Protección de imágenes Watermark

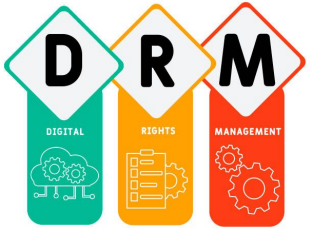
Kenny Alejandro, Lijie Yin, Javier Abella

Contenido

I.Técnicas de protección

Marcas de agua

Encriptación



Digital Rights Management

Técnicas de seguimiento

Herramientas legales

II. Watermark

¿Qué es?

- Consiste en la inserción de marcas o señales dentro de una imagen con el objetivo de protegerla o certificar su autenticidad.
- Puede ser **visible** o **invisible**.

Usos:

- Protección de derechos del autor
- Asegurar la integridad
- Rastreo de distribución
- Branding y publicidad

Visible

Invisible

A. Watermarking visible

Características:

- Perceptibles a simple vista
- Formas comunes: logos de la marca, textos de copyright, nombre de la marca o compañía, firma personal, etc.
- Debe identificar claramente al autor o propietario del contenido.

Ventajas:

- Disuasión directa
- Compatible con múltiples formatos (imagen, video)
- Marketing visual

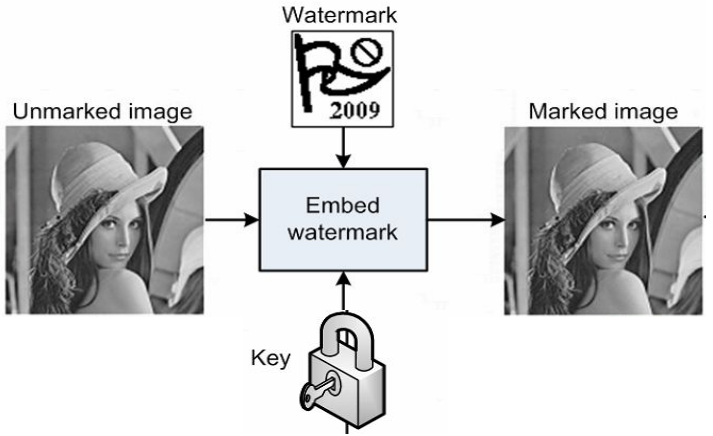
Desventajas:

- Fáciles de borrar
- Fáciles de modificar, si se encuentran en las esquinas
- Afectan la estética

B. Watermarking invisible

Características:

- Se integran/ocultan directamente en la imagen
- Son difíciles de detectar o eliminar, requiere conocer algoritmo y/o clave utilizada
- Pueden resistir transformaciones como recortes, compresión o redimensionado.
- Cualquier intento de borrarlas puede degradar la calidad del archivo, dejando evidencia de manipulación.



Ventajas:

- No afecta la apariencia visual
- Difícil de eliminar sin permiso
- Resistente manipulación

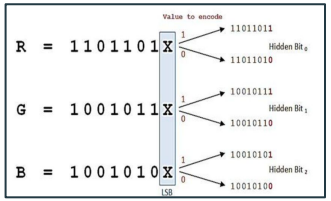
Desventajas:

- Complejidad técnica
- Difícil de comprobar a simple vista
- Posible degradación al eliminarla

C. Clasificación de técnicas watermarking invisible

1. Dominio espacial

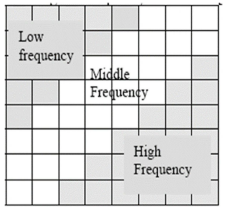
- Incrustación directa en los píxeles
- Ejemplo: **LSB**



Bit menos significativo

2. Dominio de frecuencia

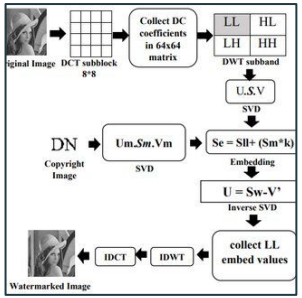
- Se incrusta la marca en los coeficientes de frecuencia (menos perceptibles al ojo humano)
- Ejemplo: **DCT**, **DWT**, **FFT**



Coefficient matrix of DCT

3. Híbridas

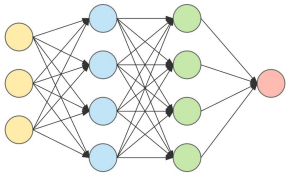
- Combinación de transformadas
- Ejemplo:
  DWT-DCT,
  DWT-SVD,
  DCT-SVD,
  DWT-DCT-SVD



DWT-DCT-SVD

4. Basadas en IA

- Uso de redes neuronales (CNN, GAN) para incrustación y extracción.

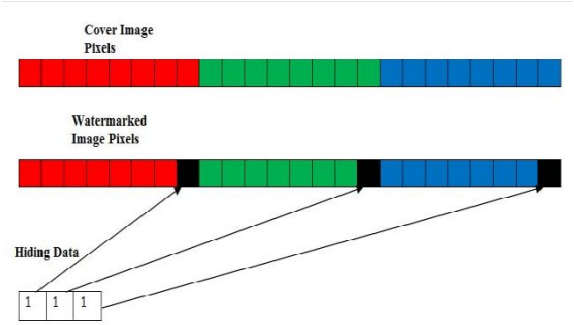


Redes convolucionales

1. Dominio espacial:

Algoritmo LSB
(*Less Significant bit*)

- **¿Qué hace?** Inserta la marca de agua modificando el bit menos significativo de cada píxel de la imagen.
- **¿Para qué sirve?**: Cada píxel de una imagen está compuesto por valores de color en RGB, y el bit menos significativo se puede alterar sin causar cambios perceptibles a simple vista.
- **Ventaja**: Fácil y rápido de implementar.
- **Desventaja**: Muy vulnerable a compresión, filtrado, o cualquier modificación de la imagen.



Se puede pensar en LSB como escribir un mensaje secreto en la arena de la playa, justo al borde del agua. Si nadie lo toca, el mensaje sigue ahí. Pero si viene una ola (como la compresión o el ruido), se borra fácilmente.

2. Dominio frecuencia:

Algoritmo DCT

(*Discrete Cosine Transform*)

- **¿Qué hace?** Convierte una imagen del dominio espacial (píxeles) al dominio de frecuencia, representando la señal como una suma de funciones coseno con diferentes frecuencias.
- **¿Para qué sirve?**: Permite insertar la marca de agua en los coeficientes de frecuencia media, que son menos perceptibles al ojo humano y más resistentes a la compresión.
- **Ventajas**: Los cambios con poco perceptibles, y que ya es muy usado y probado en compresión (ej. JPEG).
- **Desventajas**: Menos resistente a rotaciones ó escalados de imagen.

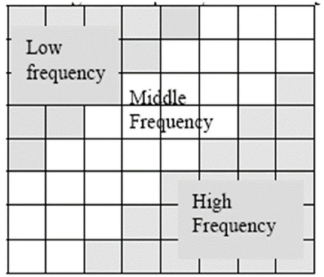
Se puede pensar DCT como un filtro que separa la imagen en distintos tipos de patrones, desde superficies suaves hasta los bordes más marcados.

$$X_{u,v} = \alpha(u)\alpha(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)\cdot\cos\left[\frac{\pi}{N}(x+0.5)u\right]\cdot\cos\left[\frac{\pi}{N}(y+0.5)v\right]$$

Donde:

- $f(x,y)$ es el valor del píxel en la posición (x,y).
- $X_{u,v}$ es el coeficiente de frecuencia para esa combinación de cosenos.
- $\alpha(u)$ y $\alpha(v)$ son factores de normalización:

$$\alpha(k) = \begin{cases} \sqrt{\frac{1}{N}} & \text{si } k = 0 \\ \sqrt{\frac{2}{N}} & \text{si } k > 0 \end{cases}$$



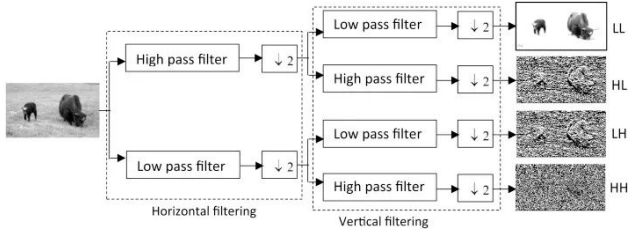
Coefficient matrix of DCT

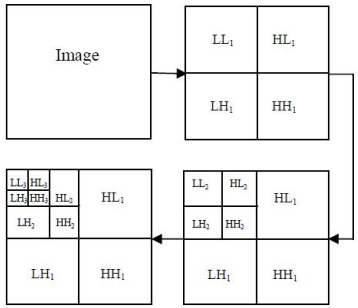
2. Dominio frecuencia:

Algoritmo DWT
(*Discrete Wavelet Transform*)

- **¿Qué hace?** Divide una imagen en **subbandas** de diferentes **resoluciones o escalas** (como si hicieras zoom).
- **Subbandas**:
  - Aproximación (LL): baja frecuencia, contiene la forma general.
  - Detalle (LH, HL, HH): alta frecuencia, contienen bordes y detalles.
- **¿Para qué sirve?**: Permite insertar la marca de agua en subbandas de detalle (como HL o HH), donde los cambios son menos visibles y más difíciles de eliminar.
- **Ventajas**: Permite aplicar DWT en múltiples niveles para tener mayor control sobre la “profundidad” de inserción
- **Desventajas**: Implementación más compleja, no es directamente compatible con formatos como JPEG

Piensa en la DWT como un zoom que muestra diferentes niveles de detalle; puedes esconder información en partes donde el ojo no detecta cambios fácilmente.





3. Híbridos:

Algoritmo SVD
(*Singular Value Decomposition*)

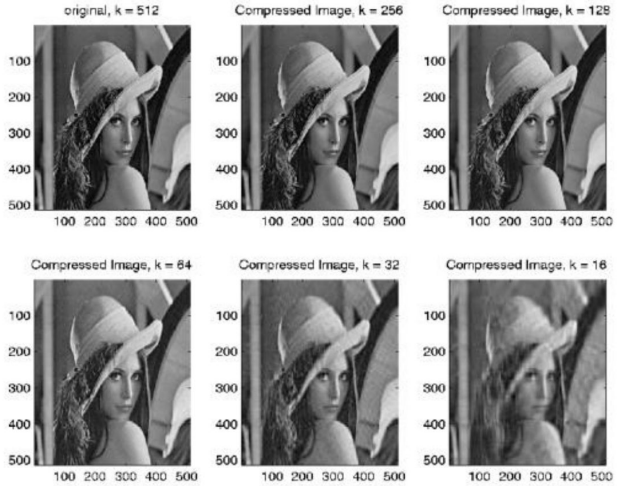
- **¿Que hace?**: Descompone la matriz (imagen) into 3 matrices:
  $A = U \times S \times V^t$, donde:
  - U y V^t son ortogonales(rotacion/estructura)
  - S contiene los valores singulares (características relevantes de la imagen)

- **¿Para qué sirve?**: Se modifican los valores de S (que no son dominantes) para insertar la marca de agua. Estos cambios son estables y no afectan mucho la calidad visual.

- **Ventajas**: Muy robusto, conserva la marca incluso con distorsiones. Es ideal para sistemas híbridos(DCT-SVD, DWT-SVD)

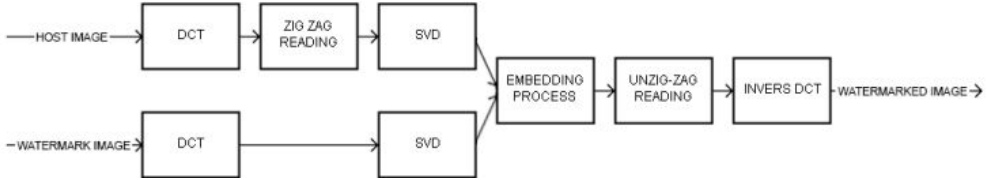
- **Desventajas**: Requiere un cómputo un poco elevado y por si sola no actúa sobre un dominio específico (requiere transformadas DCT o DWT)

Piensa en SVD como una forma de separar la "estructura interna" o core de la imagen, permitiéndole hacer pequeños ajustes sin que se note visualmente.

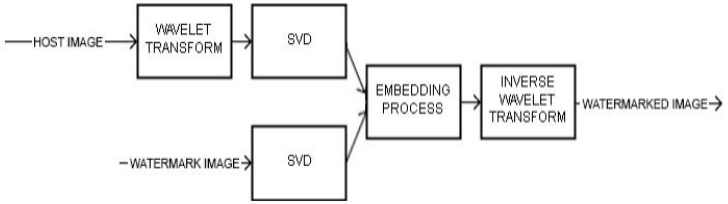


3. Híbridos:

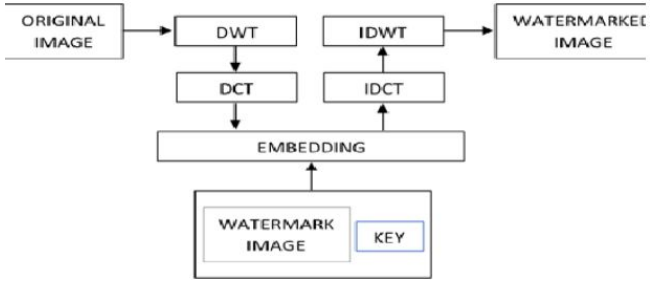
DCT-SVD



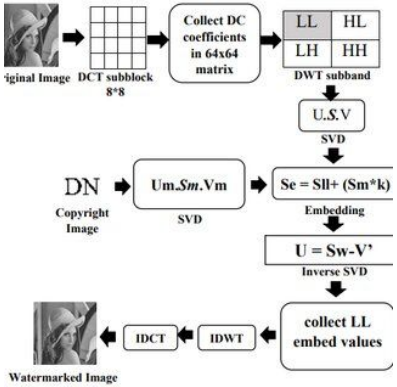
DWT- SVD



DCT- DWT



DCT-DWT-SVD

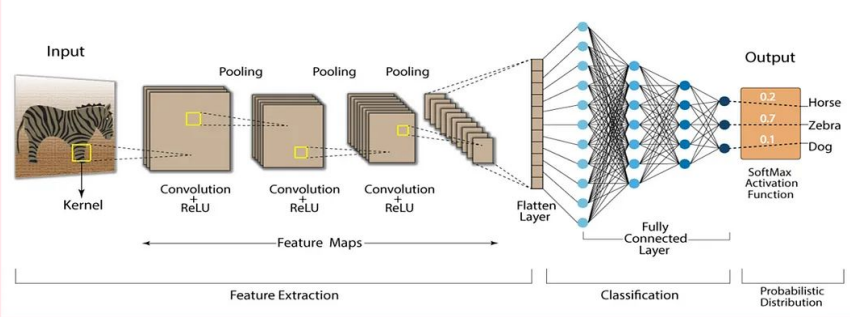


4. Basadas en IA:

CNNs

(Convolutional Neural Networks)

- **¿Qué hace?** Aprende automáticamente a incrustar y extraer marcas de agua invisibles en imágenes, simulando una codificación y decodificación como en los métodos tradicionales.
- **Funciona** : Se entrena una red **encoder** para insertar la marca en la imagen y una red **decoder** para recuperarla incluso si la imagen fue alterada (compresión, ruido, recorte, etc.).
- **¿Para qué sirve?**: Permite crear un sistema inteligente que oculte información de forma imperceptible y la recupere de forma robusta, incluso si la imagen ha sufrido modificaciones.
- **Ventajas**: Alta imperceptibilidad visual y la capacidad de adaptarse a ataques si se entrena correctamente
- **Desventajas**: Requiere una gran cantidad datos para entrenar.



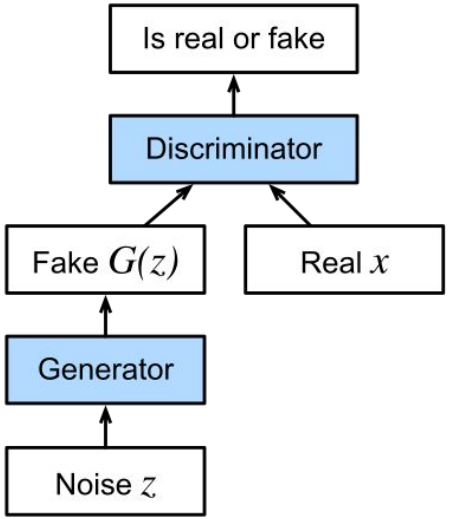
Se puede pensar en CNN como una red que aprende a esconder y encontrar marcas de agua por sí sola, ajustando la imagen de forma inteligente sin que se note visualmente.

4. Basadas en IA:

GAN

(Generative Antagonic Networks)

- **¿Qué hace?** Usa dos redes enfrentadas:
  - Un generador, que intenta ocultar la marca.
  - Un discriminador, que intenta detectarla. Ambas aprenden juntas para lograr marcas más resistentes e invisibles.
- **¿Para qué sirve?**: Para crear sistemas de watermarking que sean más difíciles de atacar, incluso por redes entrenadas para eliminarlos.
- **Ventajas**: Genera marcas altamente robustas e imperceptibles. Se adapta dinámicamente a intentos de detección o eliminación.
- **Desventajas**: Complejo y sensible a errores, alto consumo computacional.



Piensa en GAN como como un sistema de CNNs que compiten y se mejoran mutuamente.

D. Seguridad y vulnerabilidades

Seguridad

Robustez: La marca debe resistir ataques como compresión, redimensionamiento o edición.

Imperceptibilidad: No debe degradar la calidad de la imagen original.

Capacidad: Posibilidad de insertar información suficiente sin comprometer la imagen.

Fiabilidad: Debe permitir una recuperación precisa de la marca.

Vulnerabilidades

Ataques de compresión (JPEG): Pueden eliminar marcas poco robustas.

Filtros y transformaciones: Reducción, recorte, rotación o desenfoque pueden eliminar la marca.

Ataques de colusión: Comparar múltiples versiones para deducir la marca.

Eliminación por IA: Herramientas de edición avanzadas pueden borrar marcas visibles o modificar patrones invisibles.

Resumen Comparativo de Vulnerabilidades

Algoritmo	Vulnerable a	Robustez	Imperceptibilidad
LSB	Cualquier alteración (ruido, compresión, etc.)	Muy baja	Alta (pero insegura)
DCT	Redimensionado, transformaciones geométricas	Alta	Alta
DWT	Transformaciones geométricas, recorte parcial	Media - Alta	Alta
SVD	Alteraciones en subespacios, geometría compleja	Alta	Muy Alta
DWT-SVD	Alteraciones simultáneas en múltiples niveles	Alta	Muy Alta
DCT-SVD	Transformaciones + compresión severa	Alta	Alta
DWT-DCT-SVD	Ataques muy severos o combinados	Muy Alta	Muy Alta
CNN (IA)	Ataques no vistos durante entrenamiento	Alta (variable)	Muy Alta
GAN (IA)	Inestabilidad del entrenamiento, ataques dirigidos	Muy Alta	Muy Alta

III. Conclusiones

La protección de imágenes es un desafío en constante evolución.

Las técnicas de watermarking, especialmente las invisibles, ofrecen un equilibrio ideal entre protección, estética y funcionalidad.

Sin embargo, es crucial complementar estas tecnologías con buenas prácticas legales y un enfoque integral que incluya cifrado y monitoreo.

IV. Preguntas debate

Base al contenido:

1. ¿Qué técnica de watermarking invisible es la más equilibrada en términos de protección y calidad visual: LSB, DCT, DWT, SVD o una híbrida? ¿Por qué?

La técnica más equilibrada es la híbrida DWT-DCT-SVD, porque combina robustez ante ataques como compresión, recortes y transformaciones, con una alta imperceptibilidad que no afecta la calidad visual de la imagen.

2. ¿Por qué no basta con usar marcas de agua visibles para proteger una imagen en internet?

Porque se pueden eliminar fácilmente con herramientas de edición, afectan negativamente la estética de la imagen, y no ofrecen capacidades reales de rastreo ni una protección legal sólida.

Abiertas

3. ¿Debería el watermarking convertirse en un estándar de protección en plataformas de contenido como redes sociales o bancos de imágenes?

4. ¿Cuál es el mayor reto actual para lograr que el watermarking invisible sea realmente indetectable e imborrable?

5.¿Sería viable un ecosistema digital donde todas las imágenes lleven watermarking por defecto desde su creación?

6.¿Qué riesgos de privacidad o mal uso puede traer el uso masivo de watermarking digital invisible en fotografías personales?

Gracias