Capítulo 2: Técnicas de inserción de marcas de agua

2.1 Introducción

El primer paso en el estudio de un sistema de sellado digital usando marcas de agua consiste en la definición del proceso de inserción de la marca, tarea no vana, puesto que las propiedades de rendimiento de la técnica utilizada dependen, en gran medida, de la forma en que el sello se inserta en los datos originales. Es por ello que se dedica este capítulo a la descripción de algunos de los criterios básicos de clasificación de marcas de agua en una primera parte, dedicándose el resto al estudio de los diferentes procesos incluidos en la etapa de generación e inserción de la marca, mostrados en la figura 2.1.

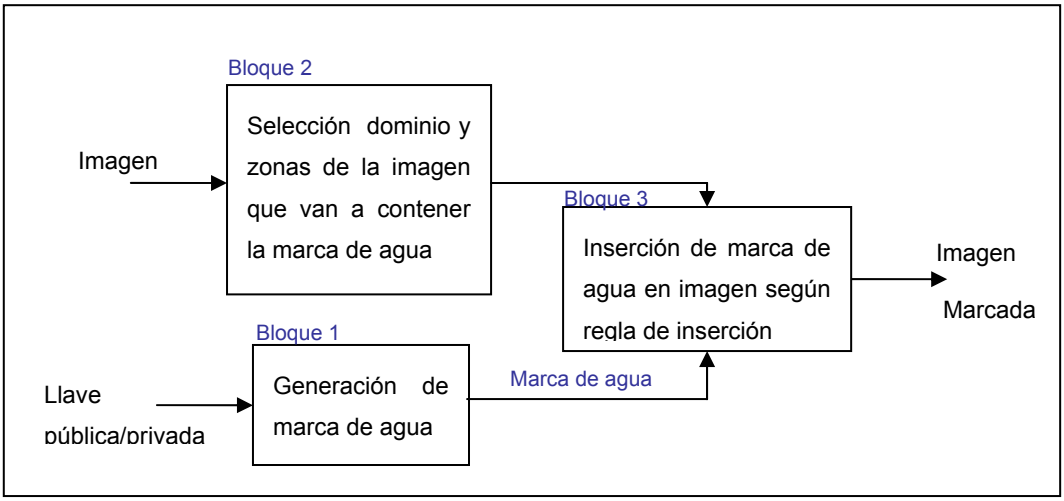

FIGURA 2.1. Diagrama de bloques genérico del proceso de sellado de la imagen.

La información de la marca de agua se puede encajar en la imagen de diversas maneras. Puede añadirse directamente a los bits menos significativos de los píxeles de la imagen, distribuirse por toda la imagen de manera aleatoria, colocarse en las zonas más ruidosas de la misma… Los métodos de inserción de marcas de agua en datos (imágenes, vídeo, audio) son prácticamente infinitos, creciendo, además, su número continuamente, de tal manera que la idea de intentar enumerarlos de una forma rigurosa resultaría tediosa y carente de sentido. Lo que aquí se pretende es indicar una clasificación más o menos general que pueda englobar los diferentes mecanismos, y quizás enumerar algunos de ellos por razones ilustrativas o históricas.

2.2 Clasificación de las técnicas de inserción de marcas de agua.

De entre los múltiples criterios de clasificación de las técnicas de sellado digital con marcas de agua seleccionamos los siguientes, dada su extendida utilización [Barni01], [Heileman98], [Comezaña01], [Kutter99], [Delfino00].

- Según su **robustez**, las marcas de agua pueden ser *robustas* o *frágiles* dependiendo de la resistencia que presenten a ataques que traten de borrarlas de manera intencionada o no. Una marca débil es aquella que se elimina con facilidad con un mínimo procesado aplicado a la imagen dejando ver aquellas zonas de la misma que han sufrido modificación, mientras que una marca robusta debe resistir ataques más duros.
- Atendiendo al **criterio de visibilidad**, las marcas de agua podrán ser *visibles* o *invisibles* según sean perceptibles o no en el documento protegido mediante esta técnica.
- En función del **dominio de representación** de la imagen elegido para la inserción de la marca de agua, las técnicas de sellado digital también se pueden clasificar según que la marca de agua se añada al documento original formulado en el *dominio temporal* espacial (píxeles de la imagen) o bien en el *frecuencial* o cualquier otro *dominio transformado*. Se comprobará que si la marca de agua se inserta en un dominio transformado de la imagen, ésta resulta ser más robusta a ciertos ataque básicos tales como el filtrado paso bajo o la compresión JPEG de la imagen marcada.
- Según **se disponga** o no, en el proceso de detección de la marca, del **documento original sin marcar**, el esquema de sellado con marcas de agua puede ser *privado* o *público*, respectivamente. Los esquemas de watermarking privados son también conocidos como métodos ciegos.
- También se clasifican los métodos atendiendo a la **dependencia** que tiene la marca de la imagen, es decir, *marcas dependientes* y *marcas independientes* de la imagen. Las marcas de agua dependientes del documento original tratan de adaptarse a las características del sistema humano que interviene en el

procesado del documento, es decir, si se trata de una señal de audio, una marca de agua dependiente de la señal se adecuaría a las características de la misma haciendo uso de las deficiencias del sistema auditivo humano, ocurriendo lo mismo para el caso de las imágenes con el sistema visual humano (referido como HVS).

2.2.1 Técnicas de sellado aplicadas a imágenes estáticas. Características

En el presente proyecto, como ya se ha mencionado, se estudiarán diferentes técnicas de sellado digital mediante marcas de agua en imágenes estáticas, con el objeto de protegerlas de usos no permitidos, por lo que es lógico pensar que las marcas de agua insertadas deberán:

- Ser **invisibles** a la percepción humana y no afectar a la calidad de la imagen de manera severa.
- Identificar de forma no ambigua al propietario. Sería crítico que una misma marca de agua se generase con diferentes claves.
- No ser detectadas mediante pruebas estadísticas.
- Ser difíciles de eliminar. Deben ser **robustas** a los diferentes algoritmos de procesado de imagen, como pueden ser filtrado, compresiones, reescalados, etcétera..., así como a intentos deliberados de eliminarla.

Los algoritmos implementados serán privados y se realizarán mediante el marcado de la imagen en algún dominio transformado de la misma, tratando de cumplir, con mejor o peor aproximación, las premisas enumeradas.

2.3 Generación e inserción de marcas de agua

En este apartado se van a considerar las características principales de los procesos de generación de marcas de agua, así como de las reglas de inserción de las mismas en la imagen [Barni01]. Véase figura 2.1 bloque uno.

2.3.1 Generación de marcas de agua

Para generar una señal de marca de agua deben tenerse en cuenta las especificaciones señaladas con anterioridad, en particular, aquella que indicaba que una marca de agua debe identificar de forma única a un determinado

propietario, pues no sería nada deseable que una misma marca se generase, por ejemplo, con diferentes claves, lo que significa que identificaría a varios propietarios y sería imposible conocer a cuál de ellos corresponde cada imagen.

La señal de marca de agua se genera, normalmente, como una cadena pseudoaleatoria de muestra independientes e idénticamente distribuidas, a partir de una semilla o clave secreta (para garantizar la seguridad del sistema) y con distribución normal de media cero y varianza unitaria ($N(0,1)$), en la mayoría de los algoritmos de inserción [Cox97], [Barni01]. La elección de esta forma deriva de las analogías encontradas entre las técnicas de watermarking y los sistemas de comunicaciones tradicionales. Según éstas la técnica que nos ocupa puede verse como la transmisión de una señal débil por un canal muy ruidoso, problema resuelto en algunos sistemas mediante el empleo de técnicas de espectro expandido. Dos de los algoritmos implementados utilizarán estas técnicas para incrustar el sello en la imagen.

Además, esta señal de marca de agua también puede ser antipodal (tomando sólo los valores *{+1, -1}*), protegida mediante el empleo de técnicas de codificación de canal…etcétera. Existen tantas opciones como ideas en la imaginación del desarrollador.

Es importante diferenciar entre los términos *señal de marca de agua* y *mensaje de la marca de agua*. El mensaje se corresponde con la información directa acerca del propietario de la imagen, número de ISBN, fecha de creación y todo aquello que se crea conveniente conocer acerca de la imagen y/o su propietario/creador. La señal de marca de agua es la que realmente se inserta en la imagen, el resultado de cifrar el mensaje original para poder insertarlo en los datos a proteger de una manera más eficaz, y es el tipo de marca de agua con el que se trabajará en este proyecto, quedando fuera de su alcance todo lo relacionado con el mensaje de marca de agua. Es decir, haciendo uso, de nuevo, de las analogías existentes entre las técnicas de marcado de agua y los sistemas de comunicaciones tradicionales, nuestro trabajo empezaría a partir de la salida del codificador de fuente, esa sería nuestra señal de marca de agua generada.

Además, si se considera que los posibles ataques a los que se puede someter una imagen para conseguir eliminar la marca que contiene, están limitados por la necesidad de no degradar demasiado la calidad del elemento marcado, se puede deducir que mientras más parecida a mencionada imagen

sea la marca de agua insertada más difícil será distinguir entre la marca que se quiere destruir y la imagen que se desea preservar. La formulación de este enunciado depende lógicamente de la métrica de distorsión empleada, por ejemplo, si se usa el criterio MSE (Mean Squared Error) se pueden obtener algunas condiciones que debe cumplir el espectro de potencia de la marca. Para conseguir la forma adecuada de la marca de agua se puede realizar un filtrado sobre una secuencia intermedia pseudoaleatoria, o bien usar secuencias caóticas. Notar, no obstante, que un método más adecuado debería considerar la percepción de la degradación por un observador humano, más que el criterio MSE [Petitcolas98], [Barni01], [Anderson98].

2.3.2 Inserción de marcas de agua

Desde un punto de vista general el proceso de inserción de la marca se lleva a cabo mediante la obtención, en primer lugar, de un conjunto de características de la imagen, que luego serán modificadas de acuerdo al contenido de la marca de agua. De manera que se requieren dos pasos para definir el proceso de sellado:

- Elección de las características de la imagen que se van a modificar con la inserción de la marca. (figura 2.1 bloque dos).
- Definición de la regla de inserción a utilizar (figura 2.1 bloque tres).

Se han propuesto variadas soluciones, resultando en diferentes tipos de sistemas de watermarking, se mostrarán las más extendidas.

2.3.2.1 Dominios espaciales y transformados de la imagen

Como ya se ha indicado, la mayoría de los sistemas de watermarking requieren un esquema en el que las modificaciones que introduce la marca no alteren de forma severa la calidad visual de la imagen original. Otro requisito importante era la robustez a las alteraciones debidas al procesado de señal, que intencionadamente o no, tratan de borrar o alterar el contenido de la marca.

La inserción de la marca puede hacerse directamente en el espacio de señal original de la imagen o en algún dominio transformado para explotar propiedades de percepción y/o conseguir robustez a ciertas transformaciones debidas a diferentes procesados de señal [Heileman98]. Muchas veces la inserción directa en el espacio de señal es deseable para conseguir reducida

complejidad, bajo coste, poco retraso o algún otro requerimiento del sistema. La regla de inserción determina la localización del píxel así como la intensidad de la señal de marca de agua que se va a añadir. La localización de la zona de inserción de la marca puede determinarse por un procesado de forma de onda de bajo nivel o bien por algún tipo de procesado de mayor nivel como detección de bordes o extracción de características.

La inserción en el dominio transformado incluye la transformada discreta del coseno basada en bloques o no (DCT) o la transformada wavelet (DWT), así como cualquier otra representación en el dominio de la frecuencia. Los esquemas de inserción de marcas de agua pueden elegirse de tal manera que se evite que la marca se añada a coeficientes que puedan ser eliminados o descartados en posteriores procesados, tales como compresión JPEG o filtrado paso de baja.

Las restricciones de diseño que tratan de garantizar la invisibilidad pueden ser incorporadas en representaciones en el dominio de la frecuencia, mediante la eliminación de las componentes de baja frecuencia para insertar la marca, en las que cualquier alteración puede traducirse en una distorsión visible. La característica de robustez también se puede conseguir mediante la elección de un adecuado dominio transformado de la imagen para insertar la marca [Barni01]. Esta elección se hará en función de los ataques frente a los que deseemos que la marca sea resistente. Se trata, por tanto, de una decisión de compromiso que deberá basarse en los objetivos y aplicaciones de la técnica de watermarking que se pretende desarrollar.

2.3.2.2 Reglas de inserción

Elegido el dominio en el que se va a insertar la marca de agua, la regla que se utilice para encajarla en las características de la imagen, elegidas para tal efecto, debe definirse. Es también importante conocer la forma de la marca a insertar, puesto que el rendimiento último de un sistema de sellado con marcas de agua depende también de esta aspecto como ya se ha referenciado en el apartado 2.3.1. [Cox97], [Barni01], [Heileman98].

2.3.2.2.1 Reglas de inserción de marcas de agua

Los mecanismos de inserción de marcas de agua en una imagen más comunes son los siguientes.

- Regla aditiva

$$I'_i = I_i + \alpha \cdot w_i \quad \textbf{(2.1)}$$

Donde:

I_i *:* componente i-ésima del vector de la imagen original, sea en el dominio espacial o transformado.

w_i *:* se corresponde con la i-ésima muestra de la marca.

α : factor que controla la intensidad de la marca.

I_i' *:* representa el i-ésimo elemento del vector correspondiente a la imagen sellada (la estego imagen), en cualquier dominio de la misma.

- Regla multiplicativa

$$I'_i = I_i + \alpha \cdot w_i \cdot I_i \quad \textbf{(2.2)}$$

Donde los símbolos utilizados tienen el mismo significado que en la ecuación 2.1.

- Regla exponencial

$$I'_i = I_i (e^{\alpha \cdot w_i}) \quad \textbf{(2.3)}$$

La regla aditiva es siempre invertible, las reglas multiplicativa y exponencial son invertibles sólo si $I_i \neq 0$. La regla aditiva puede no ser adecuada cuando los valores de I_i tienen un rango de variación alto. En este caso técnicas de inserción basadas en las ecuaciones 2.2 y 2.3 resultan ser más robustas frente a tales diferencias de escala. Las ecuaciones 2.2 y 2.3 presentan resultados similares cuando $\alpha \cdot w_i$ es pequeño [Cox97].

La principal razón para la popularidad de la regla aditiva es su simplicidad. Otra ventaja es que bajo la asunción de que las características de la imagen, en las que se va a encajar la información correspondiente a la señal de la marca de agua, siguen una distribución gaussiana y que los ataque están reducidos a la suma de ruido blanco la técnica de decodificación basada en la correlación es óptima, en ese caso la probabilidad de error o la probabilidad de pérdida, dada una tasa de falsa alarma, pueden ser minimizadas. La adopción de la decodificación por correlación permite también hacer frente a los

desplazamientos espaciales, debidos por ejemplo al cortado de la imagen. La búsqueda exhaustiva de la marca de agua por todas las posibles localizaciones espaciales de la imagen es fácil de implementar en el dominio transformado, puesto que la correlación de señales en el dominio espacial equivale a una multiplicación en el dominio transformado.

Las técnicas que operan en el dominio frecuencial suelen usar la regla de inserción multiplicativa [Cox97]. Fundamentalmente para evitar la posibilidad de que al insertar el sello se cambie el signo de la transformada, en ese caso tendríamos problemas de sincronización en la detección.

2.3.3 Características del sistema visual humano (HVS). Modelo HVS

Para saber cuál es la localización más adecuada de la marca de agua es importante conocer el funcionamiento del sistema visual humano y aprovechar sus deficiencias en cuanto a resolución de cara a insertar señales en la imagen que resulten imperceptibles para el observador y que a su vez presenten cierto grado de robustez frente a los ataques más habituales que pueda sufrir la imagen.

De manera que será interesante disponer de un modelo del sistema visual para poder realizar, y posteriormente usar, máscaras perceptuales que estén basadas en este modelo [Heileman98], [Lim90].

Un modelo del sistema visual humano se puede representar de la siguiente manera:

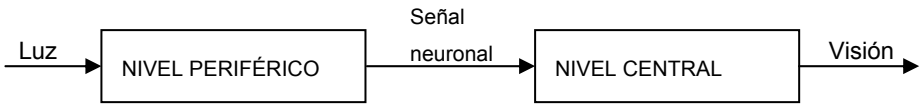

FIGURA 2.2. Modelo simplificado del sistema visual humano.

El nivel periférico es la parte del HVS donde la luz se convierte en señales neuronales. Este nivel es bastante conocido, especialmente en el caso de imágenes en blanco y negro. El nivel central procesa las señales neuronales para extraer la información y conseguir la visión, se trata de un nivel cuyo funcionamiento nos es prácticamente desconocido. Existen modelos relativamente simples para el sistema periférico pero no para el nivel central dadas su complejidad y escaso conocimiento.

2.3.3.1 Nivel Periférico

Un modelo sencillo del nivel periférico que tiene en cuenta un gran número de fenómenos visuales se muestra en la figura 2.3.

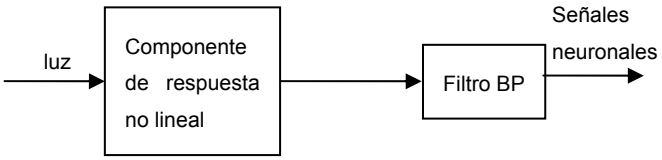

FIGURA 2.3. Modelo del nivel periférico del sistema visual humano.

El elemento no lineal lo que hace es comprimir los altos niveles de intensidad y expandir los más reducidos. El filtro paso banda se puede modelar como un sistema lineal e invariante en el espacio (sistema LSI) con una respuesta en frecuencia que tiene su valor máximo en el rango medio de frecuencias. El procesado realizado en el nivel periférico conlleva una serie de fenómenos visuales, indicamos los más importantes en relación con el tema objeto del proyecto.

- **Efecto de banda**.

  Por este efecto se observa que la percepción del brillo en regiones de intensidad uniforme no es homogénea. Gracias a este efecto no es preciso preservar con mucha exactitud la forma de los bordes.

- **Resolución en intensidad**.

  El sistema visual humano permite que pequeñas modificaciones en el valor de intensidad de los píxeles no sean percibidos. De manera que si en las técnicas de sellado invisible se insertan las marcas en los bits menos significativos (LSB) de la imagen estas pasarán desapercibidas.

- **Sensibilidad a la intensidad**.

  Hace referencia a que el incremento de intensidad en el valor de un píxel que se puede añadir de manera que no sea percibido varía en función de la intensidad de los píxeles vecinos. Es decir, el valor de ΔI que causa una diferencia en el píxel ya apreciable aumenta proporcionalmente con I. Esta propiedad también puede ser usada en las técnicas de watermarking, de manera que se añada mayor

cantidad de información adicional en regiones cuyos píxeles tengan valores altos de intensidad, y reduciendo o eliminando la inserción de marcas de agua en zonas de menor intensidad.

- **Resolución espacial**.

  Píxeles aislados pueden ser modificados sin que se perciba el cambio. Característica que presenta un problema fundamental debido a que las modificaciones que se añadan a los píxeles aislados introducirán componentes de alta frecuencia en la imagen modificada, fáciles de eliminar mediante el uso de un filtro paso bajo con mínimos efectos sobre la imagen.

- **Efecto de máscara espacial**.

  Cuando se añade ruido aleatorio a una imagen, este es más notable en regiones que tienen valores de intensidad uniforme que en aquellas donde hay un alto nivel de contraste, como regiones de borde, que se corresponden con zonas de frecuencia espacial alta. Además el ruido es menos notable en las regiones oscuras que en las claras. Se observa, por tanto, que la intensidad de los píxeles cercanos a bordes puede ser modificada en mayor medida sin que se aprecie visualmente cambio alguno en la imagen, en especial lo bordes oscuros. Además se concluye también que grandes modificaciones en regiones de intensidad o color homogéneos no deben ocurrir, sobre todo en regiones claras.

- **Intensidad del canal azul**.

  El sistema visual humano no es simétrico con respecto a cómo se perciben los cambios en las tres componentes de color básicas. Se estima que presenta entre tres y cinco veces menos sensibilidad a los cambios en la componente de color azul de la imagen que a los cambios en las otras componentes de color.

Todos estos fenómenos mencionados son todavía más acentuados cuando no se dispone de la imagen original para compararla con la imagen marcada. De entre todos ellos el más utilizado, normalmente, por las técnicas de sellado invisible es el fenómeno de resolución de intensidad.

Máscaras Perceptuales

Debido a los fenómenos visuales enumerados en el apartado anterior, se deduce que no todos los píxeles de la imagen son igualmente aptos para portar la marca de agua. Las máscaras perceptuales hacen uso de estas propiedades con el fin de conseguir marcas de agua más robustas.

La idea básica será aumentar la intensidad de una marca de agua en aquellos píxeles de la imagen donde los cambios sean menos apreciables y reducirla en las zonas que sean más sensibles a los cambios. Además, la mayor parte de la señal de marca de agua debe colocarse en las regiones más significativas, sólo de esta manera se dispondrá de una marca robusta frente a ataques que traten de eliminarla.

2.3.3.2 Nivel central

El procesado que tiene lugar en este nivel es prácticamente desconocido. Indicar simplemente que muchos elementos de apariencia artificial o no natural percibidos en las imágenes se observan, en la mayoría de los casos, gracias al procesado que tiene lugar en este nivel.

2.4 Conclusiones

En conclusión, la selección de las características de la imagen o el dominio transformado adecuado para la inserción de la marca de agua se basan, a menudo, en el conocimiento de la aplicación para la que se realiza la técnica concreta y en la elección de un espacio en el que se puedan desacoplar las componentes significativas de la imagen de aquellas que no lo son. Para conseguir que la marca sea transparente esta debería ir insertada en las componentes no significativas, sin embargo si se desea que sea robusta debería ir junto con las componentes más significativas de la imagen. Modelos de percepción del HVS pueden ayudarnos a diseñar esquemas de inserción de marcas de agua que nos permitan encontrar un compromiso entre ambos requisitos incompatibles.

En las técnicas de sellado digital de imágenes, la selección de las características puede ocurrir en el dominio espacial, donde la localización del píxel o zona espacial local a marcar es elegida basándonos en algún criterio de percepción. Un ejemplo podría ser una medida de la actividad local por bloques (como la varianza por ejemplo) decidiéndose que los bloques de datos a marcar

sean aquellos cuya varianza sea bastante superior a un determinado umbral calculado de forma empírica.

Trabajar en el dominio transformado es una buena elección para poder aplicar determinadas propiedades de percepción en el proceso de inserción de marcas de agua. Un método bastante generalizado consiste en utilizar la transformada discreta del coseno (DCT) de la imagen basada en bloques. Para tener en cuenta ciertos efectos psicovisuales se debe considerar lo siguiente:

- No insertar la marca en las bajas frecuencias porque sería visible.
- No marcar las altas frecuencias porque el sello sería fácilmente eliminado mediante procesados del tipo filtrado paso bajo o compresión con pérdidas (JPEG).
- Lo que debe hacerse es marcar un conjunto fijo o subconjunto aleatorio de frecuencias medias.

Modelos más sofisticados proponen el empleo de un coeficiente de intensidad α variable según las características de la imagen. Para conseguirlo suele usarse la magnitud de la DCT para determinar la potencia de la señal de marca de agua asociada al coeficiente correspondiente de la DCT. Para la mayoría de los esquemas de sellado que adaptan la intensidad de la marca usando información perceptual, el factor de escalado de intensidad es función de los datos de la imagen por lo que varía en función de las características de la misma, de manera que el esquema de inserción no es estrictamente aditivo, en el sentido de que el factor que escala la energía con que la marca de agua modula a la imagen es función de las características locales de la misma, de modo que su valor no es constante.

También debe considerarse la característica de sensibilidad a la frecuencia, independiente de las característica de la imagen pero relacionado con las condiciones de visión de la misma, o bien otras propiedades que incluyen características dependientes de la imagen, como sensibilidad a la luminancia (habilidad para detectar ruido considerando diferentes niveles de luminancia) o máscara de contraste (habilidad para detectar una señal en presencia de otra: características de textura, detalles de alta frecuencia).

En conclusión, se han identificado tres aspectos relacionados con la inserción de marcas de agua:

- Selección del dominio de la imagen y/o área de la misma apta para albergar la marca de agua.
- Elección de la regla de inserción.
- Explotación de los factores psicovisuales, comportamiento del sistema visual humano.