Algoritmo de marca de agua para imágenes en escala de grises basado en secreto compartido visual y puntos característicos

Por

Dulce Rosario Herrera Moro

Tesis

Sometida como requisito parcial

para obtener el grado de

MAESTRA EN CIENCIAS EN EL ÁREA DE CIENCIAS COMPUTACIONALES

en el

Instituto Nacional de Astrofísica , Óptica y Electrónica

Febrero 2008

Tonantzintla, Puebla

Supervisada por:

Dra. Claudia Feregrino Uribe, INAOE

©INAOE, 2008

Derechos Reservados

El autor otorga al INAOE el permiso de reproducir y

distribuir copias de esta tesis en su totalidad o en partes

A mis Padres:

Yrene Moro Torres y Eliazar Herrera González que son todo para mi.

A mi compañero de vida y mejor amigo:

Angel Valentin Bautista Durán por su amor y paciencia.

Agradecimientos

A mis hermanos Martha Neri y Eleazar Herrera Moro por su cariño incondicional.

Al Instituto Nacional de Astrofísica, Óptica y Electrónica por brindarme una excelente formación profesional.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por su el apoyo otorgado a través de la beca No.201765.

A la doctora Claudia Feregrino Uribe mi más sincero agradecimiento por el apoyo brindado para la realización de este proyecto.

A mis compañeros de generación por su amistad y apoyo.

Resumen

En este trabajo de tesis se presenta un esquema de protección del derecho de autor en imágenes en escala de grises basado en el esquema de secreto compartido y puntos característicos. A diferencia de los esquemas tradicionales de marca de agua, el esquema propuesto no altera la imagen original. El algoritmo propuesto extrae características de la imagen original para crear un secreto visual compartido que se utiliza para reconstruir la marca de agua. Para alcanzar robustez contra los ataques geométricos de rotación y traslación, el algoritmo propuesto utiliza puntos característicos en la imagen como referencia para crear la imagen compartida. Los resultados obtenidos muestran que el algoritmo propuesto puede reconstruir la marca después de realizar una rotación, traslación, compresión JPEG y aplicar operaciones comunes sobre la imagen. Durante los experimentos realizados se obtuvo para el 94 % de los casos una correlación entre la marca original y la marca reconstruida superior a 0.90 y una precisión en la reconstrucción de la marca arriba del 80 %. Los valores más bajos de correlación y porcentaje de precisión alcanzados durante los experimentos fueron de 0.877 y 75 % respectivamente. La característica sobresaliente del algoritmo propuesto es la reconstrucción de la marca sin tener que regresar la imagen a su posición original después de una rotación.

Abstract

This thesis presents a copyright protection scheme for grayscale images based on secret sharing and feature points. Unlike the common watermarking scheme that modifies the original image to embed a watermark, the proposed scheme does not alter the original image. Instead, the algorithm extracts features of the original image to create a visual sharing secret that is used to retrieve the watermark. To reach robustness against rotation and translation geometrical attacks the proposed algorithm uses feature points as reference to create the shared secret. The results show that the proposed algorithm can retrieve the watermark after rotations, cropping, translations, compression JPEG and common image filters, reaching correlation . In the 94 % of the cases, the results of correlation between the original mark and the mark retrieved obtained for the proposed algorithm were higher than 0.90 while the accuracy rate was higher than 79 % . On the other hand, the worst values obtained were 0.877 in correlation and 75 % for the accuracy rate . The main feature of this work is that after rotation the image does not need to return back to the original position to retrieve the watermark.

Índice

Dedicatoria III

Agradecimientos V

Resumen VII

Abstract IX

Índice de Tablas XV

Índice de Figuras XVII

1. Introducción 1

1.1. Motivación 3

1.2. Objetivo General 5

1.3. Objetivos Específicos 5

1.4. Estructura de la Tesis 5

2. Seguridad en información digital 7

2.1. Criptografía 7

2.1.1. Secreto compartido 9

2.1.2. Secreto compartido visual 9

2.2. Esteganografía 12

2.2.1. Esquema general de las técnicas esteganográficas 12

2.2.2. Características de los sistemas esteganográficos 14
2.3. Diferencia entre Criptografía y Esteganografía 15
2.4. Marcas de agua 17
2.4.1. Orígenes de las técnicas de marca de agua 17
2.4.2. Requerimientos de las marcas de agua en imágenes digitales . . 18
2.4.3. Esquema general de un sistema de marcas de agua 20
2.4.4. Clasificación de las marcas de agua 23
2.4.5. Ataques a las marcas de agua 26
2.4.6. Aplicaciones de las marcas de agua 27

3. Esquemas basados en VSS **29**
3.1. Trabajos relacionados 31
3.2. El problema de sincronización 33
3.2.1. Clasificación de los esquemas de sincronización 34
3.3. Puntos característicos de imágenes digitales 36
3.3.1. Descriptores de puntos característicos 38
3.3.2. Aplicaciones de las características invariantes 39

4. Algoritmo propuesto **41**
4.1. Fase de generación del secreto compartido 41
4.1.1. Extracción de puntos característicos y descriptores 42
4.1.2. Construcción de la marca compartida 46
4.2. Fase de reconstrucción de la marca de agua 48
4.2.1. Reconstrucción de la marca en diferentes puntos. 50
4.2.2. Mejora de la marca reconstruida usando redundancia de información. 51
4.3. Modificaciones al algoritmo propuesto 52

5. Resultados **55**
5.1. Métricas utilizadas durante la evaluación de resultados 55

5.2. Experimentos 57
5.2.1. Manipulación de los valores de pixeles 58
5.2.2. Prueba para la detección de falsos positivos 64
5.2.3. Ataques géometricos 66
5.3. Discusión de resultados 79

6. Conclusiones y trabajo futuros **81**
6.1. Conclusiones 81
6.2. Trabajo futuro 82

Referencias **85**

Índice de Tablas

4.1. Tabla de verdad del operador XOR 48

5.1. Resultados obtenidos ante la ecualización, cambio en el contraste y adición de ruido gaussiano en la imagen *Lena*. 63

5.2. Resultados obtenidos por los diferentes algoritmos cuando se intenta reconstruir la marca de agua de imágenes diferentes a la usada para construir el secreto compartido 65

5.3. Resultados obtenidos por el algoritmo propuesto después de ataques de rotación a la imagen *Lena* 67

5.4. Resultados obtenidos por el algoritmo propuesto después de ataques de rotación a la imagen *Bárbara* 68

5.5. Resultados obtenidos por el algoritmo propuesto después de ataques de rotación a la imagen *Mariposa* 69

5.6. Resultados obtenidos por el algoritmo propuesto después de ataques de rotación a la imagen *Mandril* 70

5.7. Resultados obtenidos ante ataques de croping y traslación en la imagen *Lena* 74

5.8. Resultados obtenidos ante ataques de croping y traslación en la imagen *Bárbara* 75

5.9. Resultados obtenidos ante ataques de croping y traslación en la imagen *Mariposa* 76

5.10. Resultados obtenidos ante ataques de croping y traslación en la imagen *Mandril* 77

Índice de Figuras

2.1. Criptografía Visual 10
2.2. Codificación de los pixeles de la imagen en Criptografía Visual. 11
2.3. Esquema general de Esteganografía 13
2.4. Lineamientos de la Esteganografía 15
2.5. El problema de los prisioneros 16
2.6. Relación entre robustez, capacidad y seguridad en un sistema de marca de agua 19
2.7. Fase de generación e inserción en un sistema de marca de agua 21
2.8. Fase de Detección/Extracción en un sistema de marca de agua 22
2.9. Clasificación de los ataques a marcas de agua 26

3.1. Fase de generación del secreto compartido. 30
3.2. Fase de reconstrucción de la marca. 31
3.3. El problema de sincronización 34
3.4. Clasificación de los métodos de sincronización 35

4.1. Detección de puntos característicos *potenciales*. 43
4.2. Descriptores de puntos característicos 45
4.3. Extracción de puntos de referencia 45
4.4. Extracción de información de la imagen para la generación de la imagen compartida. 46
4.5. Transformada Log-polar. 47

4.6. Construcción de las imágenes $Sh2_i$. 48
4.7. Ejemplo de imagen compartida generada por el esquema propuesto . . . 49
4.8. Localización de los puntos de referencia después de rotar la imagen de $Lena$ 45º. 50
4.9. Reconstrucción de la marca en diferentes puntos. 51
4.10. Resultados de la combinación de marcas recuperadas 52
4.11. Torus-automorphism 53
4.12. Ejemplo de la imagen compartida generada después de las modificaciones hechas al esquema propuesto. 54

5.1. Imágenes utilizadas en los experimentos 58
5.2. Marcas binarias utilizadas en los experimentos 59
5.3. Degradación de la imagen después de la compresión jpg 59
5.4. Precisión alcanzada por los algoritmos después de un ataque de compresión jpg 60
5.5. Gráfica de degradación de la imagen después de ruido gaussiano 61
5.6. Precisión alcanzada por los algoritmos después de agregar ruido gaussiano 62
5.7. Prueba de falso positivo forzando puntos en el esquema propuesto . . . 66

Capítulo 1

Introducción

En la actualidad los medios digitales como: texto, audio, imágenes y video han ido reemplazado a sus contrapartes analógicas pues ofrecen ventajas tales como:

- Facilidad al crear, editar y almacenar documentos enteros o parte de ellos.
- Facilitan la integración de diferentes medios en uno sólo (multimedia) que enriquece su contenido.
- Tener acceso aleatorio a diferentes partes del documento.
- Realizar copias de los documentos de manera masiva sin representar un costo computacional y/o económico elevado.
- Las copias que se hagan de ellos son exactamente iguales a la original pues no presentan las variaciones eléctricas y/o magnéticas de los medios analógicos.
- La información se transmite fácilmente vía Internet.

Uno de los medios digitales con más auge en los últimos años y de interés en este trabajo son las imágenes digitales, ya que debido a la gran cantidad de información que poseen han sido utilizadas en diversas áreas de investigación tales como: medicina, robótica, reconocimiento de patrones, entretenimiento, etc.

Hasta el momento se han presentado las ventajas que ofrecen los medios digitales a los usuarios. Sin embargo, algunas de ellas representan un serio problema para los creadores y distribuidores que se traducen en pérdidas de capital y es que, como se mencionó anteriormente, los medios digitales permiten hacer copias de calidad igual a la original y pueden ser fácilmente trasmitidas vía Internet, hechos que favorecen la adquisición y distribución de copias ilegales.

Ahora bien, para el intercambio masivo de información digital vía Internet se han desarrollado distintas técnicas de seguridad. La técnica principal es la Criptografía, una técnica antigua que consiste en cambiar la estructura de los datos digitales con la finalidad de evitar que usuarios no autorizados conozcan el contenido de los documentos transmitidos. Las técnicas de Criptografía cubren los siguientes aspectos de seguridad:

- **Confidencialidad.** La información sólo puede ser accedida por usuarios autorizados.
- **Autentificación.** Las partes involucradas en una transmisión segura no pueden ser suplantadas.
- **Integridad.** Se asegura que los datos no han sido modificados por entidades ajenas al sistema.
- **No repudio.** Se refiere a que no se pueda negar la autoría de un mensaje enviado.

Sin embargo, y a pesar de los diversos servicios que ofrecen, las técnicas criptográficas están limitadas con respecto a la protección del derecho de autor pues una vez que se ha quebrantado el sistema criptográfico que protege el contenido de medio digital, nada impide que éste se distribuya libremente.

Para proteger el derecho de autor se han desarrollado varias técnicas denominadas *marcas de agua* que están basadas en la antigua técnica de Esteganografía. Las técnicas de marca de agua consisten en insertar en el medio digital, imágenes en nuestro caso; información que identifique al autor o propietario legal. La información insertada puede ser texto con el nombre del autor, imágenes con el logo de la compañía, etc.

Puesto que el principal interés en las técnicas de marca de agua es detectar copias ilegales existentes en la red, es indispensable que la información insertada pueda ser recuperada después de operaciones comunes que se le hacen a las imágenes como compresión, aumento de contraste, ruido etc. Lo que deja un campo abierto a la investigación pues hasta el momento ningún método ha logrado ser robusto ante cualquier tipo de manipulación que se le haga a la imagen.

1.1. Motivación

En los últimos años, las técnicas de marca de agua han sido las técnicas más utilizadas para la protección de los derechos de autor en los medios digitales y en particular de las imágenes digitales.

Las marcas de agua pueden insertar información del propietario legal de la imagen tanto en el dominio espacial como en el dominio de la frecuencia. El problema que se presenta al insertar información, ya sea en uno o en otro dominio, es que se altera la calidad de la imagen. Aunque la calidad de la imagen no se degrade a tal punto sea posible percibirla visualmente por los usuarios comunes, el ruido generado puede afectar el resultado en aplicaciones sensibles a la calidad de la imagen.

Los algoritmos de marca de agua que trabajan en el dominio de la frecuencia alteran la calidad de la imagen más que aquellos que lo hacen en el dominio espacial. Sin embargo, los algoritmos que trabajan en el dominio espacial son más susceptibles tanto a ataques geométricos como a ataques por filtrado de la imagen. Esta es una de las razones por la que la mayoría de los algoritmos de marca de agua se enfocan ya sea a preservar la calidad de la imagen o en ser robustos ante ataques geométricos.

Otro problema que se presenta al alterar la imagen durante el proceso de inserción de la marca es que si bien el ruido no es detectado visualmente, mediante análisis estadísticos a la imagen el ruido puede ser detectado y revelar así la posición de la marca de agua. Una vez revelada la posición de la marca en la imagen, usuarios malintencionados pueden alterarla o removerla.

En años recientes, se han propuesto nuevos esquemas de marca de agua basados en el esquema secreto compartido visual que no alteran de ninguna forma la imagen original y presentan buenos resultados ante ataques por filtrado de la imagen. No obstante, estos esquemas no resisten un ataque de tipo geométrico tal como: rotación, escalado o traslación, porque no poseen un mecanismo de sincronización que les permita relocalizar la marca después un ataque de este tipo.

Los esquemas robustos a ataques geométricos, en su gran mayoría trabajan en el dominio de la trasformada y además de alterar la calidad de la imagen no es posible recuperar la marca de manera que sea reconocible visualmente por el usuario. Por tanto, usando estos esquemas, el usuario tiene que confiar en las métricas de similaridad utilizadas por el algoritmo para determinar la presencia de la marca en la imagen.

Existen otro tipo de algoritmos que trabajan en el dominio espacial y no alteran la imagen tanto como los que trabajan en el dominio de la transformada. Para relocalizar la marca después de un ataque geométrico estos algoritmos utilizan puntos de referencia. Estos puntos pueden ser insertados en forma de patrón u obtenidos a partir del contenido de la imagen. Generalmente los puntos de referencia se utilizan para determinar cual fue la transformación geométrica utilizada para atacar a la imagen y una vez que se conoce se aplica la transformación inversa para recuperar la forma de la imagen original. Esto, implica más tiempo de procesamiento y no garantiza que se pueda regresar la imagen exactamente a su forma original, hecho que afecta a muchos algoritmos de marca de agua.

La creación de un mecanismo que conserve la calidad de la imagen pero a la vez sea robusto ante ataques de filtrado así como geométricos, resulta atractivo pues sigue siendo un campo abierto a la investigación y podría emplearse en aplicaciones en donde la calidad de la imagen afecte los resultados.

1.2. Objetivo General

- Diseñar e implementar un algoritmo de marca de agua para imágenes en escala de grises que no degrade la calidad de la imagen y sea robusto ante ataques de compresión JPEG, filtrado, rotación y cropping.

1.3. Objetivos Específicos

- Analizar y diseñar procesos de inserción y extracción de la marca de agua que permitan, adicionalmente a la métrica de correlación, reconocer visualmente la marca recuperada. El proceso de inserción debe permitir robustez ante ataques de filtrado y compresión JPEG.

- Analizar y diseñar un método de sincronización de la marca basado en las características de la imagen que permita la inserción/extracción de la marca de agua en el dominio espacial. Además, este método de sincronización le debe permitir al algoritmo recuperar la marca después de un ataque de rotación o cropping.

1.4. Estructura de la Tesis

En el capítulo 2 se describen brevemente las técnicas de Criptografía, Esteganografía y Marcas de agua que son las principales para la seguridad de información digital y representan el fundamento teórico utilizado para el desarrollo del algoritmo propuesto en este trabajo.

En el capítulo 3 se presentan los esquemas de marca de agua basados en secreto compartido visual describiendo sus características y funcionamiento, así como las ventajas que presentan sobre los algoritmos de marca de agua tradicionales. En este capítulo se presentan además los algoritmos basados en secreto compartido visual que se utilizaron como base para el algoritmo propuesto.

En el capítulo 4 se describe el algoritmo propuesto, mencionando las fases de generación de secreto compartido y reconstrucción de la marca, así como el mecanismo de sincronización que permite al esquema ser robusto ante ataques de rotación y cropping.

En el capítulo 5, primero se describen las métricas y el conjunto de imágenes utilizas durante los experimentos. Después, se presentan los resultados obtenidos por el algoritmo ante diferentes ataques a la imagen. En la parte final del capítulo, se discuten los resultados obtenidos.

Finalmente, en el capítulo 6 se presentan las conclusiones a las que se llegaron durante el desarrollo de este trabajo, así como el trabajo futuro.

Capítulo 2

Seguridad en información digital

Con el creciente auge en el intercambio de información digital a través de Internet es necesaria la creación de técnicas que garanticen la seguridad de los datos así como la propiedad de los mismos. Existen diferentes métodos para la protección de datos digitales, sin embargo, la seguridad sigue siendo un campo de investigación abierto. Esto se debe principalmente a que existe una gran diversidad de aplicaciones con diferentes tipos de necesidades en lo referente a confidencialidad, derechos de acceso o integridad.

En este capítulo se describen tres de las técnicas utilizadas en la seguridad de datos digitales cuyos principios son utilizados por la mayoría de los métodos de seguridad existentes. Se hace especial énfasis en la teoría relacionada a las técnicas de marcas de agua utilizadas para proteger los derechos de autor en medios digitales. La teoría presentada en este capítulo se utilizó para crear un algoritmo de marca de agua basado en la técnica de Criptografía Visual que se describirá más adelante.

2.1. Criptografía

La palabra Criptografía proviene de las palabras griegas $krypto$ que significa ocultar y $graphein$ que significa escritura, su definición es: Arte de escribir con clave secreta o de un modo enigmático.

El principal objetivo de la criptográfica es evitar que personas no autorizadas tengan acceso al contenido de archivos, para ello modifica la información original de tal manera que sólo las personas autorizadas puedan reconstruir la información.

Cuando se habla de Criptografía, sólo se involucra el proceso de cifrado, es decir técnicas para codificar la información para hacerla ilegible. Mientras que se denomina Criptoanálisis al conjunto de técnicas que se utilizan para comprometer la seguridad de los datos encriptados.

Un criptosistema de define como una 5-tupla (M, C, K, E, D) [20], donde:

- M representa el conjunto de todos los mensajes sin cifrar (texto en claro).
- C representa el conjunto de todos los posibles mensajes cifrados.
- K representa el conjunto de claves que se pueden emplear en el criptosistema.
- E es el conjunto de transformaciones de cifrado o familia de funciones que se aplica a cada elemento de M para obtener un elemento de C.
- D es el conjunto de transformaciones de descifrado, análogo a E.

Los criptosistemas se dividen en dos tipos fundamentales:

- *Criptosistemas simétricos o de llave privada*. Son aquellos que emplean la misma clave k tanto para cifrar como para descifrar la información. El problema que se presenta en este tipo de criptosistemas es encontrar la forma de transmitir la llave a todos los usuarios involucrados, de manera segura.
- *Criptosistemas asimétricos o de llave pública*. En este tipo de esquemas se emplea una doble clave (Kp,KP), donde Kp se conoce como clave privada y KP se conoce como clave pública. La llave pública KP se utiliza para cifrar mientras que Kp se utiliza para descifrar

En las siguientes secciones se describirán dos métodos criptográficos que se utilizaron durante el desarrollo del presente trabajo de investigación.

2.1.1. Secreto compartido

Generalmente se utiliza una llave secreta para proteger el acceso a la información importante. Sin embargo, si esa llave se pierde, la información queda inaccesible para todos los usuarios.

En 1979 George Blakley [3] y Adi Shamir [28] desarrollaron paralelamente el esquema de secreto compartido que se denomina esquema de umbral m de n, donde $m \leq n$. La idea básica en el esquema es dividir el secreto S en n partes o sombras, proporcionando, de forma secreta, una sombra a cada uno de los n participantes. La división del secreto se lleva a cabo por un administrador y debe hacerse de modo que cualesquiera m participantes diferentes puedan determinar el valor del secreto S sin más que compartir las m sombras que poseen y de tal manera que ningún grupo de $m - 1$ participantes, o menos, pueda obtener información alguna sobre el valor secreto de S.

Este esquema se utiliza para administrar de manera segura la llave que protege a un documento.

2.1.2. Secreto compartido visual

Comúnmente, la seguridad de los algoritmos criptográficos se basa principalmente en la intratabilidad computacional de los problemas matemáticos que utilizan. Sin embargo el uso de técnicas matemáticas se complica cuando el mensaje que se pretende proteger no esta formado por números o letras, sino una imagen binaria. Para resolver este tipo de problemas Naor y Shamir en 1995 propusieron en [21] lo que denominaron secreto compartido visual o VSS por las siglas en ingles de visual secret sharing que se conoce también como criptografía visual.

El esquema VSS propuesto por Naor y Shamir se fundamenta en el esquema umbral m de n (sección 2.1.1) y tiene las siguientes características:

Ahora, el secreto S es una imagen binaria. El esquema codifica cada pixel de la imagen binaria para crear n diferentes versiones del mismo. Cada una de las versiones del

pixel es asignada una imagen sombra por lo que al final se tienen n imágenes sombra. Para recuperar la imagen original se realiza la superposición de m cualesquiera de las imágenes sombra y no es posible recuperar la imagen con sólo $m - 1$.

A continuación se ilustrará mediante un ejemplo como funciona lo que seria el equivalente de un esquema umbral 2 de 2 en Criptografía Visual.

Supóngase que el secreto que se desea proteger es una imagen binaria con el logo del INAOE, en la figura 2.1 se muestra el funcionamiento del esquema de Criptografía Visual para este caso. Primero, la imagen binaria se codifica para crear dos imágenes sombra S_1 y S_2.

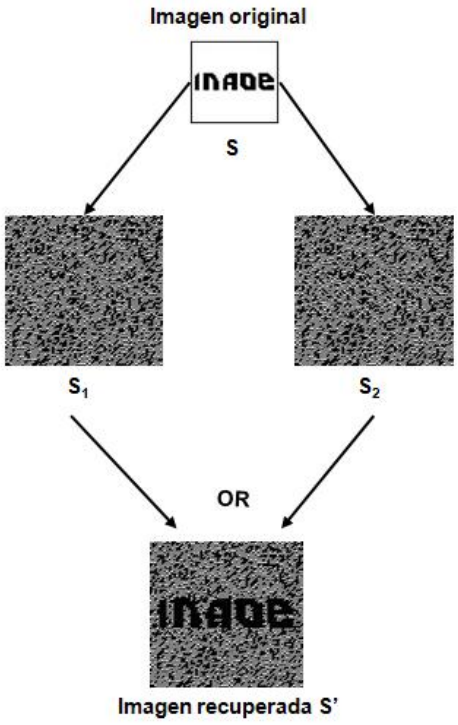


Figura 2.1: Criptografía Visual umbral 2 de 2

Para crear las imágenes sombra, Naor y Shamir propusieron un código en el que cada pixel de la imagen original es representado en las imágenes sombra por bloques de 2×2 con dos pixeles blancos y dos pixeles negros en distintas posiciones. Cuando se quiere representar a un pixel blanco de la imagen I se asignan bloques iguales a cada

sombra, mientras que para representar un pixel negro se crean bloques complementarios. La figura 2.2 muestra un ejemplo del código propuesto por Naor y Shamir.

Para reconstruir la imagen, se toman los correspondientes bloques de 2×2 pixeles en las imágenes sombra y se les aplica un operador OR tomando como 1 lógico los pixeles negros y 0 lógico los blancos (figura 2.2 columna 3). De esta manera, los bloques que son iguales en las imágenes sombra quedarán sin cambio en la imagen reconstruida mientras que los que son opuestos quedaran como bloques de pixeles negros.

Esto da como resultado que las regiones compuestas por pixeles negros en la imagen original S se reconstruyen en S' sin errores mientras que las regiones blancas presentan ruido. No obstante el ruido en la imagen reconstruida S' de la figura fig:csv se puede distinguir fácilmente la leyenda 'INAOE'.

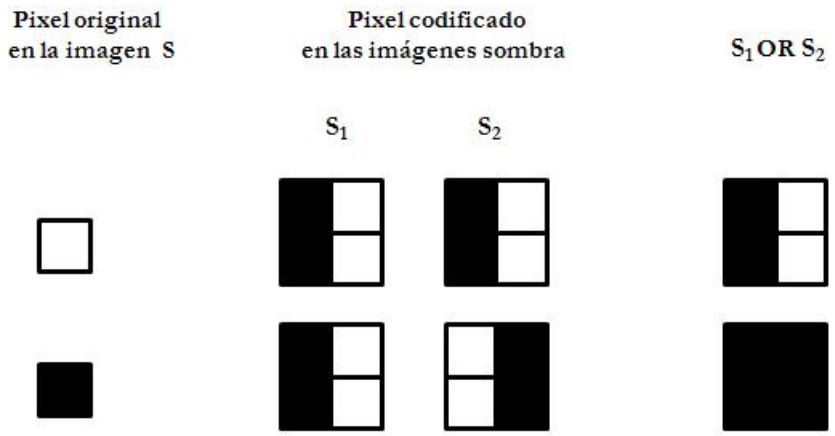


Figura 2.2: Codificación de los pixeles de la imagen en Criptografía Visual.

Como se puede observar, en este tipo de esquema no se requiere ninguna técnica criptográfica para recuperar el mensaje, basta con observar la imagen resultante de la superposición de las imágenes sombra. Obsérvese que las imágenes sombra así como la imagen recuperada S' tienen el doble de tamaño que la imagen original.

2.2. Esteganografía

La Esteganografía es una antigua técnica que se ha utilizado desde civilizaciones antiguas para intercambiar mensajes de manera segura en tiempos de guerra principalmente. La palabra Esteganografía proviene de las palabras griegas $estegos$ que significa cubierta y $graphic$ que significa escritura, por lo que Esteganografía significa literalmente escritura cubierta [14]. El principal objetivo en esta técnica consiste en esconder el mensaje en un medio de manera que no pueda ser detectado.

Los primeros registros que se tienen del uso de esta técnica se encuentran en los relatos del historiador Herodoto. En ellos se describe por ejemplo cómo el general persa Harpagus mandaba mensajes ocultos en el abdomen de liebres muertas que el mensajero transportaba disfrazado de cazador. En otra historia, relata cómo Histaieus rapó a uno de sus esclavos de confianza para tatuar en su cuero cabelludo instrucciones a sus aliados para comenzar la revolución en contra de los Persas y los pueblos mediterráneos.

En la actualidad la Esteganografía utiliza la redundancia de información que existe en los medíos digitales como audio, video, imágenes y texto para ocultar información.

Al conjunto de técnicas que desean comprometer la seguridad de un sistema esteganográfico se les denomina estegoanálisis. En Esteganografía, la seguridad del sistema se quebranta cuando: 1) Se detecta la presencia de información oculta mediante el análisis del *estego-medio* (ataques pasivos). 2) Cuando el medio se altera intencionalmente con la finalidad de destruir posibles mensajes ocultos (ataques activos).

2.2.1. Esquema general de las técnicas esteganográficas

En la figura 2.3 se muestran los elementos que conforman el esquema general de un proceso esteganográfico donde: el mensaje Y es la información que se desea ocultar, ésta puede ser: texto, imagen, video o cualquier otra información que pueda ser representada por una secuencia de bits. El *medio cubierta* o host X es el medio en el que se oculta la información. La *estego-llave* es un dispositivo de seguridad opcional

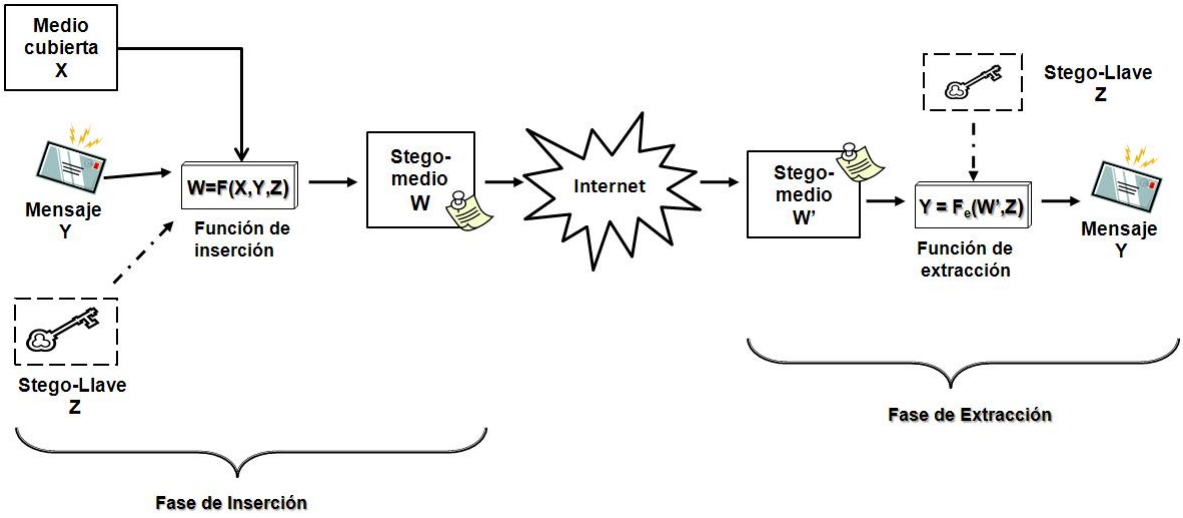


Figura 2.3: Esquema general de la Esteganografía

cuya función es controlar el proceso de inserción de manera que sólo los receptores que la conozcan puedan extraer el mensaje oculto en el medio. La función de inserción F recibe como parámetros al medio cubierta X, el mensaje M y opcionalmente la llave K (*stego-llave*) y como resultado regresa el medio con información oculta Z al que se le denomina *estego-medio*. La función F depende primordialmente de la estructura del medio cubierta y consta de dos partes: en la primera parte se seleccionan los bits en los que se insertará la información y en la segunda parte se inserta la información en los bits seleccionados.

En la fase de extracción se toma el *stego-medio* W' y la *estego-llave* Z como parámetros de la función de extracción F_e para recuperar el mensaje Y. Obsérvese que para este caso se etiquetó al estego-medio con W' y no con W como en la fase de inserción esto se hizo para indicar que pudo sufrir alteraciones a consecuencia de ataques de estegoanálisis o manipulaciones comunes en la red.

En Esteganografía la mayoría de los métodos utilizan la misma estego-llave para insertar y extraer el mensaje. Sin embargo, existen métodos en los cuales las llaves son diferentes al igual que en Criptografía.

2.2.2. Características de los sistemas esteganográficos

Cuando se diseña un sistema esteganográfico se deben considerar tres aspectos importantes:

- *Seguridad* . En los sistemas esteganográficos, la seguridad está ligada directamente a la imperceptibilidad de la información oculta.

- *Capacidad*. Es la cantidad de información que se puede insertar en un medio sin comprometer su estructura.

- *Robustez*. En Esteganografía se refiere a la capacidad de los métodos de recuperar la información después de modificar el estego-medio.

El lograr que un algoritmo cumpla con los tres aspectos mencionados anteriormente no es una tarea sencilla, puesto que al lograr uno de ellos, este compromete a los otros dos en la mayoría de los casos. Por ejemplo, cuando se diseña un algoritmo en el que se requiere ocultar una gran cantidad de información es muy difícil evitar que la información sea detectada y se conserve después de una manipulación.

De acuerdo a las características que se deseen resaltar se pueden distinguir dos lineamientos generales que siguen los esquemas derivados de la Esteganografía [25] (ver figura 2.4).

El objetivo de los esquemas de *protección contra detección* es evitar que se detecte la presencia de un mensaje oculto por lo que se utilizan principalmente para comunicación secreta. Los métodos que siguen este lineamiento deben poner énfasis en la seguridad del esquema, la segunda característica que se considera es la capacidad y por último la robustez.

En los métodos que siguen el lineamiento de *protección contra la extracción* la característica que sobresale es la robustez y su principal aplicación es la protección de la propiedad intelectual en medios digitales. En este tipo de aplicaciones se requiere evitar que la información insertada en el medio sea eliminada mediante manipulaciones co-

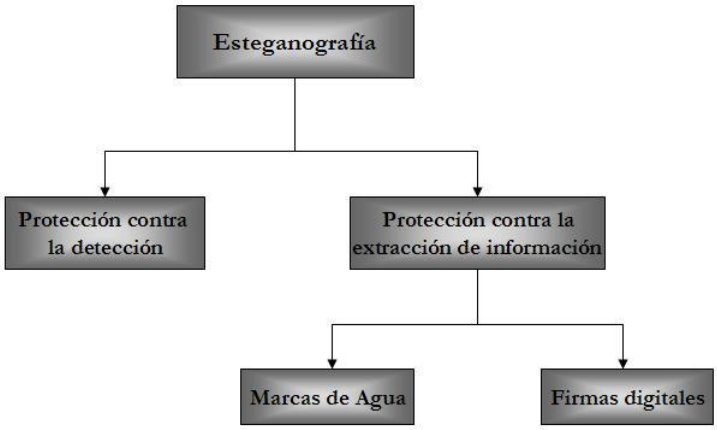


Figura 2.4: Lineamientos de la Esteganografía

munes, como compresión o cambio de brillo en el caso de imágenes digitales, así como en manipulaciones que se hagan intencionalmente para extraer la marca.

Las técnicas que siguen este lineamiento se pueden dividir de manera general en:

- ***Marcas de Agua.*** La información insertada en el medio digital identifica al dueño legítimo. Esta información puede ser texto con el nombre de la empresa, un sonido caracteristico, una imagen con el logo de la empresa etc.

- ***Fingerprinting.*** La información que se inserta se utiliza para identificar al documento de manera única mediante un número de serie, código de barras etc. Su principal aplicación está en el control de copias.

2.3. Diferencia entre Criptografía y Esteganografía

En esta sección se utilizará *el problema de los prisioneros* propuesto por Simmons en 1983 [29] para ilustrar la diferencia entre Criptografía y Esteganografía. Supongamos que Bob y Alice son dos prisioneros que desean elaborar un plan para escapar de la cárcel(Figura 2.5). La única forma que tienen para comunicarse entre sí es a través

de mensajes escritos que son revisados por el custodio Willy. Por lo tanto, Bob y Alice necesitan encontrar la manera de escribir sus mensajes de tal manera que Willy no conozca sus intenciones. Para lograr su objetivo ellos podrían emplear un método criptográfico o uno esteganográfico.

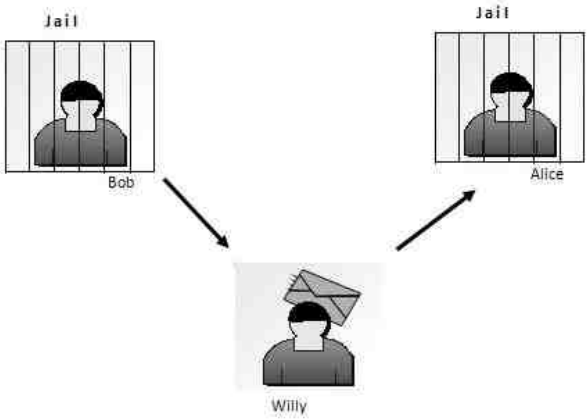


Figura 2.5: El problema de los prisioneros

Si utilizaran Criptografía para escribir sus mensajes Willy ciertamente no sabría el contenido de los mensajes pero sospecharía acerca de las intenciones de los presos al ver una escritura extraña. Más aún, impediría que los mensajes llegaran a su destino y los confinaría sin posibilidad de comunicarse.

En cambio si se ellos deciden utilizar un método de Esteganografía el plan podría estar oculto en un mensaje común que no levante sospecha.

Hasta ahora se han descrito las técnicas de *Criptografía*, *Esteganografía* y la diferencia entre ellas. En la siguiente sección se profundizará en las técnicas de marca de agua que son las técnicas más utilizadas para proteger los derechos de autor en un medio digital.

2.4. Marcas de agua

2.4.1. Orígenes de las técnicas de marca de agua

Registros históricos sigieren que las técnicas de marca de agua tuvieron su origen en Bolonia Italia alrededor de 1280. Deben su origen a los fabricantes de papel que las utilizaban como medio de identificación y como signo del prestigio de su producto. Al principio, las marcas se reducen a figuras geométricas muy simples, tales como cruces, números y signos, sin embargo, a principios del siglo XIV, las marcas de agua se vuelven artísticamente más elaboradas, iban desde escudos familiares hasta representaciones religiosas o réplicas de cuadros famosos, y eran utilizadas como medida de seguridad contra la falsificación. Para el siglo XVIII en Europa y América las marcas de agua se utilizaban como marcas de fábrica en la que registraba el día de fabricación o indicaba las dimensiones originales del papel.

En 1848 se incorpora el relieve y una densidad no uniforme en el papel que produce una escala tonal cuando se ve el papel a trasluz. En 1992 la Interpol reconoce a las marcas de agua como un elemento de seguridad altamente confiable en contra de la falsificación de documentos puesto que al fotocopiarlos, escanearlos o imprimirlos la marca no se reproduce porque ésta se forma directamente en el papel del documento original. En la actualidad las marcas de agua se utilizan comúnmente para evitar la falsificación de billetes.

Con la aparición de la computadoras los documentos que se deseaban proteger no estaban más en papel sino en medios digitales de almacenamiento o en movimiento en las redes. Este nuevo tipo de información digital es más propensa a la falsificación pues no existe diferencia física entre el documento original y la copia, además de que es muy fácil distribuir las copias ilegales a través de Internet para su posterior uso o modificación.

Desde finales de 1990 se dio un extraordinario auge en el desarrollo de técnicas para la protección de los derechos de autor en diferentes tipos de datos digitales pero

con especial énfasis en imágenes, audio y video. Una de esas técnicas retoma la idea de la marca de agua tradicional y la traslada al mundo digital dando paso a la aparición de las marcas de agua digitales.

Una marca de agua digital es un conjunto de bits adheridos al medio que representan información referente al autor o propietario intelectual del mismo. El objetivo principal de las técnicas de marca de agua digitales, es poner de manifiesto el uso ilícito de un cierto servicio digital por parte de un usuario no autorizado.

A partir de este punto y por comodidad se utilizará el término marca de agua aludiendo a las técnicas digitales.

En los últimos años, el medio en el que más se han desarrollado sistemas de marcas de agua son las imágenes digitales

En la siguiente sección se enlistan las principales características que les son exigidas a las técnicas de marca de agua.

2.4.2. Requerimientos de las marcas de agua en imágenes digitales

Al igual que en los algoritmos esteganográficos las tres principales características que se consideran al diseñar algoritmos de marcas de agua son: capacidad, robustez y seguridad pero a diferencia de los métodos esteganográficos aquí se da prioridad a la robustez dejando en segundo término a la capacidad y por último a la seguridad en la mayoría de los esquemas existentes.

Como se verá más adelante, se han desarrollado diversas técnicas de marca de agua con diferentes características de acuerdo a las necesidades de sus aplicaciones. Sin embargo, existen ciertas características que son comunes en la mayoría de ellas. En esta sección se describen esas características comunes.

- **Invisibilidad perceptiva:** Las modificaciones causadas por la inserción de una marca de agua en una imagen digital no deben degradar la calidad de la misma a tal punto que sea perceptible visualmente por el usuario. Sin embargo, las diferencias apenas visibles pueden llegar a percibirse cuando la imagen marcada se

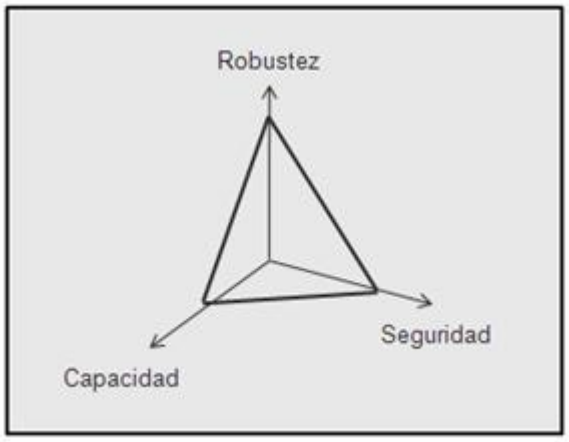


Figura 2.6: Relación entre robustez, capacidad y seguridad en un sistema de marca de agua.

compara directamente con la imagen original.

- **Detección confiable:** Las marcas de agua deben de constituir una comprobación confiable y suficiente de propiedad sobre un producto en particular por lo que la detección de falsos positivos deberían aparecer muy rara vez o nunca. Esto quiere decir que un algoritmo de marca de agua no debe detectar la presencia de la marca de agua en aquellas imágenes en las que no se insertó previamente.

- **Llave asociada:** Para aumentar la seguridad, las marcas de agua deben de estar asociadas a una clave o número denominado llave de la marca de agua. La llave de la marca de agua se usa para insertar, detectar y remover una marca de agua. Se sugiere que la llave sea privada y que sólo la conozca el dueño legal.

- **Detección/Búsqueda automatizada:** Dado que el principal uso de las marcas de agua es detectar las copias de la imagen que se distribuyen ilegalmente en la red, se desea que los algoritmos de marca de agua se puedan combinar fácilmente con procedimientos de búsqueda que exploren cualquier dominio accesible públicamente.

- **Invisibilidad estadística:** Las marcas de agua no deben revelarse usando métodos estadísticos.

- **Múltiples marcas de agua:** Cuando se utiliza la marca de agua para proteger la propiedad de un medio que es transferido a otro propietario es deseable que el sistema sea capaz de insertar múltiples marcas de agua en una sola imagen y que cada una de las marcas insertadas sean detectadas usando su correspondiente llave.

- **Robustez:** La marca de agua debe de preservarse después de las modificaciones que se realicen a la imagen, ya sea de manera deliberada usando ataques o por manipulaciones comunes del usuario tales como compresión, filtrados para la extracción de ruido, cambio de tamaño, cambios en el contraste, brillo. Esto es posible siempre y cuando la calidad de la imagen huésped permanezca dentro de límites aceptables. Además de permanecer después de alguna modificación en la imagen, las marcas de agua deben ser estadísticamente irremovibles, esto quiere decir que un análisis estadístico a la imagen no debe significar ventaja al atacante.

2.4.3. Esquema general de un sistema de marcas de agua

Todo sistema de marcas de agua consiste de al menos tres fases fundamentales: Generación, inserción y detección/extracción de la marca de agua. En las siguientes secciones se describen de manera general esas etapas:

Generación e inserción de la marca de agua

Durante la fase de generación de la marca de agua, se representa la información del propietario en un formato adecuado al algoritmo. Por ejemplo, algunos algoritmos que trabajan en el dominio de las frecuencias convierten las imágenes con logos que identifican al propietario en señales unidimensionales antes de la inserción.

Cuando se requiere que la marca de agua sea invisible se pueden considerar las características de la imagen en la fase de generación de la marca para codificar la imagen para que se adapte mejor a las características de la imagen y sea imperceptible.

Durante la fase de generación se puede usar una llave para codificación de la marca.

En la figura 2.7 se muestran las fases de generación e inserción de la marca de agua.

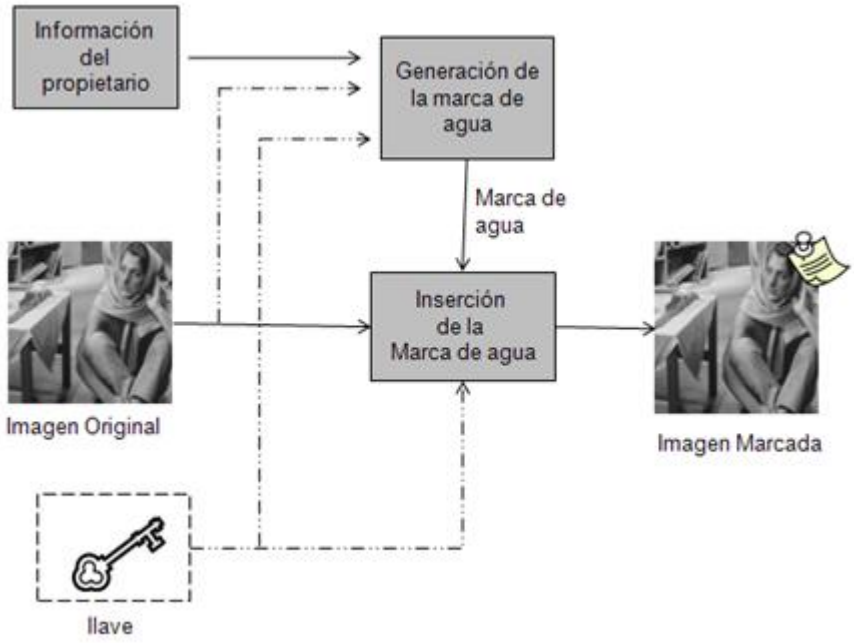


Figura 2.7: Fase de generación e inserción en un sistema de marca de agua

Durante la fase de inserción se toma la imagen original I y la marca de agua W y se genera una imagen I' que contiene a la marca. De acuerdo con la figura 2.7 se puede hacer uso de una llave durante este proceso. La llave aquí se utiliza generalmente para determinar la posición de la marca dentro de la imagen.

Formalmente, el proceso de inserción se puede representar por la ecuación (2.1).

$$I' = FI(I, W, K) \tag{2.1}$$

Donde, I es la imagen original, W es la marca que se desea insertar, FI es la función de inserción y K es la llave.

Detección/Extracción de la marca de agua

En la figura 2.8 se presenta el esquema de detección/extracción de la marca de agua. Como se puede apreciar para detectar o recuperar la marca de agua se puede usar la imagen original y/o una llave que indique ya sea la posición de la marca en la imagen o que se utilice para decodificar el contenido de la marca. En la salida de este proceso, dependiendo de la aplicación, se pueden presentar dos situaciones para: 1) que sólo se requiera determinar si la marca de agua está presente en una imagen o no sin necesidad de mostrarla en su forma original. Bajo este esquema la función de detección se puede ver como una función binaria cuyo valor 1 indica la presencia de la marca. 2) que se desee extraer la información de marca de agua y no sólo indicar si está o no.

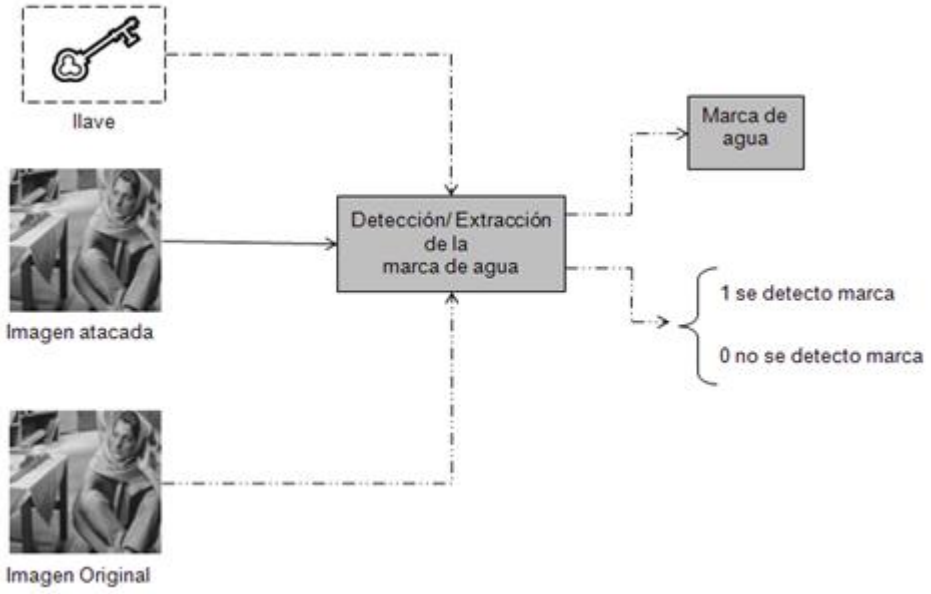


Figura 2.8: Fase de Detección/Extracción en un sistema de marca de agua

Formalmente, el proceso de detección de la marca de agua se puede definir mediante la ecuación (2.2).

$$FD(I', W, K, I) = \begin{cases} 1, \, if \, W \, existe \, en \, I'. \\ 0, \, if \, W \, no \, existe \, en \, I'. \end{cases} \tag{2.2}$$

Donde, I' es la imagen en la que se desea detectar la marca, I es la imagen original, W es la marca que se desea insertar, FD es la función de detección y K es la llave.

Por otro lado, la ecuación (2.3) define formalmente al proceso de extracción.

$$W' = FE(I', K, I) \tag{2.3}$$

En donde I' es la imagen de la que se desea extraer la marca, I es la imagen original, FE es la función de extracción, K es la llave y W' es la marca de agua recuperada.

Algunos autores consideran a la fase de detección de la marca como la más importante en un esquema de marca de agua, dado que la salida del detector debe proporcionar una evidencia sustancial y suficientemente confiable, de la pertenencia de los derechos de autor en la imagen. Este requerimiento sugiere que la detección de la marca de agua debe de ser un procedimiento públicamente conocido y globalmente aceptable. Durante esta fase se pueden presentar dos tipos de errores que se deben evitar cuando se diseña un algoritmo de marca de agua. Estos errores son:

- **Falso positivo:**La marca de agua es detectada a pesar de que no existe en la imagen.
- **Falso negativo:**La marca de agua no es detectada en la imagen a pesar de estar presente.

2.4.4. Clasificación de las marcas de agua

Existen múltiples clasificaciones de las marcas de agua de acuerdo a diversos parámetros como la función de inserción que utilizan, espacio de inserción, robustez, la función de extracción que utilizan, perceptibilidad e incluso algunos autores hacen una clasificación de acuerdo a la aplicación en que se utiliza y al medio portador de la marca.

De acuerdo al grado de resistencia que tenga la marca de agua ante las modificaciones que se le hagan a la imagen, los esquemas de marca de agua se pueden clasificar en:

- **Robustas.** Un esquema de marca de agua robusto es aquel que es capaz de recuperar/detectar la marca de agua insertada de una imagen que ha sido sometida a modificaciones ya sea intencionalmente o no. Se utilizan principalmente en la protección de derechos de autor.

- **Frágiles.** Se destruye o altera al modificar la imagen en la que es insertada. Esto, con la finalidad de poner en evidencia las alteraciones realizadas en la imagen. Este tipo de esquemas se utilizan principalmente para autentificación de la imagen.

- **Semi-frágiles.** Este tipo de marcas de agua se destruye sólo cuando el ataque es intencionado y se preserva cuando se trata de una manipulación común de la imagen [5, 18]. El mayor reto de este esquema es determinar cuando una modificación en la imagen se considera maliciosa y cuando no. Usualmente en este esquema se utilizan las propiedades o relaciones entre coeficientes de alguna transformada en la imagen [17].

De acuerdo a como percibe la marca de agua un usuario las marcas de agua pueden ser:

- **Visibles.** En este tipo de esquemas, el logo de la empresa o la información del propietario legítimo es percibida a simple vista por el usuario sin necesidad de aplicar algún filtro o hacer un análisis a la imagen. Este tipo de esquemas se utilizan principalmente en las fotografías de la revistas, billetes y en las transmisiones de programas de televisión.

- **Invisibles.** Este tipo de esquemas aprovechan las limitaciones del *sistema visual humano* o *SVH* para insertar la marca de agua de manera que ésta sólo pueda ser detectada mediante un análisis de la imagen y no por los usuarios normales. Un ejemplo de este tipo de esquemas está en [22].

- **Duales.** Este esquema es la combinación de una marca visible e invisible en la misma imagen con la finalidad de tener un respaldo invisible de la marca visible.

Muchos autores consideran que la clasificación principal de los esquemas de marca de agua es de acuerdo al tipo de dominio en el que se inserta la marca. De acuerdo a ese criterio las marcas de agua se pueden clasificar en:

- **Dominio espacial.** Los esquemas que trabajan en el dominio espacial insertan la marca de agua directamente en los pixeles de la imagen.

- **Dominio de la transformada.** Transforman la imagen a un nuevo espacio e inserta la marca en los coeficientes que representan a la imagen transformada. Ejemplos de transformadas utilizadas por estos esquemas son: transformada Discreta de coseno o DTC, Wavelet [2], Furier, y Ortogonal [27].

Las técnicas que trabajan en el dominio espacial generan menos degradación en las imágenes cubiertas. Sin embargo, estas técnicas son frágiles ante ataques geométricos o ante el filtrado de la imagen marcada. Por otro lado, los esquemas que trabajan en el dominio de transformada presentan mayor robustez ante ataques a la imagen marcada pero al contrario de las técnicas de dominio espacial, introducen más degradación en la imagen.

De acuerdo al tipo de información que necesitan para extraer la marca de agua insertada los esquemas de marca de agua se clasifican en:

- **Públicos.** No se necesita información adicional para extraer-detectar la marca de agua insertada.

- **Privados.** Requieren de la imagen original para extraer/detectar la marca insertada.

Las ventajas que presentan los algoritmos públicos sobre los privados es que pueden recuperar la marca de agua después de aplicar grandes deformaciones a la imagen. Sin embargo los esquemas que requieren la imagen original no se pueden combinar fácilmente con sistemas de búsqueda en la web o en librerías digitales [31].

2.4.5. Ataques a las marcas de agua

Las investigaciones sobre las técnicas de marcas de agua han progresado en dos direcciones: en una se investigan y desarrollan nuevas tecnologías de marcas de agua y en la otra se desarrollan diferentes ataques que comprometan la seguridad de los sistemas. Los ataques a las marcas de agua se pueden clasificar en ataques sobre esquemas de marcas de agua robustos y ataques sobre esquemas de marcas de agua frágiles.

Clasificación de los ataques a los esquemas de marcas de agua

En la actualidad existen una gran cantidad de ataques intencionales a los sistemas de marcas de agua pues cada uno se enfoca a uno de los distintos tipos de marcas de agua. En [16] se presenta una clasificación general de los ataques que se aplican a los sistemas de marca de agua robustos. Esta clasificación se muestra en la figura 2.9.

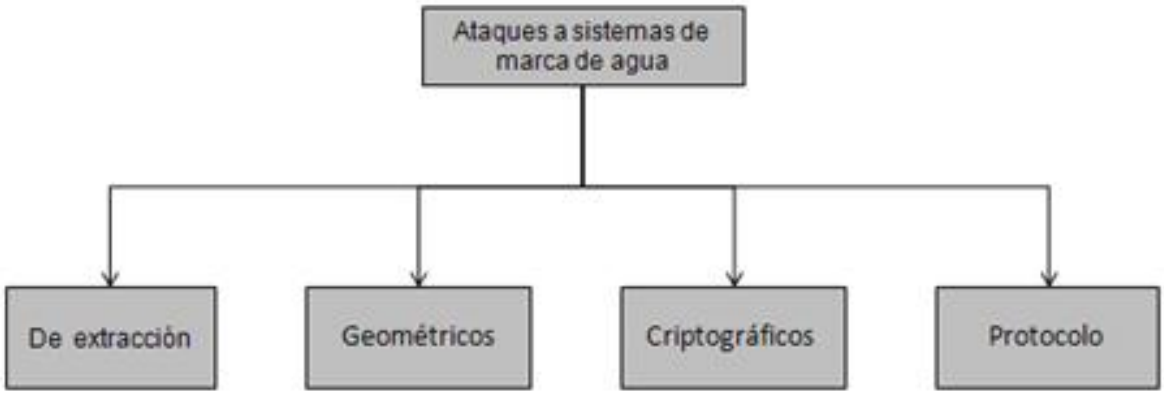


Figura 2.9: Clasificación de los ataques a marcas de agua.

Ataques de extracción

Los ataques de extracción tienen por objetivo remover completamente la marca de agua de la imagen. Este enfoque considera a la marca de agua insertada como ruido en la imagen por lo que utilizan técnicas para la eliminación de ruido en la imagen.

Ataques geométricos

A diferencia de los ataques de extracción, los ataques geométricos no pretenden remover la marca de agua insertada. El ataque consiste en distorsionar la imagen de tal manera que le sea imposible al detector localizar la marca en la imagen atacada.

Ataques criptográficos

Estos ataques son muy similares a los ataques usados en Criptografía y pueden ser de diferente naturaleza. Existen por ejemplo los ataques de fuerza bruta que intentan encontrar la llave de la marca de agua a través de una búsqueda exhaustiva. Otro ejemplo de este tipo de ataques es el llamado Oracle que crea una imagen sin la marca de agua analizando el dispositivo de detección

Ataques de protocolo

Los ataques de protocolo tienen por objetivo la destrucción de la información insertada a través de una manipulación de datos a nivel local o global.

2.4.6. Aplicaciones de las marcas de agua

Tecnologías de marcas de agua digitales para la administración de los derechos de autor

La aplicación más común de las marcas de agua es la protección de los derechos de autor. Aquí una marca escondida e invisible es insertada en la imagen.

Uso de marcas de agua digitales para la autentificación

El objetivo de esta técnica no es proteger los contenidos de un posible robo o copia sino proveer un método de autentificación de la imagen asegurando la integridad de la misma.

Tecnologías de marcas de agua reversibles y visibles para la distribución electrónica

A diferencia de las tecnologías de las marcas de agua digitales descritas anteriormente, las marcas de agua reversibles y visibles están disponibles como producto comercial. Esta forma única de tecnología de marcas de agua producida por IBM permite insertar en el contenido de los propietarios una forma visible o logo como puede ser el logo de la compañía en la parte superior de la imagen. La marca es removida solamente con la aplicación de una apropiada llave de desencriptamiento y un software removedor de marcas de agua.

Tecnologías de marcas de agua para rastreo

Las marcas de agua pueden usarse para rastrear el uso de contenido digital. Cada copia del contenido digital puede tener una marca única con metadatos que especifican el contenido de usuarios autorizados. Tales marcas de agua pueden ser usadas para identificar a los usuarios que generan réplicas de contenido de manera ilegal. La técnica de marca de agua usada para el rastreo es llamada fingerprint (huella digital).

Tecnologías de marcas de agua para monitoreo de transmisiones

Las marcas de agua pueden ser usadas para monitorear el contenido de las transmisiones de televisión y radio. Las compañías publicitarias pueden usar sistemas que detecten la transmisión de publicidad con propósitos de facturar de acuerdo al número de veces que se identifique la marca de agua en el contenido de la programación.

En este capítulo se hizo una revisión de las técnicas utilizadas para la seguridad de la información digital, en especial de las técnicas de marca de agua en imágenes digitales. Se presentaron las diferentes clasificaciones que se hacen de las marcas de agua con base en diversos parámetros, los requerimientos que se desean de un algoritmo de marca de agua y los diversos ataques a los que son sometidas las imágenes marcadas. Finalmente se presentaron algunas de las aplicaciones de las técnicas de marca de agua.

Capítulo 3

Esquemas de marca de agua basados en secreto compartido visual

En años recientes se han propuesto esquemas de marca de agua para la protección de derechos de autor en imágenes digitales basados en el esquema criptográfico de secreto compartido visual propuesto por Naor y Shamir descrito en la sección 2.1.2.

De los primeros trabajos en utilizar este esquema está el propuesto por Hou&Chen en el 2000 [34]. En este trabajo se utiliza criptografía visual para crear dos imágenes sombra de la marca que identifica al propietario, una de ellas se inserta en la imagen y la otra se le denominó *imagen compartida* y se guarda para usarla como llave para revelar la marca. Para revelar la marca basta con superponer la imagen compartida y imagen marcada.

El trabajo de Hou&Chen usaba el esquema de secreto compartido visual pero seguía el esquema tradicional de marca de agua en el que la imagen es alterada durante el proceso de inserción.

Los trabajos posteriores al propuesto por Hou&Chen incorporaron un nuevo esquema de marca de agua que a diferencia de los esquemas tradicionales no incrusta ningún tipo de información en la imagen. La idea en este esquema es usar los principios de la técnica de criptografía visual de Shamir para crear una imagen compartida a partir de la

combinación de características de la imagen original con la marca, de tal manera que al combinar la imagen compartida con las características de la imagen analizada se revela la marca de agua. El esquema consiste en dos fases principales:

1. ***Fase de generación del secreto compartido.*** En esta fase la imagen compartida S_2 se codifica a partir de las características de la imagen y la marca que se desea ocultar (figura 3.1). Esta fase reemplaza a la fase de inserción en los esquemas tradicionales de la marca de agua.

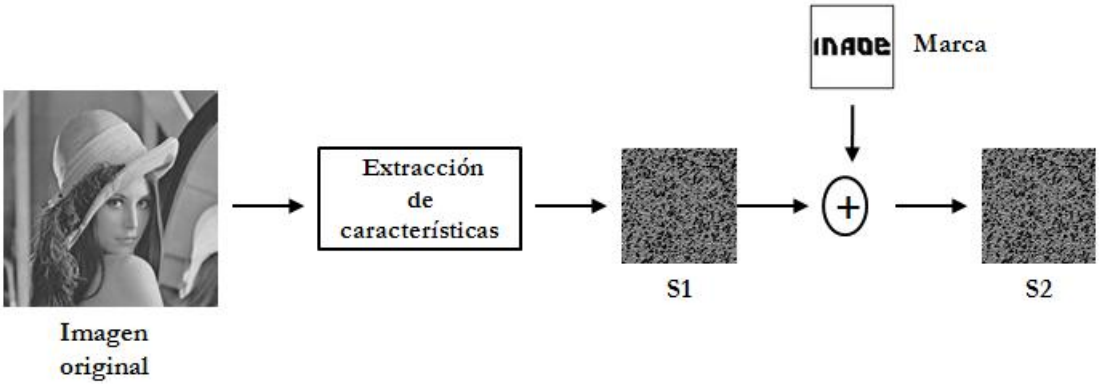


Figura 3.1: Fase de generación del secreto compartido.

2. ***Fase de reconstrucción de la marca.*** Durante esta fase se requiere de la imagen compartida $S2$ así como, la imagen $S1$ obtenida a partir de la imagen analizada para revelar la marca (ver figura 3.2). Esta fase es equivalente a la fase de extracción de la marca de agua en los esquemas tradicionales.

Los algoritmos bajo el esquema de secreto compartido visual presentan las siguientes ventajas sobre los esquemas tradicionales:

1. La calidad de la imagen no se ve alterada por la inserción de la marca de agua.

2. Al no insertar información extra en la imagen, la marca de agua no puede ser detectada por ningún análisis estadístico ni extraída por un ataque.

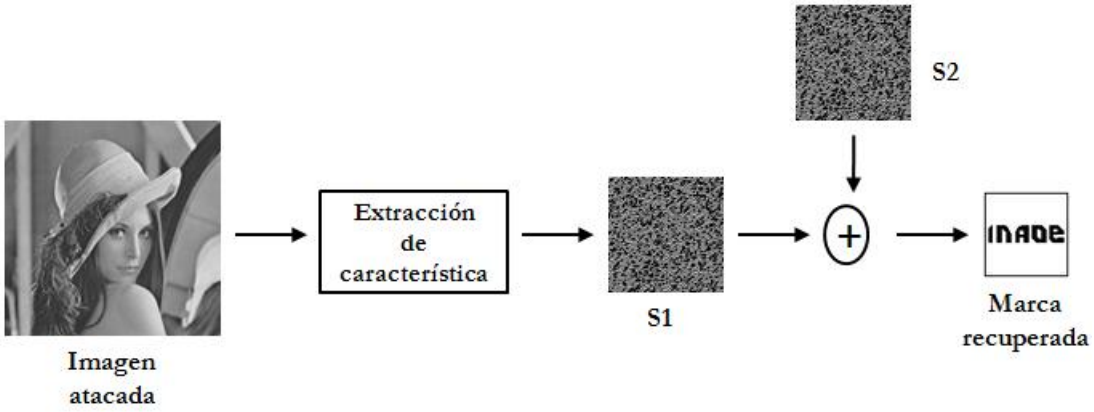


Figura 3.2: Fase de reconstrucción de la marca.

3. La seguridad del sistema se incrementa pues para recuperar la información se requiere la imagen compartida.

El algoritmo de marca de agua propuesto en este trabajo de investigación sigue el esquema basado en secreto compartido. El algoritmo propuesto se describirá en el capítulo 4. Pero antes, en la sección 3.1 se presentarán los trabajos relacionados en los que se basa.

3.1. Trabajos relacionados

En 2004 Shu-Fen&Ching-Sheng en [30] propusieron un algoritmo basado en el esquema de secreto compartido visual que utiliza el código de truncado de bloques BTC (Block Truncation Code) para crear una imagen binaria a partir de la original. Después la imagen binaria generada se combina con la marca binara que identifica al propietario mediante el operador XOR. La imagen resultante se guarda y se utiliza posteriormente en la fase de reconstrucción de la marca. Los resultados reportados en este trabajo demuestran que la marca se puede recuperar después de operaciones comunes en las imágenes digitales tales como: aumento o disminución de brillo, contraste, difuminación, compresión, ruido y al ataque de cropping. Sin embargo, no especifica

el método de sincronización utilizado para recuperar la marca después del ataque de cropping por lo que se asume que esto se hizo de manera manual.

En el año 2005 Lou et al. en [9] propusieron un código para la generación de la imagen compartida, basado en las propiedades de la transformada wavelet y los valores de los píxeles en la marca, las dimensiones de la imagen compartida resultante son el doble de la marca original. El esquema funciona en imágenes en escala de grises y reporta los resultados obtenidos ante ataques de compresión JPEG, inserción de ruido sal y pimienta, realce de contraste e incluso afirma que la marca de agua se puede reconstruir después de rotar la imagen.

En el 2005 Hsieh et al. en [11] presentan un algoritmo que al igual que el esquema anterior trabaja con propiedades de la transformada, aunque de distinto modo, para la generación de la imagen compartida. Este esquema trabaja imágenes a color para lo que propone una modificación al muestreo tradicional 4:2:2 del espacio YCbCr, por considerarlo poco robusto ante los ataques. Los resultados reportados por este trabajo son a ataques de compresión JPEG, realce de contraste, difuminación escalamiento de la imagen y adición de ruido.

En los esquemas mencionados anteriormente, a diferencia del propuesto por Hou&Chen, ninguna de la imágenes sombra se inserta en la imagen original ni es necesaria la imagen original para reconstruir la marca. En ninguno ellos existe degradación de la imagen ocasionada por la fase de inserción lo cual cumple uno los objetivos de este trabajo, sin embargo, no cuentan con una fase de sincronización por lo que los esquemas no son robustos a ataques tales como rotaciones, traslaciones e incluso cropping. En el único trabajo que se hace mención de un ataque de rotación es en [9], sin embargo, al no contar con un mecanismo de sincronización que relocalice a la marca después de la rotación se asume que el ataque reportado se realizó de la siguiente manera: la imagen se rota n grados e inmediatamente después se vuelve a rotar $-n$ grados para regresarla a su posición original y poder reconstruir la marca. Este tipo de pruebas son habitualmente utilizadas para probar la resistencia de los algoritmos ante rotaciones. Sin embargo en la práctica no se conoce a priori el número de grados que ha sido rotada una imagen.

Por lo que consideramos que los algoritmos que ocupan este tipo de test para evaluar su algoritmo no prueban que el algoritmo sea resistente a rotaciones si no más bien a una especie de filtro pasa altos pues a pesar de que la imagen se regresa a su posición original sufre una alteración debido a la transpolación realizada durante las rotaciones.

Hasta ahora se han presentado los trabajos relacionados con marcas de agua usando secreto compartido visual y que sirven de base para el algoritmo propuesto. En las siguientes secciones se hablará del problema de sincronización al que se enfrentan todos los esquemas de marcas de agua que buscan ser robustos a ataques geométricos así como los diferentes tipos de métodos que se utilizan para resolverlo.

3.2. El problema de sincronización

Cuando se utilizan técnicas en las que se inserta información adicional en imágenes digitales para tener un control de las copias existentes o proteger el derecho de autor se debe asegurar que ésta se pueda recuperar. Si no se produjera ningún cambio en la imagen después de la inserción se conocerían las coordenadas exactas dentro de la imagen en las que el proceso de extracción puede encontrar la información. Sin embargo, cuando se transmite vía internet, la imagen puede ser modificada y la posición de la información insertada se puede perder. Esto es conocido como *el problema de sincronización*. Así, los métodos de marca de agua deben implementar un mecanismo que permita hacer coincidir la imagen analizada con la imagen original o calcular la nueva posición de la marca en la imagen (figura 3.3).

Para resolver el problema de sincronización en imágenes se han desarrollado diversos trabajos cuya clasificación y principales características se describen en la siguiente sección.

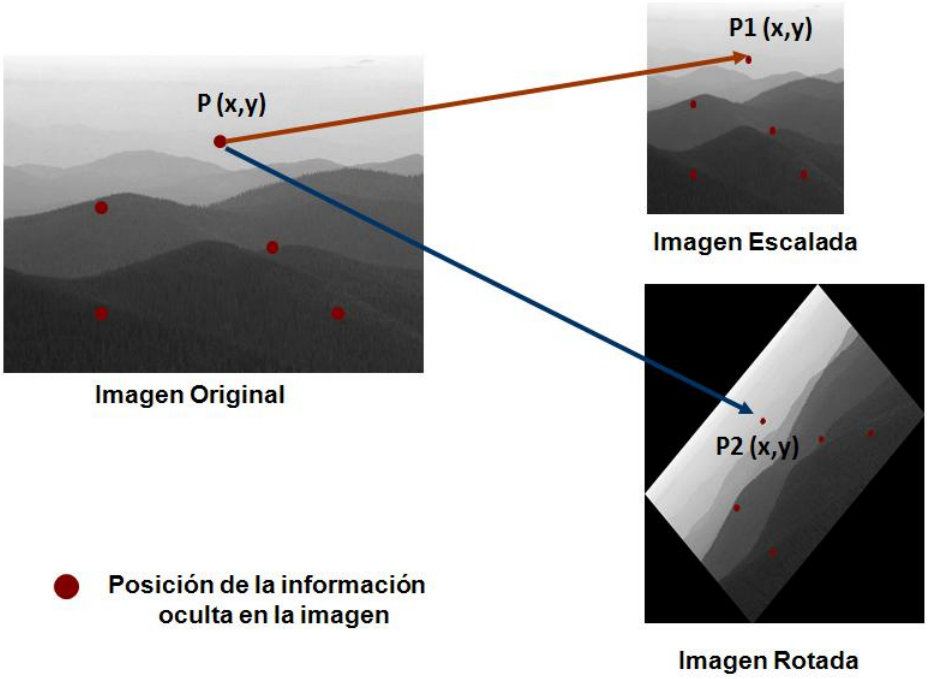


Figura 3.3: Problema de sincronización: después de rotar la imagen se tiene que determinar la nueva posición de la información oculta

3.2.1. Clasificación de los esquemas de sincronización

Las técnicas de sincronización automática se pueden clasificar como se muestra en la figura 3.4

Métodos clásicos de sincronización

De las técnicas clásicas de sincronización automática se distinguen cuatro clases [24] de acuerdo a los elementos que utilizan como referencia para la sincronizar la marca, estos elementos son:

- *Secuencias periódicas.* Este tipo de técnicas insertan la información de la marca en forma redundante en la imagen digital con el fin tener mayores probabilidades de recuperarla.

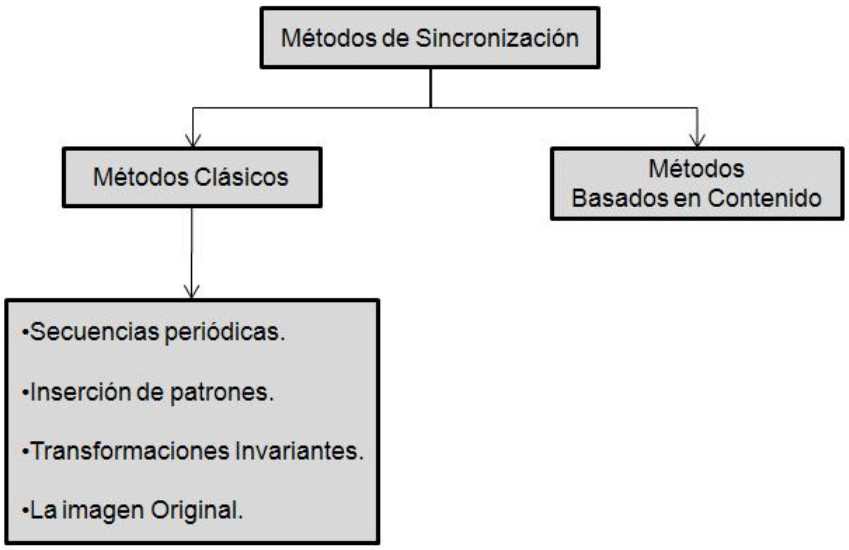


Figura 3.4: Clasificación de los métodos de sincronización

- *Inserción de patrones.* En estos métodos se insertan referencias artificiales que son recuperadas y comparadas con el patrón original para identificar la transformación afin que permita regresar la imagen a su posición original. La principal ventaja de estos métodos es que se pueden identificar tanto transformaciones locales como generales. Sin embargo se han desarrollado técnicas para remover los patrones insertados.

- *Transformaciones Invariantes.* En este tipo de métodos la imagen es transformada a un dominio que no es afectado por cambios geométricos y es ahí en donde se inserta la información deseada. Usando este tipo de sincronización, la imagen no se regresa a su posición original[6, 26].

- *Uso de la imagen Original.* La imagen original se utiliza para identificar y corregir los cambios sufridos después de aplicar ataques geométricos. Una vez corregida la imagen se extrae la marca.

Métodos basados en contenido

Como se puede apreciar, las técnicas clásicas de sincronización recurren a elementos externos a la imagen para sincronizar la información oculta, sin embargo en los últimos años se han desarrollados trabajos basados en el contenido de la imagen. En este tipo de técnicas, la posición de la información a insertar no se relaciona con coordenadas en la imagen si no al contenido de la imagen [24].

Generalmente los métodos basados en contenido usan características invariantes de la imagen para sincronizar la marca.

Las características pueden ser invariantes a nivel global, como en [7] que utiliza el centro de gravedad para sincronizar la imagen. El problema con este trabajo robusto ante rotaciones, es que se deben determinar los grados a los que ha sido rotada la imagen. Además de que no funciona cuando a la imagen se le aplica un ataque cropping o se hace una traslación.

Mientras que en [7, 24] se utilizan características locales invariantes para la sincronización de la marca después de un ataque geométrico a la imagen.

La ventaja de utilizar características invariantes locales es que permite a los esquemas ser robustos a ataques cropping y a traslaciones de la imagen.

En las siguientes secciones se presenta la teoría referente a características invariantes o puntos característicos de la imagen. Esta teoría se utilizó para crear una fase de sincronización del algoritmo propuesto.

3.3. Puntos característicos de imágenes digitales

Las características invariantes de una imagen son aquellas que se preservan después de aplicar un conjunto de alteraciones tales como: escalamiento, rotación, cambios de iluminación etc.

Sea $I = I(i, j), 0 \leq i \leq N, 0 \leq j \leq M$ una imagen en escala de grises. El valor $I(i, j)$ representa el nivel de gris en la coordenada (i, j). Con G denota un conjunto de

trasformaciones con elementos $g \in G$ que actúan sobre las imágenes. Así una imagen transformada se denota por gI [4].

Una característica invariante es una función $F(I)$ que es invariante a la acción del conjunto G de transformaciones sobre I. Esto se define mediante la ecuación (3.1).

$$F(gI) = F(I) \tag{3.1}$$

Estas características pueden ser invariantes a nivel global, cuando describen a la imagen como un todo sin importar su contenido, o a nivel local, cuando ofrecen información detallada del contenido de una imagen, razón por la cual son las más utilizadas. Las características invariantes locales son comúnmente llamadas ***puntos o regiones de interés***.

La propiedad principal que debe cumplir un punto de interés *bueno* es la ***repetitividad***, esto es, dadas dos imágenes I, I' de la misma escena tomadas de tal forma que, I' presenta variaciones geométricas y/o de iluminación con respecto a I, un alto porcentaje de los puntos característicos detectados en I deben aparecer en I'.

Los métodos que se utilizan para extraer los puntos de interés de una imagen, se pueden clasificar en:

- ***Métodos basados en bordes.*** En estos métodos los puntos característicos se determinan a partir de la información de curvatura. En un contorno: puntos de elevada curvatura (discontinuidad en la orientación del vector gradiente). En uniones: punto de intersección de varias aristas, normalmente con un significado geométrico (oclusión, etc).

- ***Métodos basados en intensidad.*** Este tipo de métodos utiliza directamente la intensidad del nivel de gris alrededor del pixel candidato para encontrar evidencia de que hay una esquina. Ejemplos de este tipo de métodos son los detectores de puntos Moravec, Kitchen&Rosenfeld, Harris y Foerstne.

- ***Métodos bioinspirados.*** Estos métodos se basan en procesos biológicos, ejemplos

de este tipo de métodos es el propuesto por Heitger et al. [10] inspirados en mecanismos neurobiológicos y el propuesto por Reisfeld et al. [8] basado en experimentos psicofísicos.

- ***Métodos basados en modelos.*** Esta clase de métodos se basan en modelos teóricos de puntos característicos, así por ejemplo se tienen modelos de diferentes intersecciones de múltiples contornos. Los modelos son parametrizados de tal manera que la señal teórica es similar a la real. Las diferencias se minimizan durante la fase de detección. Para lograr una mayor precisión, se debe dar una buena inicialización de los parámetros.

- ***Métodos multi-escala.*** Este tipo de métodos extraen puntos de interés en varias escalas para representar a una imagen, esto incrementa la repetitividad de los detectores de puntos. Sin embargo, cuando una estructura de la imagen está presente en un rango de escalas, el detector encuentra varios puntos que representan a la misma esctructura pero con pequeñas diferencias en posición y escala por lo que disminuye la probabilidad de re-localizar los puntos después de una modificación.

- ***Métodos invariantes a escala.*** Este tipo de métodos utilizan el espacio de las escala, que es un conjunto de imágenes convolucionadas con la función Gauss diferentes desviaciones estándar, para encontrar puntos invariantes. Ejemplos de este tipo de métodos son los detectores Harris-Laplacian [23] y SIFT [19].

3.3.1. Descriptores de puntos característicos

Los descriptores de puntos característicos proporcionan información que permite distinguir a un punto en etapas posteriores. Por lo tanto, los descriptores deben obtener una completa pero a la vez compacta descripción del punto característico que pueda ser medida mediante funciones de similaridad. Una de las métricas más utilizada para calcular la similaridad es la *correlación cruzada*. Esto, para poder distinguir y localizar

a un punto característico específico en dos imágenes.

Existe una gran diversidad de descriptores que enfatizan diferentes propiedades de la imagen tales como: la intensidad de los pixeles, color, textura, contornos etc. El descriptor más simple es un vector que contiene los pixeles de la imagen. Sin embargo, un *buen* descriptor debe tener las siguientes propiedades:

- Dadas dos características locales cuyos patrones de intensidades sean similares o representen al mismo objeto en la imagen, sus respectivos descriptores deben ser lo más parecidos posible.

- El descriptor debe ser compacto es decir, no debe contener información redundante ni datos altamente correlacionados. Un descriptor grande es más difícil de manejar, puede contener información irrelevante y requiere más capacidad de almacenamiento. Mientras que descriptor compacto no sólo reduce la capacidad de almacenamiento, sino que reduce el tiempo de procesamiento.

- Puesto que una imagen puede tener arriba de 1000 puntos característicos el descriptor debe ser eficiente.

3.3.2. Aplicaciones de las características invariantes

Los métodos de extracción de características invariantes de una imagen, se han aplicado a diferentes campos tales como: reconocimiento de texturas, clasificación de escenas, navegación de robots, minería visual de datos, entre otros.

En el caso de marca de agua, existen algoritmos que utilizan puntos característicos para localizar las posiciones de las marcas de agua después de rotar o trasladar la imagen [12, 24].

En el algoritmo propuesto en este trabajo, se utilizan los puntos característicos de la imagen así como sus descriptores, como referencia para reconstruir la marca de agua después de que la imagen es rotada o trasladada.

En este capítulo se presentaron los esquemas de marca de agua basados en secreto compartido. La ventaja principal de este tipo de esquema sobre los esquemas tradicionales es que no alteran la calidad de la imagen. Esto es por que no insertan la marca de agua directamente en la imagen.

En este capítulo también se hizo un breve resumen de los algoritmos utilizados como base para el diseño del algoritmo propuesto en este trabajo y que será presentado en el siguiente capítulo.

Capítulo 4

Algoritmo propuesto

En este capítulo se propone un algoritmo de marca de agua basado en secreto compartido. Aunque el algoritmo propuesto sigue el esquema de los algoritmos presentados en el capítulo 3, éste incorpora una fase de sincronización para obtener robustez ante rotaciones, translaciones y cropping. La fase de sincronización se hace con base en características locales de la imagen que se desea proteger (ver secciones 3.2, 3.3 y 3.4).

4.1. Fase de generación del secreto compartido

Para la generación del secreto compartido se requieren las siguientes entradas:

- Imagen I en escala de grises de dimensiones $N \times M$ pixeles.
- Marca W en forma de imagen binaria de $S \times S$ pixeles.

Al final de esta fase se tienen la siguiente salida:

- El secreto compartido Shd formado por la imagen compartida binaria Is de $N \times M$ pixeles y el archivo Ad con los descriptores de puntos característicos usados como referencia.

La fase de generación del secreto compartido se conforma de dos etapas principales:

1. Extracción de puntos característicos y descriptores de la imagen.

2. Creación de la imagen compartida.

4.1.1. Extracción de puntos característicos y descriptores

Para lograr robustez ante los ataques geométricos de rotación, traslación y cropping, el algoritmo propuesto incorpora una fase de sincronización que consiste en extraer características locales invariantes o *puntos característicos* que sirven como referencia para codificar la imagen compartida.

El algoritmo seleccionado para extraer los puntos característicos fue el propuesto por Lowe at.el. en el 2004 [19] denominado $SIFT$ o *Scale-invariant Features Transform*. El algoritmo de Lowe extrae características invariantes a escala, rotación y parcialmente a perspectiva y cambios en la iluminación. El $SIFT$ se ha utilizado principalmente en las áreas de visión por computadora, reconocimiento de patrones y matching.

Se escogió el algoritmo $SIFT$ con base en el análisis de los diferentes métodos de extracción de puntos característicos realizado en [15] y en la experimentación. Además, particularmente, los descriptores del algoritmo $SIFT$ permiten a el algoritmo propuesto reconstruir la marca sin la necesidad de regresar la imagen a su posición original después de un ataque de rotación.

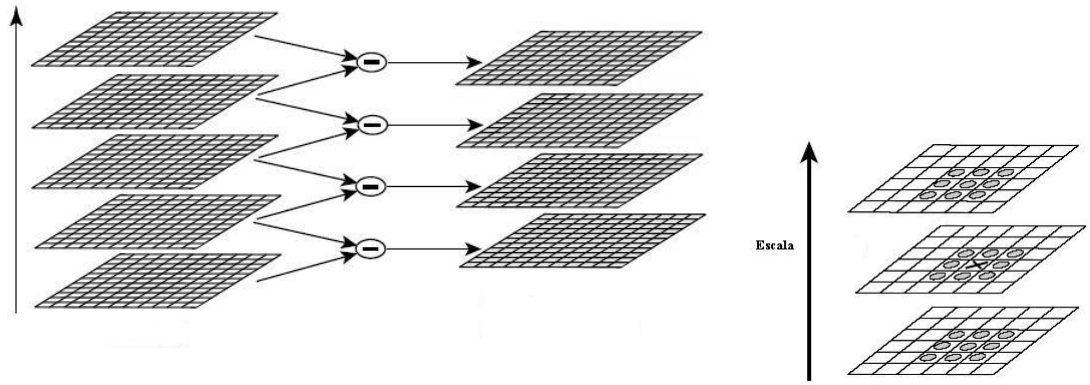
A continuación se describe brevemente el algoritmo $SIFT$.

Como primer paso se detectan los *potenciales* puntos de interés. Esto se hace calculando los valores extremos en el espacio de las escalas.

El espacio de las escalas está formado por un conjunto de imágenes que resultan de la convolución de una imagen dada con el kernel de la función Gauss, usando diferentes desviaciones estándar. En otras palabras, la escala se refiere a la desviación estándar utilizada.

Para localizar los extremos en el espacio de escalas, $SIFT$ utiliza la diferencia de imágenes en escalas contiguas tal y como se muestra en la figura 4.1-a). Después, cada uno de sus pixeles en las imágenes resultantes es comparado con sus 26 vecinos, 9

vecinos en la escala superior, 9 en la escala inferior y 7 en la misma escala. El pixel se considera *potencial* punto de interés si tiene un valor máximo o mínimo con respecto a sus vecinos(figura 4.1-b).



a) b)

Figura 4.1: Detección de puntos característicos *potenciales* a)Diferencia entre escalas contiguas. b) Comparación del pixel analizado y sus 26 vecinos para encontrar valores máximos y mínimos.

En la siguiente fase del algoritmo se evalúa la estabilidad de los *potenciales* puntos de interés para eliminar los puntos con bajo contraste o que se encuentran sobre los bordes. Esto es debido a que estos puntos se pierde fácilmente cuando se agregan pequeñas cantidades de ruido a la imagen.

La estabilidad de los puntos se calcula mediante la siguiente ecuación.

(4.1).

$$St = \frac{(D_{ij} + D_{ij})^2}{D_{ii}D_{jj} - D_{ij}^2} < \frac{(e+1)^2}{e} \tag{4.1}$$

Donde, D_{ii}, D_{jj} y D_{ij} son derivadas calculadas mediante la diferencia entre el pixel analizado y sus pixeles vecinos. Finalmente, e es el radio entre el valor del eigenvalor mayor y el menor de la matriz Hessian, para el algoritmo SIFT $e = 10$.

Como siguiente fase, a cada punto característico seleccionado se le asigna una magnitud y una dirección. Para ello, se toma una región alrededor del punto considerado y

para cada punto en esta región se asigna una magnitud mediante la ecuación 4.2 y una dirección mediante la ecuación 4.5. La magnitud y dirección asignadas finalmente al punto característico corresponde a la magnitud y dirección predominantes de la región.

$$m(i,j) = \sqrt{m1 - m2} \tag{4.2}$$

Donde

$$m1 = L(i+1,j) - L(i-1,j)^2 \tag{4.3}$$

$$m2 = L(i,j+1) - L(i,j-1j)^2 \tag{4.4}$$

$$\theta(i,j) = tan^{-1}\left(\frac{L(i,j+1) - L(i,j-1)}{L(i+1,j) - L(i+1,j)}\right) \tag{4.5}$$

Donde, L es la imagen de $N \times M$ pixeles en la escala en donde se localizó el punto característico.

Finalmente, el algoritmo $SIFT$ asigna a cada punto característico un descriptor que permite distinguir a cada punto de manera particular.

El descriptor se forma de la siguiente manera: se toma un conjunto de pixeles alrededor del punto característico, después se toman subregiones y a partir de ellas se crean histogramas con la información de magnitud y dirección del gradiente en cada pixel de las subregiones. El descriptor es un vector en el que cada entrada contiene las magnitudes de los histogramas. En la figura 4.2 se muestra un ejemplo de la generación de los descriptores, en el se generan 4 histogramas a partir de subregiones de 4×4 pixeles. Los descriptores en el algoritmo $SIFT$ son vectores de 128 valores que corresponden un arreglo de 4×4 histogramas formados por subregiones de 16×16 pixeles y 8 direcciones.

Una vez que se localizan los puntos característicos de la imagen (figura 4.3-a)el algoritmo propuesto selecciona aquellos localizados en un radio igual a S que es la dimensión de la marca, pero que además, se encuentren en el rango de escalas de 2 a 10. Este rango de escala se establece debido a que se ha observado por medio de la

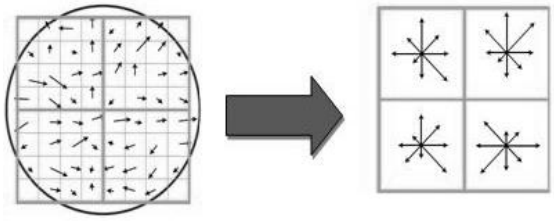


Figura 4.2: Descriptores de puntos característicos

experimentación que los puntos que se localizan en escalas menores a 2 se pierden con facilidad cuando se modifica la imagen y los puntos que se localizan en escalas mayores a 10 son difíciles de relocalizar [12]. En la figura 4.3-b) se muestran los puntos filtrados por el algoritmo propuesto para la imagen $Lena$.

a) b)

Figura 4.3: Extracción de puntos de referencia . a) Utilizando el algoritmo $SIFT$ se extraen puntos característicos de la imagen. b) Puntos filtrados por el algoritmo propuesto.

Una vez que se encuentran los puntos de referencia a utilizar sus descriptores se guardan en el archivo Ad.

4.1.2. Construcción de la marca compartida

En la segunda etapa de la fase de generación del secreto compartido se utilizan los puntos característicos filtrados como referencia para la extracción de información de la imagen usada en la codificación de la imagen compartida.

En el algoritmo propuesto se extraen pixeles alrededor de los puntos característicos en un radio S. La dirección establecida por el algoritmo $SIFT$ para cada punto de referencia se utiliza para extraer los pixeles en sentido inverso a las manecillas del reloj (figura 4.4-a). Los pixeles extraídos se utilizan para crear imágenes de tamaño de la marca de agua (figura 4.4-b). Una vez que se extraen los pixeles y se forman las imágenes de $S \times S$, las imágenes resultantes se binarizan de la siguiente manera: se divide la imagen en bloques de 3×3 y se calcula la media de cada bloque, a los pixeles con valor inferior a la media se les asigna un valor de cero y en caso contrario se asigna un valor igual a 1 (figura 4.4-c).

Cada una de las imágenes binarias para futuras referencias se etiquetan con $Sh1_i$ donde $i = 1...K$, K es el número de puntos característicos usados como referencia.

a) b) c)

Figura 4.4: Extracción de información de la imagen para la generación de la imagen compartida.

Para formar las imágenes $Sh1_i$ a patir de los parches circulares alrededor de los

puntos de referencia, se utiliza la transformada inversa *logpolar* que convierte las coordenadas polares a cartesianas (figura 4.5). La transformada *logpolar* inversa se define por la ecuación (4.6).

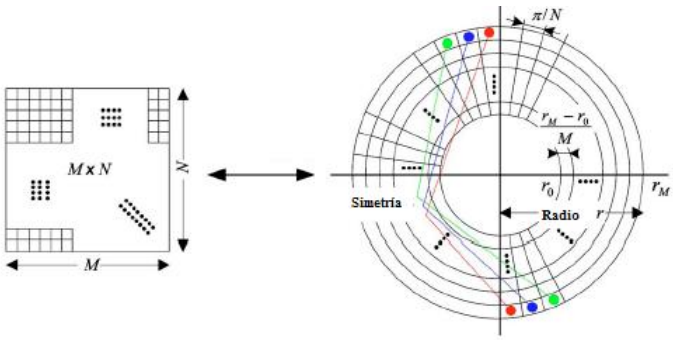


Figura 4.5: Transformada Log-polar.

$$\begin{aligned} x &= \tfrac{r_i - r_0}{r_M - r_0} \cdot M, y = \tfrac{\theta}{\pi} \cdot N, if 0 \leq \theta < \pi. \\ x &= \tfrac{r_i - r_0}{r_M - r_0} \cdot M, y = \tfrac{\theta}{\pi} \cdot N, if \pi \leq \theta < 2\pi. \end{aligned} \tag{4.6}$$

Para construir la imagen compartida se combina cada una de las imágenes binarias $Sh1_i$ con la marca W. Con esto tenemos un nuevo conjunto de imágenes binarias $Sh2_i$ de $S \times S$. Para combinar las imágenes $Sh1_i$ con la marca W se utiliza el operador XOR cuya tabla de verdad se muestra en la tabla 4.1. En términos de pixeles un valor 0 en la tabla representa un pixel negro mientras que un valor igual a 1 representa a un pixel blanco. Se utilizó este operador porque como se puede observar en la tabla 4.1 al realizar $P1 \oplus P2$ para crear P3 es posible recuperar P1 al mediante $P2 \oplus P3$.

En la figura 4.6 se ilustra la creación de las imágenes $Sh2_i$ a partir de los parches circulares alrededor de los puntos característicos seleccionados.

Finalmente, los pixeles en las imágenes $Sh2_i$ se transforman a coordenadas polares utilizando la transformada inversa *logpolar* y se colocan en la imagen compartida Is. En la figura 4.7 se muestra un ejemplo de imagen compartida Is generada por el algoritmo propuesto.

P1	P2	$P1 \oplus P2$
0	0	0
0	1	1
1	0	1
1	1	0

Tabla 4.1: Tabla de verdad del operador XOR.

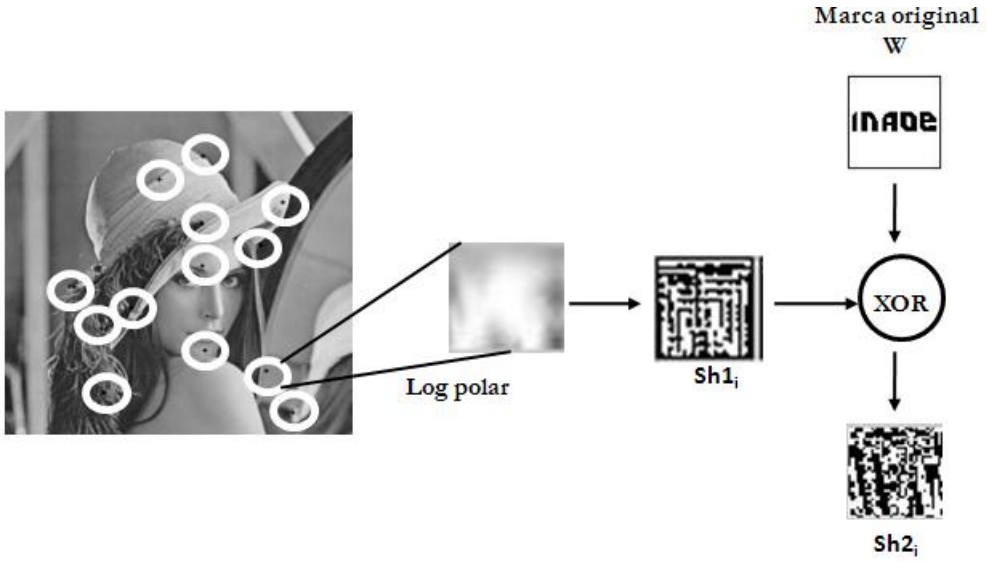


Figura 4.6: Construcción de las imágenes $Sh2_i$

4.2. Fase de reconstrucción de la marca de agua

El proceso de reconstrucción de la marca es muy parecido a la fase de creación del secreto compartido. Las entradas requeridas durante este proceso son:

- Imagen atacada I' en escala de grises de dimensión $N \times M$ pixeles que se desea analizar.

- El secreto compartido Shd formado por la imagen compartida binaria Is de $N \times M$ pixeles y el archivo Ad con los descriptores de puntos característicos usados

Figura 4.7: Ejemplo de imagen compartida generada por el esquema propuesto.

como referencia.

La salida en esta fase es una marca de agua W' reconstruida de $S \times S$ pixeles.

Las fase de reconstrucción de la marca de agua consiste de las siguientes etapas.

1. Sincronización de la marca.
2. Reconstrucción de la marca en diferentes puntos.
3. Mejora de la marca reconstruida usando la redundancia existente de información.

Sincronización de la marca.

Primero se utiliza el algoritmo $SIFT$ para extraer los puntos característicos de la imagen I' con sus respectivos descriptores. Aquí, a diferencia de la fase de construcción del secreto compartido no se filtran los puntos, pues al ser atacada la imagen los puntos pueden cambiar. Después, para distinguir los puntos de referencia utilizados en la fase de generación del secreto compartido se utilizan los descriptores del archivo Ad y los descriptores de los puntos característicos en I'. La correspondencia entre puntos se realizó calculando la distancia entre los vectores de descriptores. En la figura 4.8 se

muestra un ejemplo de la localización de los puntos de referencia después de rotar la imagen de $Lena$ 45º.

Figura 4.8: localización de los puntos de referencia después de rotar la imagen de $Lena$ 45º.

4.2.1. Reconstrucción de la marca en diferentes puntos.

Una vez localizados los puntos de referencia en la imagen I' se reconstruye la marca a partir de los diferentes puntos. El proceso de reconstrucción de la marca en los diferentes puntos es similar al de generación de la imagen compartida. En la figura 4.9 se observa como se reconstruye la marca usando parches circulares alrededor de los puntos de referencia.

Primero, se extraen los pixeles en I' alrededor de los puntos de referencia para generar las imágenes binarias $Sh1_j$ de la misma manera que en la fase de generación de la imagen compartida. Las imágenes $Sh2_j$ son imágenes binarias de $S \times S$ generadas al tomar los pixeles de la imagen compartida alrededor de los puntos de referencia. Después, las imágenes $Sh1_j$ y $Sh2_j$ se combinan usando el operador XOR para formar una versión de la marca W'_j por cada punto.

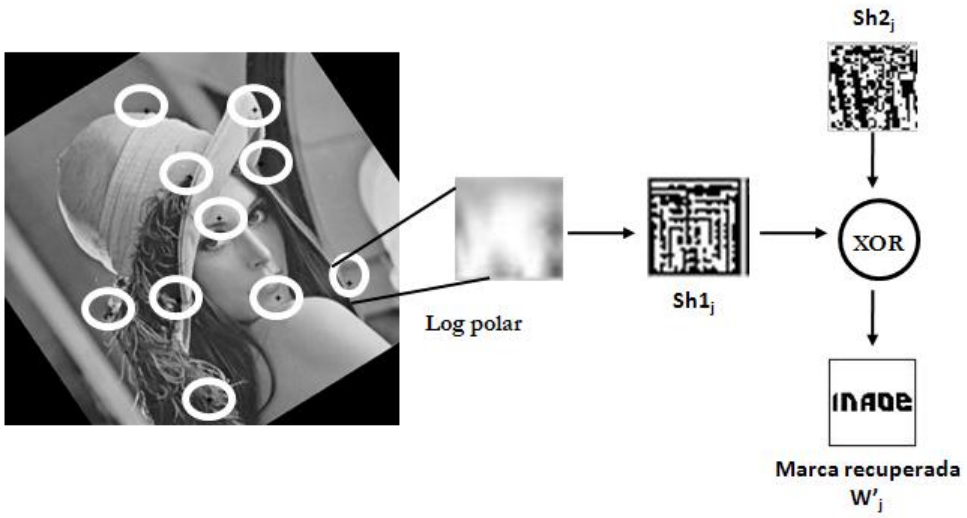


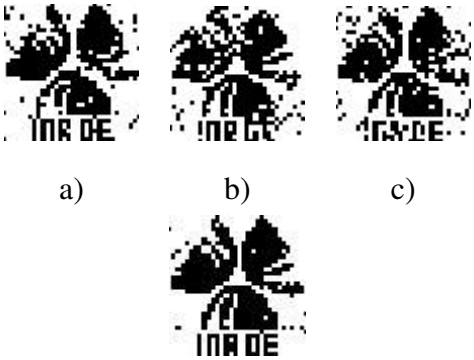
Figura 4.9: Reconstrucción de la marca en diferentes puntos.

4.2.2. Mejora de la marca reconstruida usando redundancia de información.

Al final del algoritmo propuesto se regresa sólo una marca reconstruida W'. Para mejorar la calidad de la marca reconstruida W', se combinan las tres marcas reconstruidas W'_y, W'_t y W'_z, donde: $y \neq t \neq z$, con los mayores valores de correlación entre la imagen reconstruida y la marca original. La combinación se realiza colocando en cada pixel de la marca reconstruida W' , el valor predominante de los pixeles corres pondientes en las marcas reconstruidas W'_y, W'_t y W'_z.

En la figura 4.10 se presenta un ejemplo de los resultados obtenidos al combinar tres marcas recuperadas en diferentes puntos de la imagen. La marca reconstruida que se muestra en la imagen 4.10-a) tiene una correlación igual a 0.9413, la de la marca en 4.10-b) es 0.92456 y 0.9344 en 4.10-c). Obsérvese que la imagen resultante mostrada en 4.10-d) alcanza una correlación de 0.97945.

A pesar de que el método $SIFT$ usado para la extracción de puntos característicos ha reportado un alto grado de repetitividad en los puntos extraídos, a veces debido a la intensidad del ataque realizado a la imagen, no es posible recuperar ni tres puntos



a) b) c)

d)

Figura 4.10: Resultados de la combinación de marcas recuperadas. En los incisos a)-c) se muestran las marcas recuperadas en diferentes puntos. La marca resultante de la combinación de las tres marcas en a),b) y c) se muestra en el inciso d).

necesarios para la reconstrucción de las imágenes W'_y, W'_t y W'_z o las imágenes recuperadas están demasiado degradadas y su combinación no mejora el resultado obtenido por una de ellas. En estos casos, el algoritmo regresará la marca reconstruida con mayor similitud a la original.

4.3. Modificaciones al algoritmo propuesto

Durante las primeras fases de experimentación se detectó que se podía mejorar el esquema propuesto, por lo que se hicieron los siguientes cambios:

- Durante la etapa de extracción de características y generación de la imagen compartida, al considerar los valores originales de los pixeles de la imagen, al igual que el algoritmo BTC, el esquema se vuelve sensible al ruido y a cambios como blurring y aumento de contraste. Por lo que, en lugar de tomar las características de la imagen original, una vez localizados los puntos en la imagen original, las características se extraen de la aproximación del nivel 3 de la trasformada wavelet que es más estable.

- Con la finalidad incrementar la seguridad del sistema, la marca es desordenada utilizando la técnica de caos propuesta por Voyatzis y Pitas en [32] denominada transformada *torusautomorphism*. Así, además de tener la imagen compartida y los descriptores de los puntos se deberá tener una llave numérica para reordenar la marca de agua recuperada. La transformada *torus automorphism* se define por la ecuación m.

$$\begin{bmatrix} x_{i+1} \\ y_{i+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ k & k+1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} mod(N) \tag{4.7}$$

  En la figura 4.11 se muestra un ejemplo del uso de la transformada torus automorphism para mezclar la marca de agua antes de la fase de creación del secreto compartido.

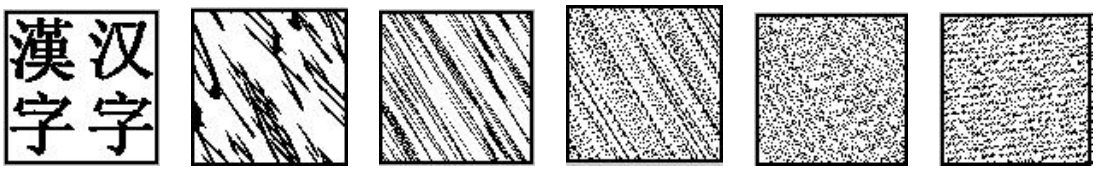


Figura 4.11: Pixeles de la marca de agua mezclados usando Torus-automorphism. De derecha a izquierda se presentan el resultado de reordenaciones sucesivas de la marca usando torus-automorphism

- La tercera modificación se hizo en la forma de crear la imagen compartida. El problema era que una vez combinadas las imágenes $Sh1_i$ con la marca original, los pixeles las imágenes binarias resultantes $Sh2_i$ se transformaban a coordenadas polares y se colocaban alrededor de los puntos de referencia en una imagen compartida con las mismas dimensiones que la original. El cambio consistió en formar la imagen compartida ordenando las imágenes $Sh2_i$. En la figura 4.12 se muestra un ejemplo de la nueva imagen compartida generada al considerar los mismos 12 puntos de referencia considerados para construir la imagen compartida de la figura 4.7. La imagen resultante de la modificación es de 384×32

pixeles por lo que necesita 13 kb para ser almacenada que es muy inferior a los 257 kb que se necesitaron para almacenar la imagen compartida usando el esquema original. En este caso la imagen compartida se formó con la unión consecutiva de las imágenes $Sh2_i$, sin embargo, se puede hacer diferente arreglos con estas imágenes.

Figura 4.12: Ejemplo de la imagen compartida generada después de las modificaciones hechas al esquema propuesto.

En este capítulo se presentó el algoritmo propuesto en este trabajo de investigación. No se sigue el esquema tradicional de los algoritmos de marca de agua en los que la marca se inserta en la imagen ya sea en el dominio espacial o de las frecuencias. El algoritmo se basa en el esquema de criptografía visual propuesto por Naor and Shamir para crear una imagen compartida que al combinarla con características de la imagen revela la marca de agua binaria que identifica al propietario. Al no insertar la marca en la imagen, el algoritmo propuesto no degrada la imagen e incrementa la seguridad, pues para reconstruir la marca se necesita tener la imagen compartida; por lo que además de protección de copyright el algoritmo puede emplearse para transmisión segura de mensajes.

Como ya se mencionó anteriormente, el algoritmo que se propuso sigue el esquema de los algoritmos basados en secreto compartido, sin embargo incorpora una fase de sincronización basada en puntos característicos de la imagen original que hace al algoritmo robusto a ataques geométricos de rotación y traslación sin necesidad de regresar la imagen a su posición original. Además, el algoritmo permite reconstruir la marca de agua después de un ataque de cropping al usar redundancia de información.

Capítulo 5

Resultados

En este capítulo se presentan los resultados obtenidos por la implementación del algoritmo propuesto ante diferentes ataques de manipulación de la imagen como, igualación, cambios en el contraste, compresión JPEG, etc. Los resultados obtenidos por el algoritmo propuesto son comparados con los obtenidos por los trabajos relacionados descritos en el capítulo 3.

5.1. Métricas utilizadas durante la evaluación de resultados

Durante los experimentos se evaluaron dos aspectos:

- El primer aspecto evaluado fue la semejanza entre la marca de agua recuperada y la marca original. Para medir el grado de similitud entre la marca reconstruida y la original se utilizó la métrica de correlación normal NC que se define mediante la ecuación (5.1) y tiene un rango de valores entre 0 y 1. En donde un valor igual a cero indica que no existe relación entre las imágenes analizadas y un valor igual

a uno indica que las imágenes son iguales.

$$NC = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} I(i,j)I'(i,j)}{\sum_{i=1}^{M}\sum_{j=1}^{N} I(i,j)^2} \tag{5.1}$$

La métrica de correlación es una de las métricas más utilizadas para evaluar los resultados obtenidos en los algoritmos de marca de agua.

- El segundo aspecto que se evaluó fue la precisión del algoritmo para reconstruir la marca, es decir; exactamente cuál es el porcentaje de pixeles de la marca que se reconstruyeron correctamente después de que la imagen es atacada. La métrica para evaluar la precisión del algoritmo fue la métrica denominada *Accuracy Rate o AR* y definida por la ecuación (5.2).

$$AR = \frac{CB}{NB} \tag{5.2}$$

En donde NB es el número total de píxeles en la marca original y CB es el número de pixeles recuperados correctamente.

El rango de valores de la métrica AR va de 0 a 1, en donde un valor igual a 1 indica que todos los pixeles de la imagen se recuperaron correctamente.

Otra métrica que se utilizó para mostrar los resultados fue la métrica $PSNR$ o *Peak Signal to Noise Ratio*. $PSNR$ mide la relación entre la señal original y una modificación de la misma señal en la que se ha afectado la calidad. La definición de la métrica PSNR está dada por la ecuación (5.3) y se expresa en decibeles.

$$PSNR = 10log_{10}\frac{255^2}{MSE}(dB) \tag{5.3}$$

En donde MSE es el error cuadrático medio cuya fórmula se presenta en la ecuación (5.4).

$$MSE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(I(i,j) - I'(i,j))^2 \tag{5.4}$$

La medida $PSNR$ es comúnmente usada como medida de calidad en la reconstrucción de imágenes o para medir la calidad de una imagen comprimida. Los algoritmos tradicionales de marca de agua utilizan esta métrica para medir la calidad de la imagen después del proceso de inserción de la marca de agua. En este caso, puesto que en el algoritmo propuesto no existe un proceso de inserción de la marca en la imagen, la calidad de la imagen original no se ve comprometida en ninguna forma. Se utiliza la métrica $PSNR$ para evaluar la calidad de la imagen después de un ataque.

5.2. Experimentos

En la figura 5.1 se presentan las imágenes que fueron utilizadas durante los experimentos. Las imágenes son en escala de grises de 512×512 píxeles y son imágenes estándar de prueba que se utilizan comúnmente en el procesamiento digital de imágenes.

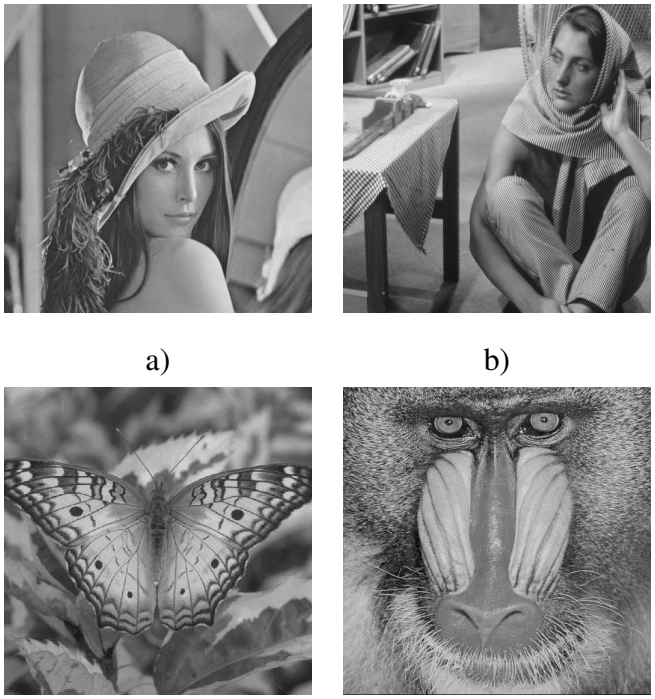
En la figura 5.2 se presentan las marcas binarias utilizadas durante los experimentos. La primera imagen (5.2-a) es de 32×32 pixeles y contiene el logotipo del INAOE, la segunda imagen (5.2-b) es de 64×64 pixeles y contiene el nombre 'INAOE', por último, la tercera imagen (5.2-c) es de 60×60 píxeles y es el símbolo chino de la tortuga.

Los experimentos realizados se pueden dividir en tres clases:

- Ataques que alteran a los valores de los pixeles de la imagen tales como compresión y adición de ruido.

- Prueba de falsos positivos.

- Ataques geométricos en los que se alteran las posiciones de los pixeles de la imagen. Para el algoritmo propuesto los ataques que se consideraron son: rotación, translación y cropping



a) b)

c) d)

Figura 5.1: Imágenes de 512x512 pixels en escala de grises utilizadas en los experimentos. a)Lena. b) Bárbara. c) Mariposa. d) Mandril.

5.2.1. Manipulación de los valores de pixeles

Compresión JPEG

En esta sección se muestran los resultados obtenidos por el algoritmo propuesto ante el ataque de compresión JPEG. El experimento se realizó utilizando la imagen de Lena y el logo del INAOE y consistió en reconstruir el logo del INAOE a partir de versiones comprimidas de la imagen con porcentajes de calidad que van del 95 % al 25 %. Estos porcentajes indican la cantidad de información que se conserva de la imagen original después de la compresión $JPEG$.

Para realizar el ataque de compresión se utilizó Microsoft Photo Editor. Durante este experimento se emplea también la métrica AR para medir la precisión con la que el algoritmo reconstruye la marca. Los resultados obtenidos fueron comparados por los

a) b) c)

Figura 5.2: Marcas binarias utilizadas en los experimentos. a) logo INAOE de 32×32 pixeles. b) Imagen con el texto 'INAOE' 64×64 pixeles. c) Símbolo chino de la tortuga de 60×60 pixeles.

obtenidos por [30, 11, 9] en el mismo experimento.

En la figura 5.3 se muestra la gráfica de degradación de la imagen comprimida medida en decibeles usando $PSNR$. Como se puede observar, la calidad de la imagen se degrada rápidamente de 44 dB a 34dB. La gráfica 5.3 proporciona un indicativo de la severidad de los ataques realizados en este experimento.

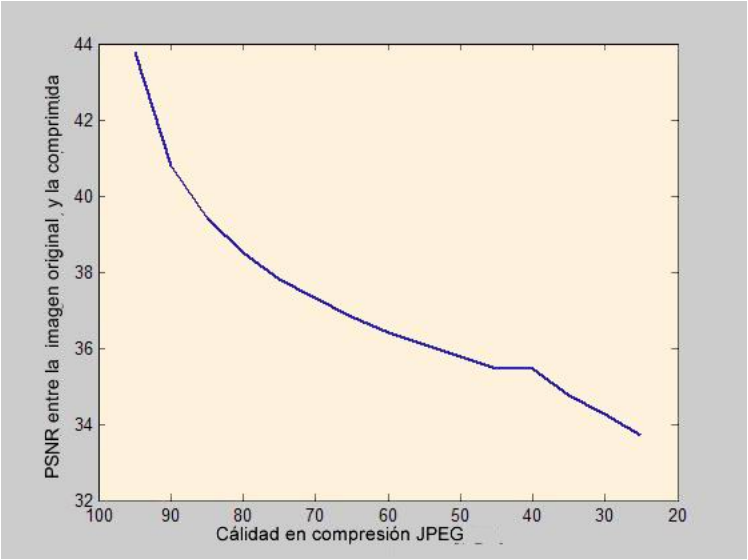


Figura 5.3: Degradación de la imagen después de la compresión JPEG.

En la figura 5.4 se muestra la gráfica de precisión alcanzada por los algoritmos ante diferentes calidades de compresión JPEG.

Como se puede observar en la figura 5.4, en general los algoritmos que trabajan con

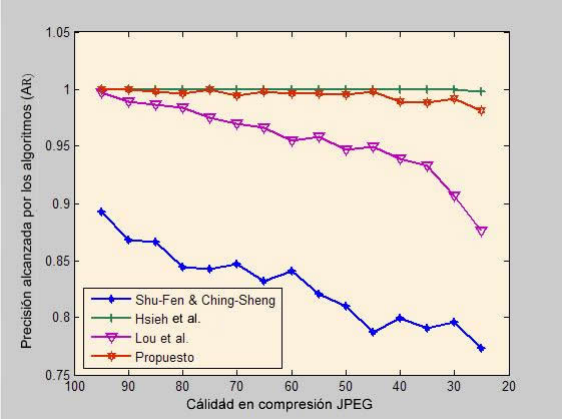


Figura 5.4: Precisión alcanzada por los algoritmos después de un ataque de compresión JPEG.

las caracteristicas de la transformada wavelet que son: el propuesto por Hsieh et al., el propuesto por Lou et al. y el propuesto en este trabajo tienen un mejor desempeño que el algortimo propuesto por Shu-Fen que trabaja con las características de los valores de pixeles originales.

El mejor de los algoritmos, fue el propuesto por Hsie et al. que reconstruyó el 100 % de la marca en la mayoría de los casos excepto en el último en donde recuperó el 99.8 % de la marca. El algoritmo propuesto tuvo un buen desempeño al reconstruir en promedio el 99.47 %. Por el alto porcentaje de precisión obtenido durante este experimento podemos concluir que el algoritmo propuesto es robusto ante ataques de compresión JPEG.

Ruido gaussiano

El segundo experimento tuvo como objetivo mostrar el comportamiento del algoritmo propuesto ante un ataque en el que se agrega ruido gaussiano a la imagen y compararlo con los trabajos relacionados descritos en la sección 3.1. La distribución de este tipo de ruido se asemeja a una distribución gaussiana definida por la media y varianza dadas.

Para agregar el ruido a la imagen se utilizó matlab 7.0. Durante esta prueba se aplicaron diferentes cantidades de ruido en la imagen para observar el comportamiento de los algoritmos. La media de la función gaussiana se estableció en 0 para todos los casos, mientras que las desviaciones estándar van de 0.000006 a 0.01 en pasos de 0.0004997. Los valores de la desviación estándar utilizados ocasionan que la imagen atacada sufra una caída en la calidad de la imagen que va de 51.3883 dB a 20.08 dB.

La figura 5.5 muestra la degradación que sufre la imagen cuando se le aplica ruido gaussiano. Como se puede observar los valores de PSNR son más bajos y descienden con mayor rapidez que los observados durante el ataque de compresión $JPEG$. Por lo tanto, se considera a este ataque más agresivo que la compresión $JPEG$.

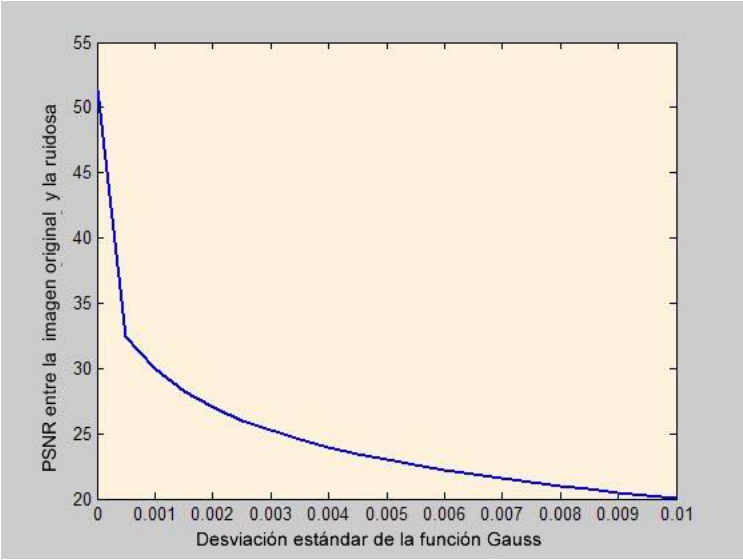


Figura 5.5: Gráfica de degradación de la imagen después de ruido gaussiano.

Los resultados obtenidos por los diferentes algoritmos se muestran en la gráfica comparativa de la figura 5.6 en donde se representa la precisión (AR) con la que los algoritmos recontruyen la marca.

En la gráfica de la figura 5.6 se aprecia que el algoritmo propuesto por Shu-Fen & Ching-Shen es el más afectado por la presencia de ruido gaussiano. Esto se debe a que construye la imagen compartida a partir de la media de los valores originales de pixeles

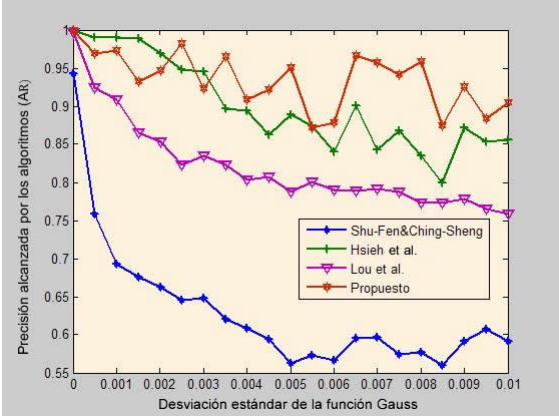


Figura 5.6: Precisión alcanzada por los algoritmos después de agregar ruido gaussiano

en bloques de 3×3 y estos valores son alterados fácilmente por la presencia de ruidos.

Los trabajos restantes obtienen mejores resultados en la reconstrucción de la marca después de agregar ruido gaussiano a la imagen que los obtenidos por Shu-Fen & Ching-Shen. Esto es debido a que utilizan, aunque de distinta forma, la transformada wavelet para la generación de la imagen compartida para reducir los efectos del ruido en la imagen.

Así, el método de codificación del algoritmo propuesto por Hsieh et al. supera al de Lou et al por que utiliza coeficientes wavelets mientras que, el algoritmo de Lou et al. utiliza una imagen obtenida de la diferencia entre dos imágenes reconstruidas a diferentes niveles de la trasformada.

En el caso del algoritmo propuesto se obtuvieron resultados en su mayoría superiores al de Hsieh et al. debido principalmente a la redundancia de información en diferentes puntos de la imagen. Los picos irregulares en la gráfica del algoritmo propuesto se deben a que el ruido además de afectar la generación de la imagen sombra, altera también la fase de sincronización pues esta se basa en el contenido de la imagen. Sin embargo, no altera de la misma forma a todos los puntos y al combinar marcas en diferentes puntos estas pueden dar mejores resultados para un ruido de mayor intensidad que para otro de menor intensidad. Sin embargo, el valor promedio de la precisión alcanza-

da por el algoritmo durante este experimento fue de 0.9348, por lo que se concluye que el algoritmo propuesto es robusto a ruido gaussiano.

Resultados obtenidos ante ecualización, aumento de contraste y ruido sal y pimienta

El tercer experimento consistió en aplicar ecualización y aumento de contraste a la imagen. Además se realizaron pruebas agregando ruido sal y pimienta a la imagen.

Las modificaciones a la imagen se realizaron utilizando funciones para el procesamiento de imágenes de Matlab 7.0. En la tabla 5.1 se muestran los valores de la métricas NC y AR alcanzados por los diferentes algoritmos durante este experimento.

Ataque		Shu-Fen&Ching-Sheng	Hsieh et al.	Lou et al.	Propuesto
Ecualización PSNR(17.993)	NC	0.9632	0.9374	0.9832	0.9803
	AR	0.942	0.9707	0.9883	0.9914
Aumento de Contraste PSNR(15.79)	NC	0.9946	0.9949	0.9783	0.9846
	AR	0.9600	0.9971	0.9844	0.9795
Ruido sal y pimienta PSNR(26.55)	NC	0.9830	0.8882	0.9685	0.9814
	AR	0.979	0.9	0.977	0.979
Ruido sal y pimienta PSNR(18.5)	NC	0.818	0.749	0.876	0.938
	AR	0.781	0.759	0.910	0.9248

Tabla 5.1: Resultados obtenidos ante la ecualización, cambio en el contraste y adición de ruido gaussiano en la imagen *Lena*.

En general, los algoritmos de Shu-Fen&Chin-Sheng, Hsieh et al. y Lou et al. obtienen buenos resultados ante los ataques de ecualización y aumento en el contraste. De ellos, el que se ve más afectado por la presencia de ruido en la imagen es el de Lou et al. En el caso de los algoritmos de Shu-Fen&Chin-Sheng y Lou et al. inicialmente el ruido no les afecta, sin embargo; cuando el ruido aumenta ambos algoritmos reducen su precisión significativamente.

En el caso del algoritmo propuesto, los resultados obtenidos en todos los ataques analizados en este experimento fueron en su mayoría los mejores. Además se observa que ante un aumento en el ruido sal y pimienta el comportamiento de nuestro algoritmo no desciende considerablemente como en los otros algoritmos. Esto se debe a que utiliza una representación de la imagen que no se ve alterada por la presencia de ruido (aproximación nivel 3 de la transformada wavelet).

5.2.2. Prueba para la detección de falsos positivos

En esta sección se describe el experimento realizado para conocer la suceptibilidad de los algoritmos a cometer errores de falsos positivos. Este tipo de error es conciderado por muchos autores como uno de los más graves para los esquemas utilizados en protección de la propiedad intelectual. El experimento consistió en generar la imagen compartida a partir del logo del INAOE y la imagen *Lena* para después intentar recuperar la marca utilizando el resto de las imágenes de prueba.Los resultados de este experimento se muestran en la tabla 5.2, en la primera columna se indica el algoritmo que se está evaluando mientras que en las columnas 2-5 se presentan la marca recostruida y los resultados de NC y AR que evalúan su relación con la marca original.

Como se puede observar en la tabla 5.2 el algoritmo propuesto por Hsieh et al. a pesar de tener los resultados más altos durante el ataque JPEG y tener un buen desempeño en los ataque presentados anteriormente, comete errores de falsos positivos. El ejemplo más claro, es la marca recontruida de la imagen *Mariposa* en donde visualmente se puede apreciar el gran parecido con la marca original, se obtiene una correlación de 0.9177 y los pixeles de la marca reconstruida coinciden en un 93.16 % con los de la marca original.

En cambio, visualmente las marcas reconstruidas por los algoritmos propuestos por Shu-Fen & Ching-Sheng no se parecen en nada a la marca original y las métricas de NC y AR son bajas.

En el caso del algoritmo propuesto no se muestran resultados puesto que al hacer

Algoritmo		Imágenes		
		Bárbara	Mariposa	Mandril
Shu-Fen & Ching-Sheng	NC	0.5250	0.5934	0.6019
	AR	0.4521	0.5088	0.5283
Hsieh et al.	NC	0.8105	0.9177	0.6400
	AR	0.8418	0.9316	0.6377
Lou et al.	NC	0.5778	0.5635	0.5668
	AR	0.5352	0.5146	0.5234

Tabla 5.2: Resultados obtenidos por los diferentes algoritmos cuando se intenta reconstruir la marca de agua de imágenes diferentes a la usada para construir el secreto compartido.

la correspondencia entre los puntos de referencia usados y los puntos caracteristicos de las imágenes analizadas durante la fase de sincronización no se encontró coincidencia y por lo tanto no se inició el proceso de reconstrucción de la marca. Cabe mencionar, que esto es una ventaja del algoritmo propuesto ya que a partir de una imagen escogida arbitrariamente no se puede reconstruir la marca.

Como se mencionó anteriormente en ninguna de las imágenes de prueba se localizaron puntos característicos que coincidieran con los encontrados en *Lena*. Sin embargo, se realizó un experimento adicional en el que se forzó al algoritmo a reconstruir la marca utilizando una región de la imagen *Mandril* para probar la suceptibilidad del proceso de reconstrucción a falsos positivos. En la figura 5.7 se muestra la marca re-

cosntruida en este experimento que obtuvo valores de NC=0.5965 y AR=0.5225.

A pesar de no coincidir los puntos, se realizó un experimento adicional en el que se omitió la fase de sincronización y se forzó al algoritmo para que reconstruyera la marca utilizando una región de la imagen *Mandril* para probar la susceptibilidad del proceso de reconstrucción a falsos positivos. En la figura 5.7 se muestra la marca reconstruida durante este experimento. Como se puede observar en la figura 5.7, visualmente no se distingue el logo del INAOE y se obtuvieron valores de NC=0.5965 y AR=0.5225.

Figura 5.7: Prueba de falso positivo forzando puntos en el esquema propuesto

5.2.3. Ataques géometricos

Resultados obtenidos ante el ataque de rotación

En la tablas 5.3 a 5.6 se presentan los resultados obtenidos por el algoritmo propuesto ante el ataque de rotación usando cada una de las imágenes y marcas de prueba. El experimento consistió en rotar cada una de las imágenes de pruebas a 33º, 45º, 60º,90º y 100º y fue realizado utilizando la función *imrotate* de Matlab.

La primera columna de las tablas indica los grados que fueron rotadas las imágenes, la segunda columna de la tabla 5.3 muestra a la imagen rotada de la que se reconstruye la marca. Como se puede observar, la imagen de la que se reconstruye la marca es directamente la imagen rotada y no existe ningún tipo de proceso que regrese la imagen a su posición original antes del proceso de reconstrucción de la marca. Finalmente en las tres últimas columnas se presentan los valores de NC y AR obtenidos al reconstruir las marcas. Sólo se muestra la imagen rotada en la tabla 5.3 así como la imagen de la marca recuperada. Sin embargo, en todos los casos la marca se reconstruyó a partir de la imagen rotada.

Rotación	Imagen atacada	Marca reconstruida		
		Marca 1	Marca 2	Marca 3
33º	NC	0.926	0.921	0.980
	AR	0.912	0.790	0.887
45º	NC	0.93	0.932	0.982
	AR	0.881	0.823	0.892
60º	NC	0.951	0.931	0.981
	AR	0.929	.813	0.893
90º	NC	0.987	0.970	0.993
	AR	0.973	0.929	0.9385
100º	NC	0.918	0.973	0.983
	AR	0.877	0.929	0.884

Tabla 5.3: Resultados obtenidos por el algoritmo propuesto después de ataques de rotación a la imagen *Lena*

Rotación	Imagen atacada	Marca reconstruida		
		Marca 1	Marca 2	Marca 3
33º	NC	0.9602	0.9477	0.9877
	AR	0.9463	0.8583	0.9280
45º	NC	0.9583	0.9780	0.9786
	AR	0.9414	0.9375	0.8645
60º	NC	0.9700	0.9760	0.9908
	AR	0.9463	0.9344	0.9231
90º	NC	0.9845	0.9861	0.9932
	AR	0.9775	0.9619	0.9321
100º	NC	0.9701	0.958	0.9700
	AR	0.9482	0.885	0.8291

Tabla 5.4: Resultados obtenidos por el algoritmo propuesto después de ataques de rotación a la imagen *Bárbara*

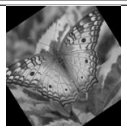
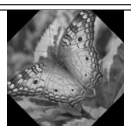
Rotación	Imagen atacada	Marca reconstruida		
		Marca 1	Marca 2	Marca 3
33º				
	NC	0.9303	0.9553	0.9550
	AR	0.9082	0.8878	0.8099
45º				
	NC	0.9230	0.9578	0.9802
	AR	0.8877	0.8869	0.8479
60º				
	NC	0.9334	0.9688	0.9700
	AR	0.9092	0.9117	0.8091
90º				
	NC	0.9676	0.9843	0.9886
	AR	0.9619	0.9453	0.9133
100º				
	NC	0.9181	0.9525	0.9684
	AR	0.8945	0.8531	0.7854

Tabla 5.5: Resultados obtenidos por el algoritmo propuesto después de ataques de rotación a la imagen *Mariposa*

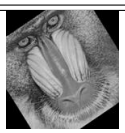
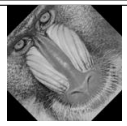
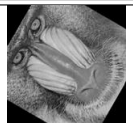
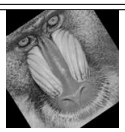
Rotación	Imagen atacada	Marcas reconstruidas		
		Marca 1	Marca 2	Marca 3
				
33º	NC	0.9705	0.9774	0.9888
	AR	0.9521	0.9433	0.9241
			—	
45º	NC	0.9651	—	0.9870
	AR	0.9434	—	0.9187
				
60º	NC	0.9592	0.9692	0.9891
	AR	0.9541	0.9128	0.9258
90º	NC	0.9932	0.9794	0.9914
	AR	0.9932	0.9456	0.9507
				
100º	NC	0.9665	0.9313	0.9883
	AR	0.9492	0.8178	0.9114

Tabla 5.6: Resultados obtenidos por el algoritmo propuesto después de ataques de rotación a la imagen *Mandril*

En la tabla 5.3 se muestran los resultados obtenidos por el algoritmo propuesto ante los ataques de rotación hechos sobre la imagen de *Lena*. Como se puede observar en todos los casos fue posible reconstruir la marca, en la mayoría de los casos se obtuvieron mejores resultados con la marca más pequeña de 32×32 pixeles. Esto es debido a que al ser pequeña la marca se consideraron más puntos de referencia y hubo más probabilidades de encontrar puntos menos afectados por la rotación. La marca que presento menor precisión y correlación fue el símbolo de la tortuga, pues a pesar de poder reconocerse visualmente la presencia del símbolo a pesar de que el fondo fue muy ruidoso.

En todas las imágenes rotadas 90º se obtuvieron los valores más altos de AR. Esto se debe a que al rotar la imagen 90º grados la imagen sufre menos alteraciones por interpolación. Otra cosa que se puede observar es que, aplicando el mismo ataque a diferente imagen no se obtienen los mismos valores de NC y AR. Esto es por que tanto la fase de sincronización como la codificación de la imagen compartidas usadas po el algoritmo dependen totalmente de las características de cada imagen. Por lo tanto, no se puede garantizar que el algoritmo obtenga ciertos valores de AR y NC para un derminado ataque en todas las imágenes.

En la tabla 5.6 fila 3 columna 4 no se presentan resultados debido a que los puntos usados como referencia no se pudieron relocalizar después de la rotación. En general cuando se usaron las marcas 2 y 3 de 60×60 y 64×64 pixeles respectivamente, el número de puntos usados como referencia fue menor debido al tamaño. Así, mientras que para la marca 1 se utilizaron 12 puntos de referencia en las marcas 2 y 3 se utilizaron 8 y 6 respectivamente. Esto reduce la posibilidad de encontrar algún punto de referencia en la imagen atacada para recuperar la marca. En este caso al rotar la imagen $Mandril$ 45º los puntos usados como referencia para la marca 2 se perdieron y no se pudo reconstruir la marca.

El número total de veces que se realizó la reconstrucción de las imágenes a partir de una imagen rotada fue de 60, que corresponden a las combinaciones obtenidas a partir de *cuatro* imágenes, *tres* marcas y *cinco* diferentes grados de rotación. De los

cuales sólo en un caso no se pudo recuperar la marca (tabla 5.6) debido a que no se relocalizaron los puntos de inserción en la imagen rotada. Para las marcas recuperadas los valores de NC están arriba de 0.9 y la de AR arriba de 0.8 por lo que podemos concluir que el algoritmo propuesto es robusto a ataques de rotación.

Resultados obtenidos ante ataques de cropping y translación

El último experimento realizado tiene como objivo mostrar el comportamiento del algoritmo propuesto ante ataques de cropping y traslación. El ataque de cropping consiste en eliminar cierta cantidad de filas y columnas de la imagen por lo que el tamaño de la imagen cambia. Sin embargo, a diferencia de un ataque de escala, el tamaño de los objos dentro de la imagen no se altera después de un ataque de cropping.

El ataque traslación consiste en desplazar los pixeles de la imagen Dx número de posiciones en dirección x y Dy en dirección y. La ecuación (5.5) describe el ataque de translación.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x + Dx \\ y + Dy \end{bmatrix} \tag{5.5}$$

Como se puede apreciar las distancias Dx y Dy pueden no ser iguales. En las tablas 5.7 a 5.10 se muestran los resultados obtenidos por el algoritmo propuesto ante los ataques realizados en este experimento.

En la primera fila de las tablas 5.7 a 5.10 se muestra el resultado de un ataque de cropping aplicado a cada una de las imágenes prueba. El ataque consistió en eliminar 20 de las primeras columnas y 112 de las últimas para un total de 132 columnas eliminadas, también fueron eliminadas 30 de las primeras filas y 92 de las últimas para dar un total de 122 filas eliminadas. En la segunda fila de las tablas se muestran los resultados obtenidos al rotar la imagen resultante del ataque anteriormente mencionado 45º. Las últimas dos filas de las tablas muestran diferentes combinaciones de ataques.

En la tabla 5.7, fila 3, se muestra el resultado de trasladar la imagen *Lena* a distancias de $Dx = -55$ y $Dy = -25$ píxeles. En la fila 4 de la tabla se presentan los

resultados de rotar la imagen trasladada 33º.

En la tabla 5.8, fila 3, se muestra el resultado de eliminar las 10 primeras y 31 últimas columnas y las 50 primeras y 111 últimas filas de la imagen *Bárbara*, después se traslada la imagen cortada a distancias $Dx = 45$ y $Dy = 25$ píxeles. En la fila 4 de la tabla se presentan los resultados de rotar la imagen atacada 90º.

En la tabla 5.9, fila 3, se muestra el resultado de trasladar la imagen *Mariposa* a distancias $Dx = 212$ y $Dy = 62$ píxeles. En la fila 4 de la tabla se presentan los resultados de rotar la imagen trasladada 60º.

En la tabla 5.10, fila 3, se muestra el resultado de trasladar la imagen *Mandril* a distancias $Dx = 20$ y $Dy = 50$ píxeles.En la fila 4 de la tabla se presentan los resultados de rotar la imagen trasladada 45º.

Los ataques de traslación y la combinación de ataques traslación+rotación fueron diferentes para cada imagen con la finalidad de probar el algoritmo propuesto ante una mayor diversidad de ataques.

Ataque	Imagen atacada	Marcas reconstruidas		
		Marca 1	Marca 2	Marca 3
Cropping				
	NC	0.972	0.9481	0.9903
	AR	0.9658	0.8689	0.9268
Cropping + rotación				
	NC	0.96	0.9381	0.9923
	AR	0.942	0.8464	0.9138
Traslación				
	NC	0.9879	0.9831	0.9935
	AR	0.9805	0.9431	0.9463
Tralación + rotación				
	NC	0.9629	0.9520	0.9870
	AR	0.9385	0.8822	0.9028

Tabla 5.7: Resultados obtenidos por el algoritmo propuesto ante ataques cropping y traslaciones en la imagen *Lena*.

Ataque	Imagen atacada	Marcas reconstruidas		
		Marca 1	Marca 2	Marca 3
Cropping				
	NC	0.9813	0.9811	0.9952
	AR	0.9795	0.9544	0.9707
Cropping + rotación				
	NC	0.9672	0.9503	0.9748
	AR	0.9551	0.8686	0.8210
Traslación				
	NC	0.9830	0.9766	0.9954
	AR	0.9824	0.9414	0.9648
Tralación + rotación				
	NC	0.9809	0.9786	0.9795
	AR	0.9688	0.9319	0.9351

Tabla 5.8: Resultados obtenidos por el algoritmo propuesto ante ataques cropping y traslaciones en la imagen *Bárbara* .

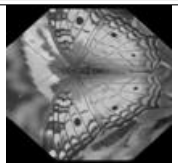
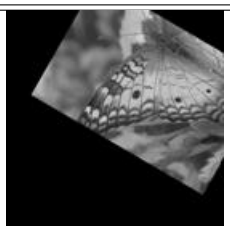
Ataque	Imagen atacada	Marcas reconstruidas		
		Marca 1	Marca 2	Marca 3
Cropping				
	NC	0.9515	0.9688	0.9822
	AR	0.9355	0.9119	0.8669
Cropping + rotación				
	NC	0.9316	0.9223	0.9595
	AR	0.8945	0.7989	0.7444
Traslación				
	NC	0.9478	0.9763	0.9402
	AR	0.9287	0.8188	0.8406
Traslación + rotación				
	NC	0.87719	0.9008	0.9711
	AR	0.8457	0.7578	0.8037

Tabla 5.9: Resultados obtenidos por el algoritmo propuesto ante ataques cropping y traslaciones en la imagen *Mariposa*.

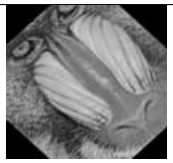
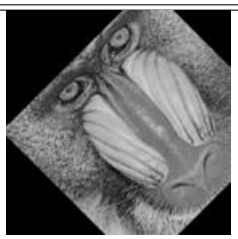
Ataque	Imagen atacada	Marcas reconstruidas		
		Marca 1	Marca 2	Marca 3
Cropping				
	NC	1	0.9890	0.9945
	AR	1	0.9764	0.9595
Cropping + rotación				
	NC	0.9230	0.9516	0.9798
	AR	0.8984	0.8553	0.8779
Traslación				
	NC	1	0.9930	0.9954
	AR	0.9980	0.9794	0.9685
Traslación + rotación				
	NC	0.9125	0.9623	0.9884
	AR	0.8955	0.9144	0.9167

Tabla 5.10: Resultados obtenidos por el algoritmo propuesto ante ataques cropping y traslaciones en la imagen *Mandril*.

De los resultados obtenidos se observa que el ataque que afecta menos al algoritmo propuesto es el ataque de cropping. Como se mencionó anteriormente el ataque de cropping fue el mismo en todas las imágenes. Sin embargo, mientras que en el caso de la imagen $Mandril$ se obtuvo una precisión y correlación de 1, en la imagen mariposa se logró una correlación de 0.95 y una precisión de 0.93. Esto se debe a que nuestro esquema cambia con las características de cada imagen. En el caso particular del ataque de cropping, a pesar de que no se cambian los valores de píxeles y en teoría se puede recuperar la marca perfectamente, el problema se presenta en la fase de sincronización; que de igual forma depende de la información de la imagen. En particular la localización de los puntos de referencia depende de los pixeles que los rodean y que se ven alterados por el cropping.

El segundo ataque que se realizó bajo condiciones iguales a todas las imágenes consistió en rotar 45º la imagen resultante del ataque cropping. Como se observa, en todos los casos los valores AR y NC disminuyen con respecto a los obtenidos en el primer ataque de cropping. Esto se debe a que un ataque de rotación no solo afecta la fase de sincronización sino que además altera los valores de los pixeles de la imagen debido a la interpolación que realiza.

Los ataques de traslación igual que los de cropping sólo afectan al algoritmo en su fase de sincronización, pues los pixeles son alterados sólo en su posición y no en su valor. Obsérvese por ejemplo, que en la imagen $Mandrill$ cuando se intenta reconstruir el logo del INAOE después de un ataque de traslación se obtienen los segundos mejores resultados de la prueba con valores de NC y AR de 1 y 0.998 respectivamente. Sin embargo, cuando la imagen trasladada se rota 45º las métricas de NC y AR de las marcas recuperadas disminuyen a 0.912 y 0.8995.

El ataque más fuerte realizado se presenta en la última fila de la tabla 5.9. Durante este ataque la imagen fue trasladada 212 columnas a la derecha y 62 pixeles hacia abajo y después fue rotada 60º. La primera columna de la tabla 5.9 muestra la imagen atacada en la que se puede observar la gravedad del ataque. No obstante, a pesar de deformar considerablemente la imagen es posible distinguir las marcas reconstruidas,

y los valores de NC y AR son de 0.87 y 0.84.

En total se realizaron 48 pruebas entre ataques de cropping, traslación y la combinación traslación + rotación. Este número resultó al considerar para las pruebas, cuatro imágenes de prueba, tres marcas y cuatro ataques por imagen. En todos los casos las marcas se pudieron recuperar con valores de NC arriba de 0.90 y AR arriba de 0.84. Por lo tanto se puede concluir que el algoritmo propuesto es robusto a ataques de cropping y translación.

5.3. Discusión de resultados

De los resultados obtenidos ante los ataques de compresión $JPEG$, adición de ruido gaussiano y sal y pimienta, aumento de contraste, y ecualización, mostrados en las figuras 5.4 y 5.5, así como en la tabla 5.1; se puede concluir que el algoritmo propuesto tiene buenos resultados en la reconstrucción de la marca después de este tipo de ataques alcanzando una precisión arriba de 0.90 en la mayoría de los ataques, excepto en dos casos durante la adicción de ruido gaussiano en los que descendió a 0.87. En general, se observó que la adición de ruido gaussiano fue el ataque que más afecto al algoritmo propuesto. Esto se debe a que este ataque en particular afecta al algoritmo $SIFT$ la localización de los puntos característicos y por lo tanto además de las alteraciones que sufren los pixeles de la imagen también afecta a la fase de sincronización.

Con respecto a los algoritmos propuestos por Shu-Fen&Chin-Sheng, Lou et al. y Hsieh et al. [30, 9, 11] el algoritmo propuesto obtuvo mejores resultados en la precisión de la marca reconstruida para la mayoría de los ataques. Se observó que debido a la forma en que codifican la imagen compartida, los algoritmos se ven afectados más por algún ataque en particular. Así, por ejemplo, mientras que el algoritmo obtiene una precisión de 1 en 15 de las 16 pruebas realizadas en compresión $JPEG$ (figura 5.4) se ve más afectado por la ecualización que por el aumento en el contraste de la imagen (tabla 5.1). En el caso del algoritmo propuesto, durante las pruebas realizadas con los ataques de compresión, ecualización, contraste y adición de ruido obtuvo para todos

ellos uno de los 2 valores más altos en la precisión de la marca reconstruida.

En la sección 5.2.2 se mostró que el esquema propuesto es menos susceptible a errores de falsos positivos que los algoritmos base. Esto se debe a que es difícil localizar un juego de puntos característicos similares en 2 imágenes distintas, pero a demás el método de codificación que utiliza el algoritmo propuesto presenta menos problemas.

En la sección 5.2.3 se muestran los resultados obtenidos ante ataques de rotación, traslación, cropping y algunas combinaciones entre ellos. Para este tipo de ataques no se presenta una comparación con los esquemas base, porque éstos no cuentan con una fase de sincronización que les permita relocalizar la marca después de este tipo de ataques.

En el algoritmo propuesto la fase de sincronización usa como referencia los puntos característicos de la imagen. Al ser estos puntos diferentes para cada imagen, no se puede conocer a priori cuantos puntos de referencia habrá y cuantos de ellos podrán ser localizados después de un determinado ataque. Sin embargo, el algoritmo $SIFT$ seleccionado para la extracción de puntos característicos ha reportado buenos resultados en la repetitividad de sus puntos. Así, de un total de 108 pruebas realizadas con ataques de traslación, rotación y cropping sólo en un caso (Tabla 5.6) no se pudieron relocalizar los puntos de referencia.

Finalmente, de los resultados obtenidos en todas las pruebas realizadas se puede concluir que el algoritmo propuesto no sólo tiene las ventajas de los algoritmos basados en secreto compartido visual, sino que además se logró establecer una fase de sincronización que le permite ser robusto a los ataques de rotación, translación y cropping.

Capítulo 6

Conclusiones y trabajo futuros

6.1. Conclusiones

En el presente trabajo de investigación se propuso e implementó en software un nuevo esquema de marca de agua que trabaja con imágenes en escala de grises. El esquema propuesto no sigue el esquema tradicional en el que la marca de agua se inserta dentro de la imagen, causando una degradación en mayor o menor medida según la función de inserción utilizada. En cambio, el esquema propuesto utiliza la técnica criptográfica de secreto compartido visual para crear una imagen sombra (imagen compartida) a partir de la combinación de características de la imagen y la marca de agua original. La imagen compartida se utiliza posteriormente para reconstruir la marca a partir de la combinación de ésta con las características de la imagen atacada.

Al no insertar la marca de agua en la imagen no se altera de ninguna forma la calidad de la misma. Otra ventaja que implica el no insertar la marca es que, ni los ataques visuales ni los estadísticos pueden detectar la marca. Además de que tampoco se puede extraer la información de la imagen pues no se ha insertado nada.

Al usar un esquema de secreto compartido se incrementa la seguridad del sistema pues se necesita la imagen compartida para reconstruir la marca de agua.

Para lograr que el algoritmo propuesto fuera robusto a rotación, traslación y cropping

se utilizaron características locales invariantes de la imagen durante la fase de sincronización.

Los resultados demuestran que el algoritmo propuesto es robusto a manipulaciones comunes en la imagen como lo son JPEG, ecualización, aumento en contraste y adición de ruido, que para nuestro caso las pruebas se realizaron con ruido gaussiano y ruido sal y pimienta.

En general, el algoritmo propuesto tiene las principales características de los esquemas basados en secreto compartido, estas son: 1) La imagen no se altera 2) La marca es indetectable e irremovible, 3) La seguridad del sistema es mayor a la de los esquemas tradicionales pues para reconstruir la marca se necesita la imagen compartida. Sin embargo, como lo demuestran los resultados de los ataques mostrados en el capítulo anterior, nuestro algoritmo además es robusto a rotación, traslación y cropping. Y más aún, no necesita calcular los grados que se ha rotado la imagen o la distancia con la que se desplazó para regresar la imagen a su posición original para poder recuperar la marca. En nuestro algoritmo la nueva posición de la información requerida para reconstruir la marca se determina mediante los descriptores de puntos característicos de la imagen.

6.2. Trabajo futuro

Indudablemente, el presente trabajo tiene limitaciones que podrían ser superadas en un futuro cercano. A continuación se presentan tres de los aspectos que se pueden mejorar como parte de un trabajo futuro:

- El algoritmo propuesto trabaja sólo a imágenes en escala de grises, por lo que el siguiente paso es adecuar el algoritmo para que trabaje con imágenes a color. Esto se podría hacer aplicando el algoritmo para cada plano del RGB, sin embargo; se considera que si se encuentra la manera de fusionar la información de los planos en una sola imagen se lograrían mejores resultados ante ataques de filtrado de la imagen.

- Una característica del algoritmo propuesto que se podría considerar como una desventaja, es el hecho de necesitar información extra (secreto compartido) para la reconstrucción de la marca. Esta característica es inevitable para los esquemas que usan secreto compartido, sin embargo, se puede reducir la cantidad de almacenamiento necesario para guardar el secreto compartido. Así, se podría proponer un nuevo método de codificación de la imagen compartida que redujera su tamaño o trabajar con los descriptores de los puntos característicos para reducir su tamaño.

- Finalmente, se piensa que se puede trabajar en una nueva forma de codificación de la imagen compartida que permita al algoritmo recuperar la marca después de un ataque de escalamiento de la imagen sin tener que regresar la imagen a su tamaño original.

Referencias

[1] A. Pfitzmann A. Westfeld. "Attacks on steganographic systems.". In *Third International Workshop on Information Hiding*, pag. 61 - 76. Springer-Verlag London, UK, 1999.

[2] K. HyungJun. B. KeunDuck, L. Sangjín. "A watermarking method using quantization and statistical characteristics of wavelet transform.". In *Sixth Conference on parallel and Distributed Computing Applications and Thecnoligies, 2005*.

[3] G. R. Blakley. "Safeguarding cryptographic keys.". *in proceedings of the National Computer Conference*, vol.48, pag. 313317, 1979.

[4] S Siggelkow H Burkhardt. "Local invariant feature histograms for image retrieval." Internal report, AlbertLudwigsUniversitat Freiburg, IIFLMB, January 1998.

[5] C.-T. C. K., Li. "Semi-fragile watermarking scheme for authentication of jpeg images." *Proceeding of the IEEE international Conference on Information Technology: Coding and Computing.*, pag. 7-11, 2004.

[6] J. Bloom I. Cox-M. Miller C. Lin, M. Wu and Y. Lui. "Rotation, scale, and translation resilient watermarking for images." *IEEE Transactions on Image Processing*, vol. 10, pag. 767-782, 2001.

[7] Ling-ge J. Hong-xia W. D. Ke, Chen H. "Geometrically invariant watermarking based on gravity center." *IEICE Trans. Fundamentals*, vol. E87-A (2), 2004.

[8] H. Wolfson D. Reisfeld and Y. Yeshurun. "Context free attentional operators: the generalized symmetry transform." *International Journal of Computer Vision*, vol. 14(2), pag. 119-130, 1995.

[9] H.K. Tso D.C Lou, J.M. Shie. "Copyright protection scheme based on chaos and secret sharing techniques." *Optical Engineering SPIE*, vol.44, 2005.

[10] R. von der Heydt-E. Peterhans O. Kubler. F. Heitger, L. Rosenthaler. "Simulation of neural contour mechanisms: from simple to end-stopped cells." *Vision Research*, vol. 32(5), pag. 963-981, 1992.

[11] S.L. Hsieh, L.Y. Hsu and I.J. Tsai. " A copyright protection scheme for color images based on image secret sharing and wavelet transformation." *TRANSACTIONS ON ENGINEERING, COMPUTING AND TECHNOLOGY*, vol. 10, pag. 1723, 2005.

[12] H.K H.Y. Lee, H. Kim. "Robust image watermarking using local invariant features." *Society of Photo-Optical Instrumentation Engineer*, vol. 45(3), 2006.

[13] R. Du J. Fridrich, M. Goljan. "Reliable detection of lsb steganography in color and grayscale images." In *Proc. of the ACM Workshop on Multimedia and Security*, pag. 27-30, 2001.

[14] R. Du J. Frifrich. "Secure steganographic methods for palette images." *Information Hiding, 3rd nt. Workshop*, pag. 47-60, 2004.

[15] C. Schmid K. Mikolajczyk. "A performance evaluation of local descriptors." *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 2(10), 2005.

[16] Martin Kutter, Sviatoshlav Voloshynovskiy, and Alexander Herrigel. "The watermark copy attack." In *Electronic Imaging 2000, Security and Watermarking of Multimedia Content II*, 2000.

[17] Chang-Tsun Li. "Digital watermarking for multimedia security management." Technical report, Department of Computer Science University of Warwick Coventry CV4 7AL UK, 2004.

[18] Chang S.-F. Lin, C.-Y. "Semi-fragile watermarking for authenticating jpeg visual content." *Proceeding of the SPIE conference on Security and Watermarking of Multimedia Contents.*, pag. 140-151, 2000.

[19] D. G. Lowe. "Distinctive image features from scale-invariant keypoints." *Int. J. Comput.*, vol.60(2):91110, 2004.

[20] M.J. Lucena-López. *Criptografía y Seguridad en Ordenadores*. 2003.

[21] A. Shamir M.Ñaor. "Visual cryptography." *Advanced in Cryptology, EUROCRYPT94, LNCS 950*, pag. 1-12., 1995.

[22] A.H Tewfik M.D. Swanson, Zhu Bin. "Transparent robust image watermarking." In *Image Processing, 1996. Proceedings., International Conference on*, vol. 3, pag. 211-214, 1996.

[23] Cordelia Schmid Mikolajczyk, Krystian. "Indexing based on scale invariant interest points." In *International Conference on Computer Vision, Vancouver*, pag. 525-531, July 2001.

[24] B. Macq P. Bas, J.M. Chassery. "Geometrically invariant watermarking using feature points." *IEEE Trans. Image Processing*, vol. 11, pag. 1014-1028, 2002.

[25] Richard Popa. "Steganography: An analysis of steganographic techniques." Technical report, The Politehnica University of Timisoara Faculty of Automatics and Computers Department of Computer Science and Software Engineering, 1998.

[26] J. ' O. Ruanaidh and T. Pun. "Rotation, scale and translation invariant digital image watermarking." In *IEEE International Conference on Image Processing 1997 (ICIP 97), Santa Barbara, CA, USA*, 1997.

[27] T. Pun S. Pereira, Joseph J. K. Ó Ruanaidh. "Secure robust digital watermarking using the lapped orthogonal transform." In *in SPIE 11th Annual Symposium on Electronic Imaging: Science and Technology EI'99*, pag. 21-30, 1999.

[28] A. Shamir. "How to share a secret." *Communications of the ACM*, vol. 22(1), pag. 612613, 1979.

[29] G.J. Simmons. "The prisoners' problem and the subliminal channel." *CRYPTO*, pag. 411-431, 1983.

[30] HSU Ching-Sheng TCU Shu-Fen. "Btc-based watermarking scheme for digital images." *INFORMATION & SECURITY . An International Journal*, vol. 15(2), pag. 216-228, 2004.

[31] G. Voyatzis, N.Ñikolaidis, and I. Pitas. "Digital watermarking: an overview." In *9th European Signal Processing Conference (EUSIPCO'98)*, Island of Rhodes, Greece, pag. 8-11, 1998.

[32] I. Voyatzis, G. Pitas. "Applications of toral automorphisms in image watermarking." In *Image Processing, 1996. Proceedings., International Conference on*, vol. 2, pag. 237–240, Sep 1996.

[33] T.Tan X. Yu, Y. Wang. "On estimation of secret message length in jsteg-like steganography." In *Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04)*, pag. 673 - 676. IEEE Computer Society Washington, DC, USA, 2004.

[34] P.M. Chen Y.C. Hou. "An asymmetric watermarking scheme based on visual cryptography." *in Proc. IEEE Int. Conf. on Signal Processing*, vol. 2, pag.992995, 2000.