Capítulo 4. Detección y recuperación de la marca de agua

4.1 Introducción

En este capítulo se van a caracterizar algunas de las técnicas más utilizadas en la actualidad para la detección y recuperación de la información contenida en las marcas de agua. En primer lugar se verá una posible analogía entre los sistemas de sellado digital con marcas de agua y los sistemas de comunicaciones tradicionales para fundamentar el empleo de ciertas técnicas de detección ya utilizadas en estos sistemas tradicionalmente. En apartados siguientes se verán los dos mecanismos básicos para extraer de la imagen sellada la marca de agua que tiene incrustada, tales métodos básicamente incluyen sustracción de la imagen original a la marcada con decisión basada en correlación y aplicación de las técnicas de procesado de Análisis de Componentes independientes (ICA) a las observaciones disponibles. Finalmente se expondrán unas conclusiones.

4.2 Analogías entre los sistemas de sellado digital y los sistemas de comunicaciones

Si examinamos cualquier esquema de watermarking, podemos apreciar que es muy parecido a un sistema de comunicaciones tradicional. El objetivo de los sistemas de watermarking es, en esencia, el mismo: introducir cierta información en un medio, llamado canal (que en nuestro caso sería la imagen), para luego intentar extraerla de la manera más fiable posible. Un esquema que muestra tal analogía se ilustra en la figura 4.1, [Hernández01].

En el esquema de la figura se observa que el componente encargado de generar la marca podría ser visto como el transmisor en el sistema de comunicaciones, el extractor de marca de agua sería el receptor y el conjunto de la imagen a proteger y los posibles ataques que pudiera sufrir la misma se correspondería con el conocido canal de comunicaciones.

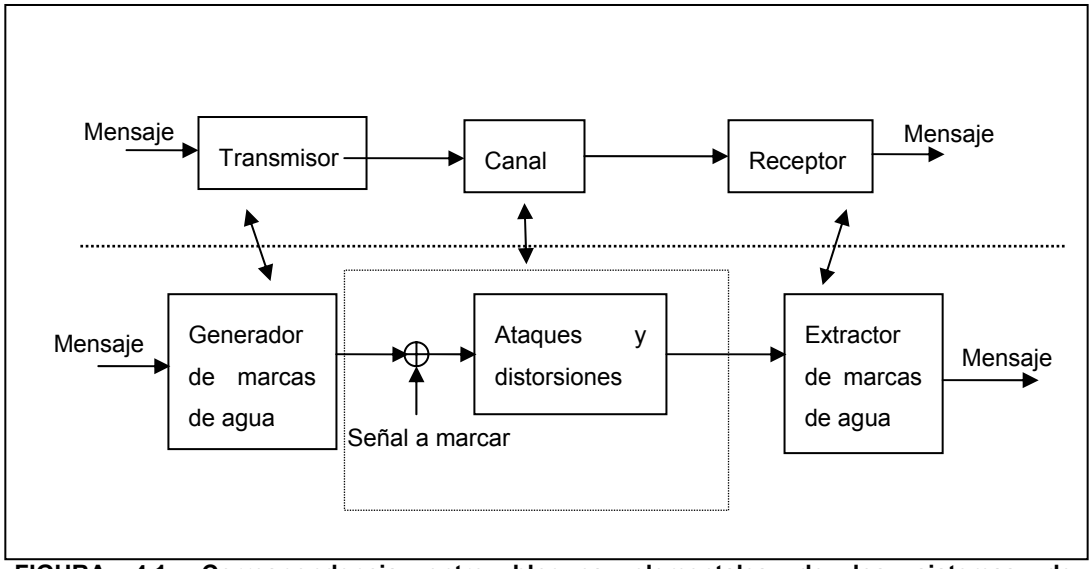

FIGURA 4.1. Correspondencia entre bloques elementales de los sistemas de comunicaciones y de watermarking.

El modelo de canal depende del escenario de la técnica de sellado y de la cantidad de información disponible en el proceso de extracción de la marca de agua. Por ejemplo si la imagen original no está disponible en la extracción de la marca de agua, entonces debe ser considerada como ruido aditivo. Pero si se dispone de una copia de la señal original sin marcar entonces tenemos un canal del que disponemos de información suficiente para mejorar el proceso de extracción.

En una aplicación real de las técnicas de sellado digital de imágenes que nos ocupan, la señal marcada puede sufrir alteraciones causadas por manipulaciones de procesado de señal durante la distribución y almacenamiento, u otro tipo de distorsiones, como ya se ha explicado de un modo algo más detallado en el capítulo anterior. Todas estas fuentes de distorsión pueden ser incluidas en nuestro modelo de comunicaciones como parte del canal.

Otro parecido entre las técnicas de watermarking y los sistemas de comunicaciones es que en ambos casos el medio impone ciertas restricciones en la señal que porta la información. En los sistemas de comunicaciones, generalmente tenemos restricciones de máxima potencia de pico o media, asociadas a limitaciones físicas en los transmisores. Como ya es sabido, en watermarking, estas restricciones están relacionadas con el hecho de que la calidad visual de la imagen marcada no debe verse degradada.

Hemos identificado una correspondencia entre elementos en un sistema de watermarking y bloques en un sistema de comunicaciones. Por lo tanto, se puede pensar que los conceptos de capacidad del canal podrían aplicarse al contexto del sellado digital de imágenes. De manera que podremos saber, por ejemplo, la cantidad de bits que se pueden insertar en la imagen de manera fiable, es decir, sin que se perciba que la imagen difiere de la original y siendo la marca robusta frente a diversos ataques. Resulta obvio que no todas las imágenes pueden admitir la misma cantidad de información insertada de manera invisible. Por ejemplo la cantidad de información que se podrá insertar en una imagen de luminancia constante será mucho menor que la que cabría en una imagen de diferentes texturas o gran cantidad de bordes.

Sin embargo, podemos encontrar algunas diferencias con respecto a los sistemas de comunicaciones clásicos. La principal es que la imagen sellada puede sufrir distorsiones no sólo debidas a procesados que tratan de mejorar la imagen, sino también a ataques maliciosos cuyo objetivo consiste en destruir la marca de agua o conseguir que no sea detectable en el proceso de extracción de la misma. Los ataques también se pueden suceder en el caso de aplicaciones militares en comunicaciones de radio, en ese caso un enemigo podría transmitir una señal parecida a ruido de alta potencia en la misma frecuencia utilizada para la comunicación. Entonces, ¿cuál sería la diferencia entre un sistema de comunicaciones tradicional y un sistema de sellado digital de imágenes con marcas de agua? Pues bien, la diferencia está en que en los sistemas de comunicaciones los ataques se limitan a señales aditivas puesto que el atacante no puede alterar el medio de transmisión. En el caso de watermarking, un atacante puede aplicar cualquier clase de algoritmo de procesado de señal a su disposición para alterar la señal marcada. El ataque podría modelarse como un canal puesto que toma la imagen sellada como entrada y da como resultado una versión alterada de la misma. Además el atacante podría siempre intentar encontrar un ataque que resulte más efectivo para cada tipo de algoritmo de sellado empleado. Su única restricción es que la imagen alterada que genera como resultado no debe diferir demasiado de la imagen original desde el punto de vista de percepción.

Otra diferencia es que en el caso de watermarking se dispone de mayor cantidad de información sobre el canal que en los sistemas de comunicaciones tradicionales. Esta situación es comparable con el hecho de conocer a priori, es decir, en el transmisor, el ruido máximo que puede introducir el canal de

comunicaciones. Esta valiosa información puede utilizarse para mejorar el rendimiento del sistema de sellado digital de imágenes con marcas de agua.

A pesar de las diferencias señaladas, se ha comprobado que la semejanza con los sistemas de comunicaciones es una característica importante y valiosa para el diseño de sistemas de watermarking. De hecho muchas de las técnicas aplicadas en sistemas de comunicaciones digitales han sido exitosamente realizadas en el diseño de sistemas de sellado de imágenes con marcas de agua.

4.3 Estructura general de un detector/extractor de marcas de agua.

El detector/extractor de marcas de agua persigue dos propósitos principales:

- Decidir si la imagen bajo prueba contiene o no una marca de agua.
- Extraer o decodificar el mensaje que la marca de agua podría contener.

El siguiente diagrama de bloques muestra un esquema simplificado de la estructura general de un detector de marcas de agua.

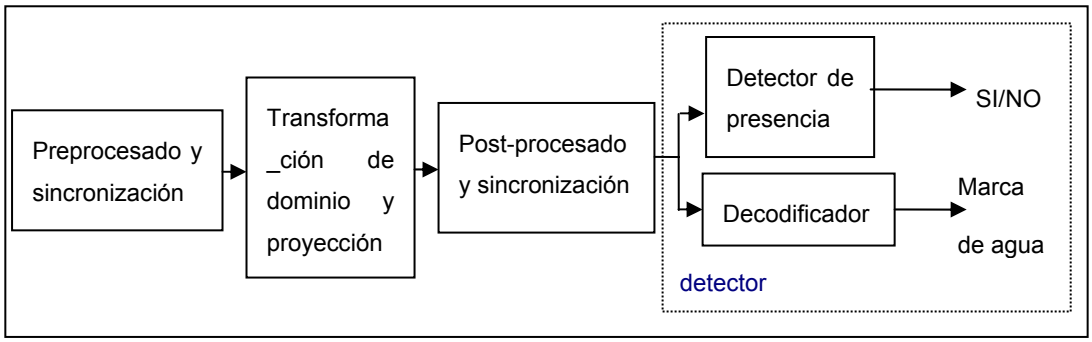

FIGURA 4.2. Estructura general de un sistema detector de marcas de agua.

En el esquema de arriba se aprecian las diferentes etapas de procesado que componen el sistema de detección de marcas de agua, que son:

- ***Preprocesado***. Realizado para eliminar contribuciones innecesarias de la imagen original que podrían distorsionar el proceso de detección. En este procesado pueden encontrarse contramedidas frente a diferentes ataques que suponemos sufrirá la imagen.
- ***Transformada de la señal*** al dominio en que se realizó el proceso de inserción de la marca de agua y proyección en un subespacio de la

misma mediante el empleo de la clave secreta utilizada en el proceso de inserción de la marca de agua en la imagen.

- ***Post-procesado***. Cuyo objetivo es mejorar las condiciones o características de las señales de manera que se facilite la detección aumentando además la robustez frente a ataques.

4.3.1 Detector de marcas de agua

Se trata del último elemento incluido en el bloque encargado de extraer las marcas de agua. Vemos el estado actual de este sistema de decisión [Hernández01], [Heileman98], [Cox97], [Cachin98].

Se pueden realizar dos tipos de pruebas, la de detección de presencia, que consiste en decidir si podemos asegurar que la imagen bajo prueba contiene una marca de agua generada a partir de una llave pública determinada, y la de decodificación de marcas de agua, que se emplea sólo cuando la marca insertada en la imagen contiene un mensaje digital (empleando una técnica de modulación digital) y consiste en extraer mencionado mensaje.

El test de detección de presencia de la marca de agua es importante en aplicaciones de protección de los derechos de autor, puesto que se utiliza para probar que el propietario de la llave es realmente el autor de la imagen protegida. Otras aplicaciones en las que el test de detección de presencia de marca es importante están relacionadas con el control de copias, en sistemas como el DVD, por ejemplo. El test de decodificación de marcas de agua es un elemento clave cuando la marca lleva información que puede utilizarse para determinar la responsabilidad sobre una copia ilegal.

Como ya se ha visto anteriormente, los ataques, el procesado de señal, distorsiones no intencionadas y la propia imagen, constituyen fuentes de incertidumbre que aparecen entre las etapas de inserción de la marca de agua y posterior detección de la misma. Para encontrar un algoritmo de sellado eficaz es preciso encontrar modelos estadísticos que caractericen apropiadamente estas fuentes de incertidumbre. Estos modelos resultarán útiles además para el desarrollo de buenas medidas del rendimiento.

En algunos casos puede ser complicado encontrar un modelo estadístico que describa con precisión una fuente de señal. Actualmente, por ejemplo, no existe tal modelo para la caracterización apropiada de las señales de luminancia y crominancia de imágenes estáticas. Una forma de poder eliminar este

problema consiste en usar un modelo muy conservativo, como por ejemplo una distribución uniforme de los valores de la luminancia. Encontrar un modelo que se adapte mejor al comportamiento de la imagen estática podría permitirnos derivar estructuras del extractor de marcas de agua con un rendimiento superior. Por ejemplo, cuando la marca de agua se inserta en un subespacio de señal, en muchas situaciones, la proyección de la imagen e incluso de las distorsiones introducidas por los ataques en el subespacio de señal puede ser mejor modelada estadísticamente.

Si introducimos una descripción estadística de todas las distorsiones que la marca de agua ha podido sufrir (incluyendo la suma de la señal original) hasta llegar al detector de marcas de agua, entonces podemos formular un test de la presencia de detección como un test de decisión de hipótesis binaria estadística. Las dos hipótesis son:

$H1$: "La señal bajo prueba contiene una marca de agua generada con la clave *K*."

$H0$: "La señal bajo prueba no contiene una marca de agua generada con la clave *K*."

Este test de decisión estadística puede ser interpretado para decidir si podemos considerar que la entrada del detector es un proceso aleatorio con la función densidad de probabilidad condicionada a $H0$ y $H1$. La estrategia de decisión óptima es la regla de Neyman-Pearson, puesto que maximiza la probabilidad de detección P_D (la probabilidad de no tener un falso negativo) para cada valor de la probabilidad de falsa alarma P_{FA} fijado (probabilidad de un falso positivo). Consiste en comparar la razón entre la función densidad de probabilidad (pdf) condicionada a $H1$ y la condicionada a $H0$, frente a un umbral.

$$\frac{f(y \mid H_1)}{f(y \mid H_0)} \geq \eta \qquad \textbf{(4.1)}$$

Donde *y* es la señal bajo prueba y η el umbral.

Cuando el canal de comunicaciones equivalente entre las etapas de inserción y recuperación de la marca de agua puede modelarse como un canal de ruido gaussiano aditivo, entonces la función de densidad de probabilidad (f.d.p.) condicionada a $H0$ es una gaussiana centrada en el origen. Si la marca contiene un mensaje, entonces la f.d.p. condicionada a $H1$ es una suma

ponderada de gaussianas centradas en todos los puntos que corresponden a una marca válida, una para cada mensaje posible.

El test de decodificación de la marca de agua puede formularse como un test de hipótesis M-ario, donde M es el número de posibles mensajes. La marca de agua será decodificada solo si la prueba de detección de la misma proporciona un resultado positivo. Por lo tanto solo se usa la función de densidad de probabilidad condicionada a $H1$. En este caso la prueba óptima sigue el criterio de máximo a posteriori (MAP) equivalente al principio de máxima verosimilitud (ML), muy usado en comunicaciones digitales, donde se supone que todos los mensajes tienen las mismas probabilidades a priori. Si aplicamos el test ML, entonces, el mensaje insertado será aquel para el que la función de densidad de probabilidad de la señal bajo prueba condicionada a ese mensaje es máxima:

$$D = \arg \max_{i \in \{1 \ldots M\}} f(y \mid m_i) \qquad \textbf{(4.2)}$$

Donde:

m_i: mensaje i-ésimo.

$f(y \mid m_i)$: es la función de densidad de probabilidad de la señal bajo prueba condicionada al mensaje.

D: índice entre 1 y M.

En muchos casos el canal de ruido gaussiano aditivo es un modelo preciso para caracterizar los efectos introducidos por la imagen original (desconocida para el extractor de la marca de agua) y algunos tipos de distorsiones y ataques. Incluso si un buen modelo estadístico no se puede definir para describir la señal principal en el dominio original en que está disponible, el modelo gaussiano se considera en muchos casos en el dominio transformado. Cuando éste es el caso, podemos aplicar la experiencia obtenida en el estudio de comunicaciones digitales a través de canales gaussianos para mejorar el rendimiento del decodificador de marcas de agua y hacerlo más robusto frente a ataques.

Por ejemplo, técnicas de codificación de error para transferir información de manera fiable a través de un canal ruidoso pueden aplicarse al contexto de watermarking para mejorar el rendimiento del extractor de marcas de agua mediante la inserción de redundancia. Los esquemas de codificación, tales como

códigos de bloques o códigos convolucionales son, por tanto, herramientas de utilidad en el campo de sellado digital de imágenes con marcas de agua, así como el empleo de técnicas de espectro expandido de secuencia directa. Dependiendo del tipo de código de canal utilizado, el correspondiente decodificador de canal puede ser insertado como parte del decodificador de marcas de agua.

4.3.2 Métodos de extracción de marcas de agua empleados

En los algoritmos implementados (cuya especificación y características se detallan en el capítulo 6) se han realizado dos mecanismos básicos para la extracción de las marcas de agua (de manera que se puedan comparar). El primero se basa en la detección mediante sustracción directa de la imagen original a la atacada, de manera que el resultado será la marca de agua estimada, pues la regla de inserción utilizada es la aditiva (veáse capítulo 2). El segundo método de extracción hace uso de un procesado de imagen de gran interés en la actualidad, como es el Análisis de Componentes Independientes, que básicamente, lo utilizamos para separar las dos señales contenidas en la imagen marcada atacada de manera que se pueda estimar la información correspondiente a la marca insertada, (un estudio más detallado de esta técnica puede verse en el capítulo 5, en el capítulo 6 se ilustra su adaptación al problema de detección en sistemas de watermarking). El test de detección utilizado para determinar la presencia/ausencia de marca de agua en la imagen atacada procesada, consiste en una medida de la correlación entre la marca de agua estimada y todas las posibles marcas de agua válidas, la justificación de su empleo y expresión formal de la misma, pueden verse en el capítulo 6, donde el umbral η arriba expuesto es referido como *umbral_sim*, además, podrá observarse que, con el objeto de mejorar las características de detección, se emplea otro umbral, referido como *umbral_relativo*, que marca la mínima distancia que debe haber entre el valor del estadístico (que será el máximo obtenido) para la marca de agua que el detector determina como presente en la imagen y el siguiente máximo local obtenido en dicho estadístico al comparar la estima con cualquier otra marca de agua.

4.3.3 Sincronización

Para una extracción de la marca de agua necesitamos acceder a los mismos coeficientes que fueron modificados en la inserción [Hernández01].

El problema de la sincronización es especialmente importante en las técnicas de sellado digital de imágenes y video cuando se requiere cierta robustez a transformaciones geométricas. Métodos para desarrollar la sincronización hay muchos pero los más destacados por sus resultados en el dominio que nos ocupa son básicamente tres.

- **Uso de dominios invariantes**. Consiste en utilizar dominios o transformadas que resulten invariantes frente a las variaciones geométricas que sabemos pueda sufrir la imagen, gracias al uso de este tipo de dominios no se requiere sincronización en el proceso de detección. Un buen dominio para este propósito lo constituye el módulo de la transformada de Fourier, que es invariante a traslaciones espaciales. El de Fourier-Mellin, que consiste en una combinación entre el módulo de la transformada de frecuencia y un mapeo log-polar, resulta invariante a las rotaciones y escalado de las imágenes… Sin embargo no se ha encontrado, hasta la fecha ningún tipo de transformación que resulte invariante a cualquier tipo de variación geométrica.

  El problema será por tanto identificar el dominio invariante que satisfaga los requisitos de robustez dependiendo de la aplicación.

- **Templates**. La definición de template en este contexto es una constelación predefinida de características que se añade a la imagen. Una posibilidad consiste en añadir un conjunto de picos de frecuencia, es decir, se aumenta la amplitud de un conjunto definido de frecuencias por encima de un umbral. En la detección se detectan los templates y su localización se utiliza para determinar la transformación geométrica que ha sufrido la imagen. Estos picos se posicionan en círculos, cuadrados y ejes diagonales o bien se determina su posición en función de la clave. El inconveniente en este caso consiste en la necesidad de insertar información adicional en la imagen, lo que provoca más distorsión.

- **Referencias propias**. Surgen para eliminar las desventajas del mecanismo de sincronización basado en templates, que requería la inserción de una señal adicional en la imagen, con el consiguiente aumento de la distorsión. En este caso la propia marca contiene la referencia geométrica para la sincronización, y la idea básica consiste

en la inserción múltiple de la misma marca de agua. En la detección se calcularía la autocorrelación del espectro de potencia de la imagen, la localización de los picos distribuidos que se obtengan determinará el tipo de transformación que ha sufrido la imagen.

4.4 Conclusiones

La idea principal que debe extraerse de todo lo mencionado se basa en que, dadas las analogías existentes entre los sistemas de comunicaciones y los mecanismos de protección de imágenes mediante sellado invisible de las mismas con marcas de agua, muchas de las técnicas de detección conocidas de los sistemas tradicionales pueden aplicarse con éxito en los esquemas de watermarking, además de mecanismos de codificación de canal o de espectro expandido con el objeto de mejorar la robustez de la marca de agua insertada.