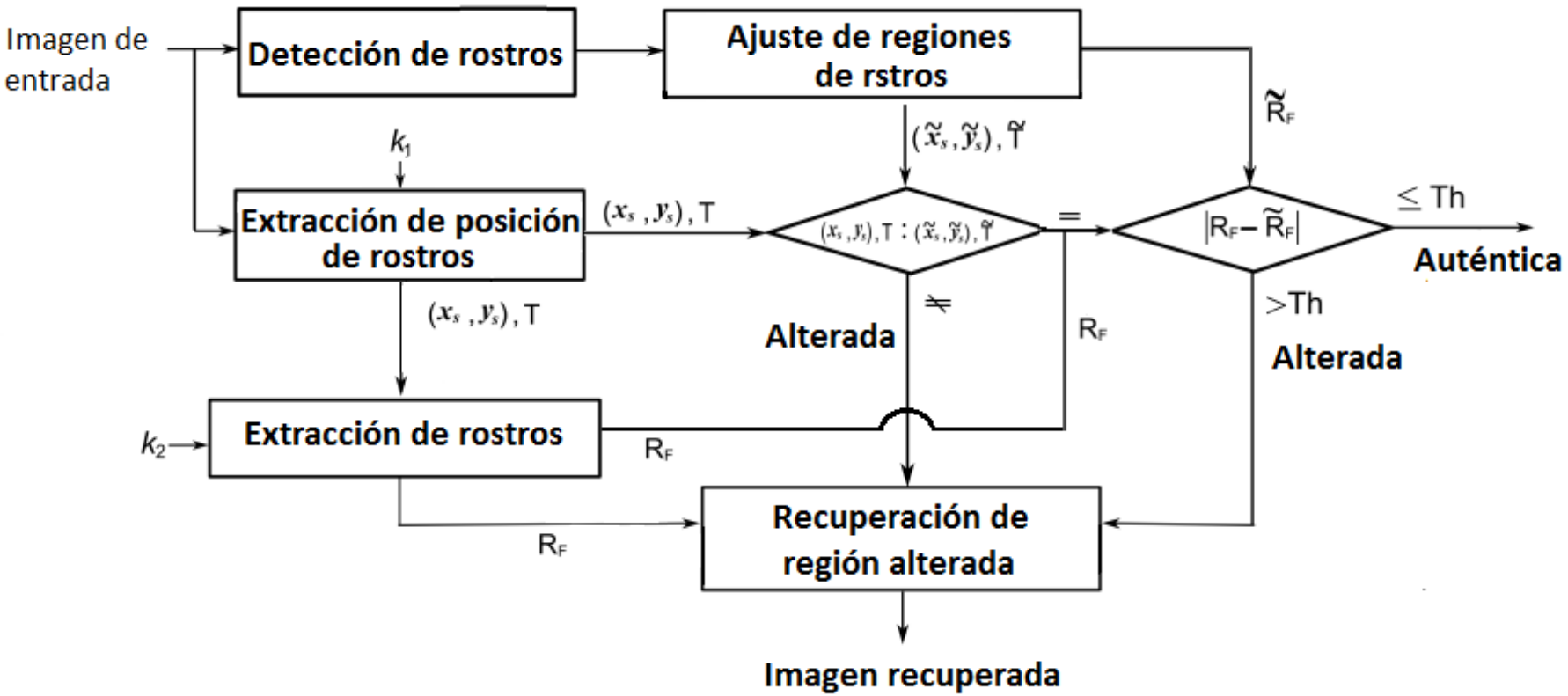
Coeficientes de baja frecuencias

	1	5	6				
2	4	7	12				
3	8	11					
9	10						

$$cw_n = Q(c_n + d(n, m_n), \Delta) - d(n, m_n), \qquad n = 1,2, \dots, L$$

$$d(n,1) = \begin{cases} d(n,0) + \frac{\Delta}{2}, & d(n,0) < 0 \\ d(n,0) - \frac{\Delta}{2}, & d(n,0) > 0 \end{cases}, \qquad Q(x,\Delta) = round(\frac{x}{\Delta})\Delta$$

Etapa de Autenticación y Recuperación



Extracción de secuencia de marca de agua

Secuencia de marca de agua

- **Información de ROIs**

  (posición y tamaño)

- **Rostros codificados**

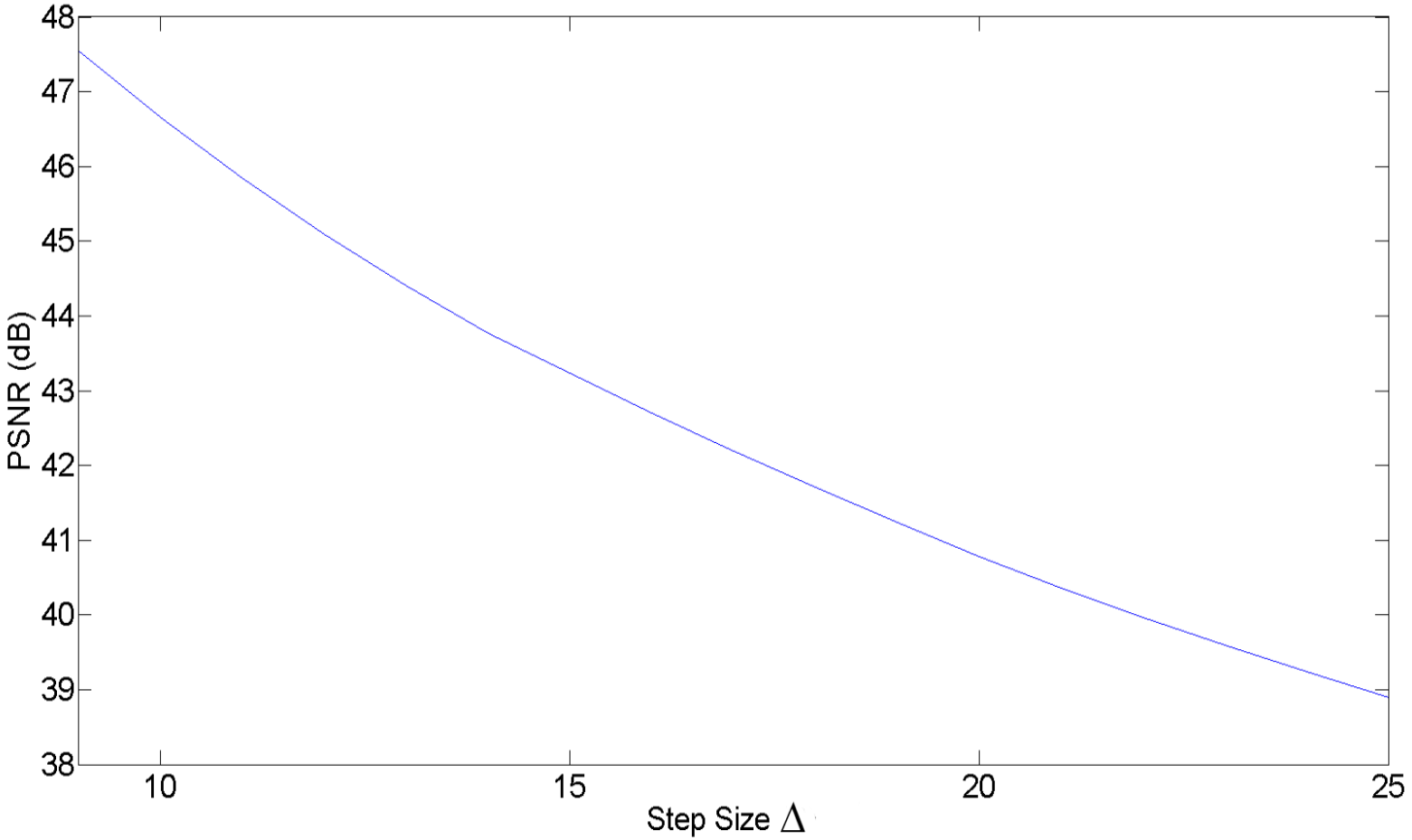
$$w_n = \underset{l \in 0,1}{\arg\min} (r_n - S_r(n,l))^2$$

$$S_r(n,0) = Q(r_n + d(n,0), \Delta) - d(n,0)$$

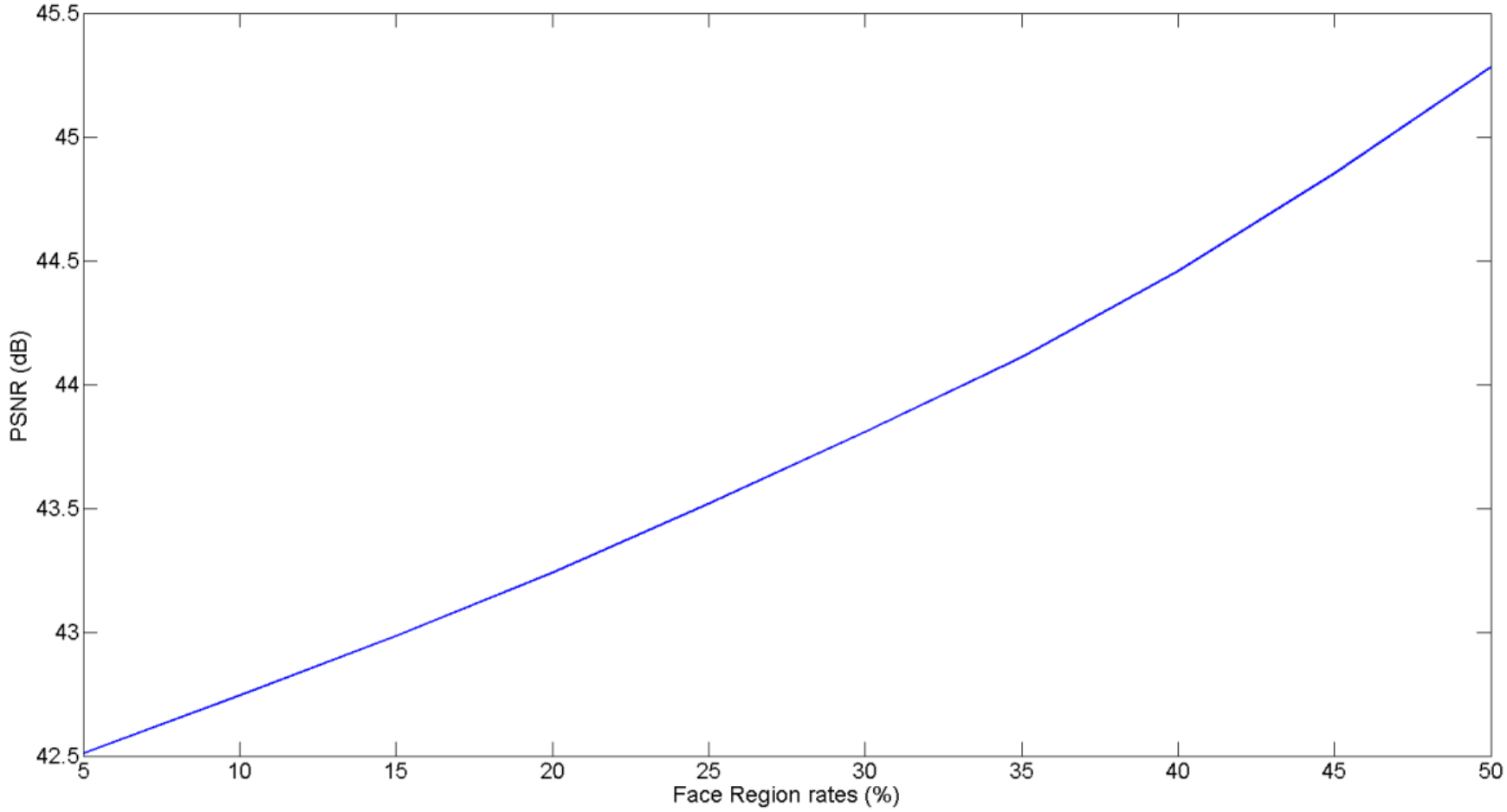
$$S_r(n,1) = Q(r_n + d(n,1), \Delta) - d(n,0)$$

r_n: n-ésimo coeficiente de DCT

Imperceptibilidad de marca de agua
PSNR: Tamaño de Paso Δ



Imperceptibilidad de marca de agua
PSNR: Tamaño de Rostro



46.69 dB

46.71 dB

Original

Marcada

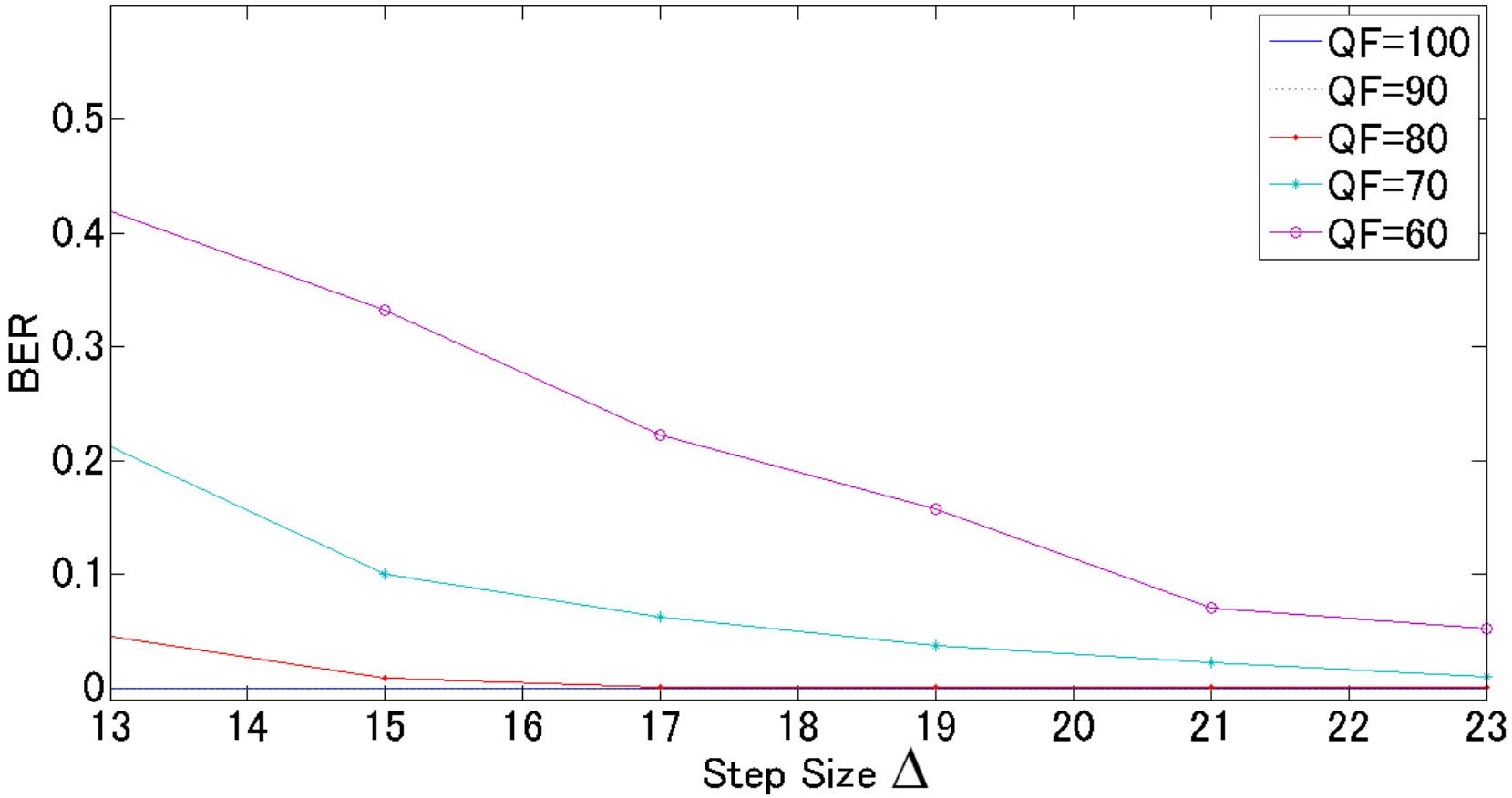
Original

Marcada: 45.55 dB

Original

Marcada: 45.55 dB

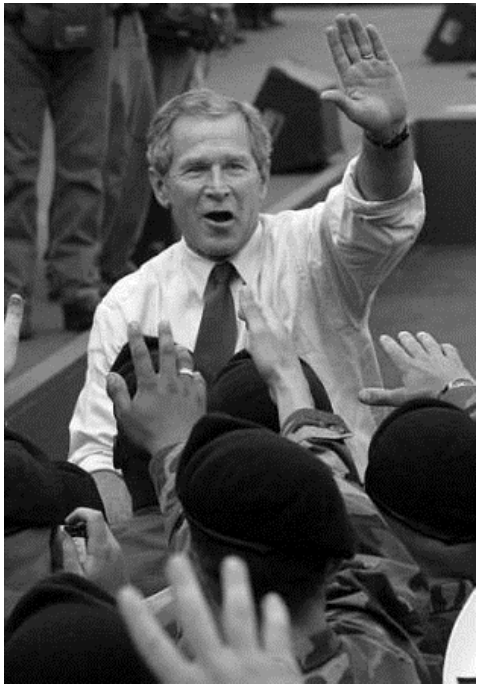
Robustez contra compresión JPEG



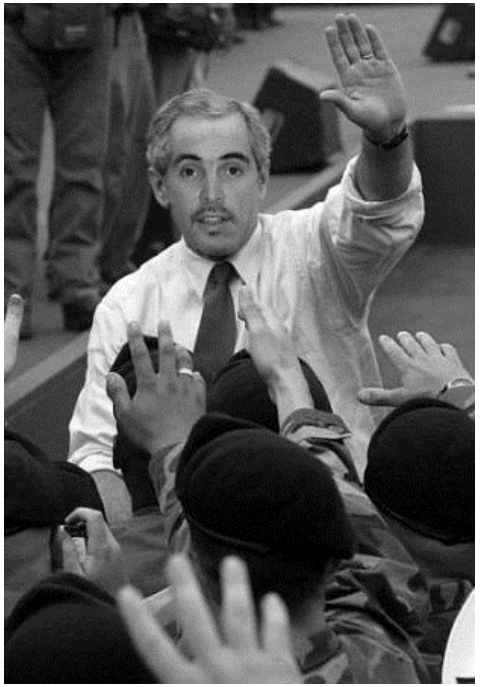
Calidad de imágenes recuperadas

Tamaño de rostro (%)	**<5%**	**5-15%**	**15-25%**	**>25%**
Imagen completa	55.09dB	45.60dB	42.92dB	32.84dB
Región de rostros	42.78dB	38.83dB	35.45dB	28.15dB

Ejemplos



Original



Alterada

Recuperada

PSNR=51.30 dB

Original

Alterada

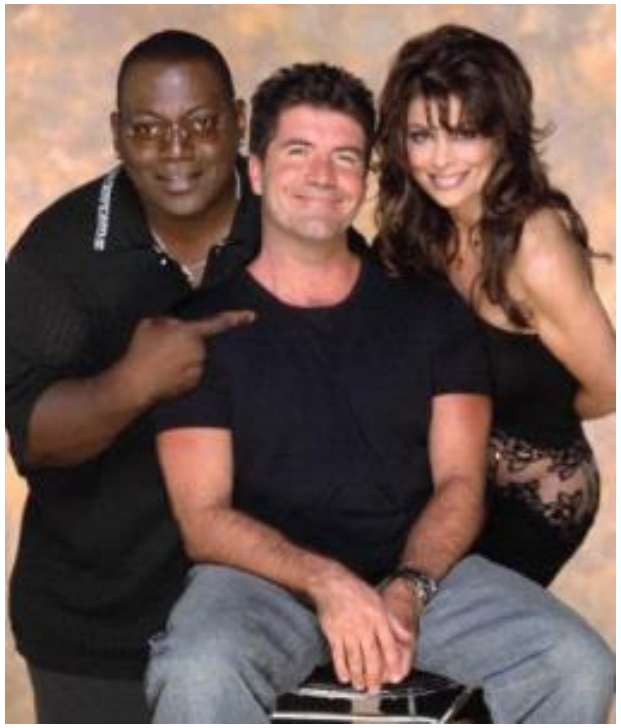
Recuperada

PSNR=71 dB

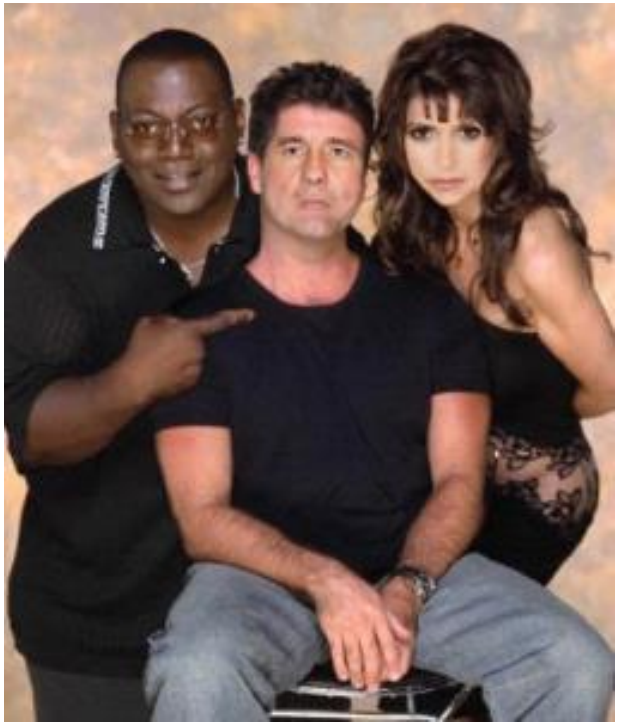
Original

Alterada

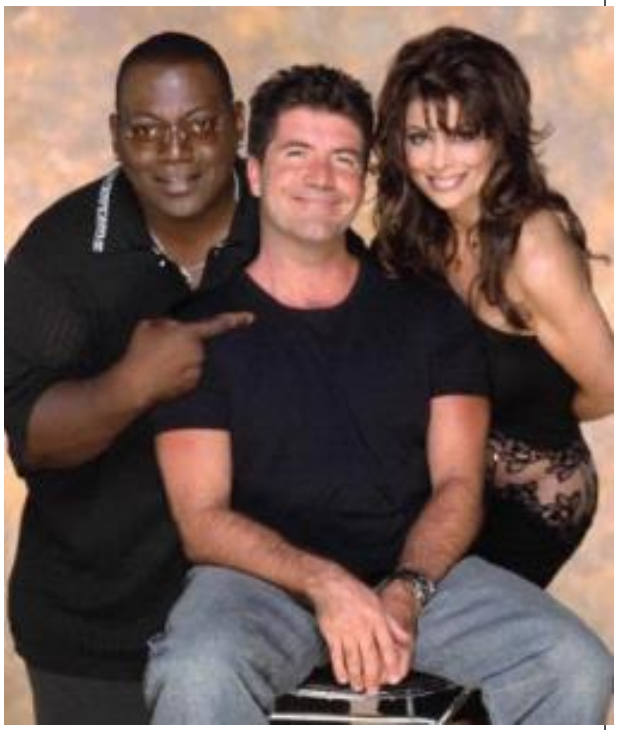
Recuperada
PSNR=43.93 dB



Original



Alterada



Recuperada
PSNR=47.09dB

Original

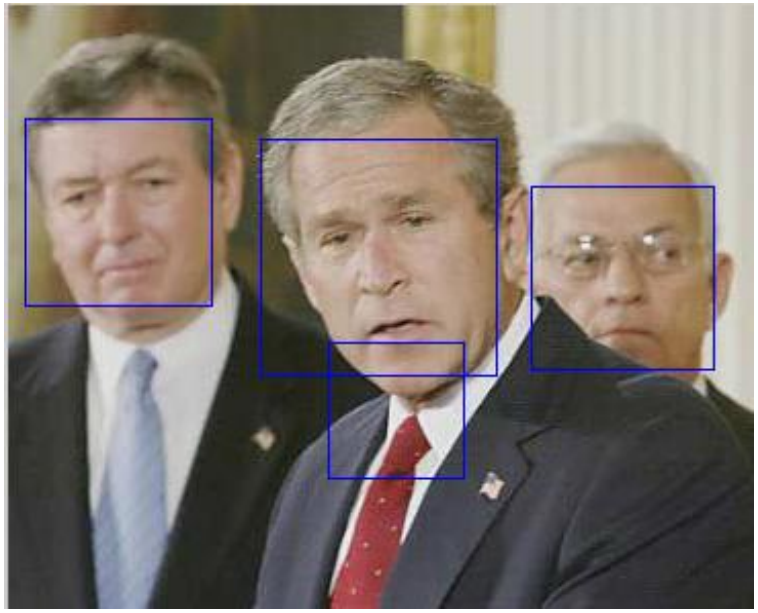
Alterada

Recuperada
PSNR=41.87 dB

Limitaciones

- Limitación de algoritmo de detección de rostro

(Algoritmo de Viola & Jones)



Limitaciones

- Tamaño de rostro es demasiado grande que el algoritmo no detecta rostro.

- Tamaño de rostro es demasiado grande que no se puede codificar.

..... Gracias .....