Ingeniería y Ciencia
ISSN:1794-9165 | ISSN-e: 2256-4314
ing. cienc., vol. 14, no. 27, pp. 53–74, enero-junio. 2018.
http://www.eafit.edu.co/ingciencia

Detección de manipulaciones en evidencia digital con fines forenses

Héctor Duvan Ortiz[1], Diego Renza[2] y Dora M. Ballesteros L.[3]

Recepción: 24-08-2017 | Aceptación: 10-04-2018 | En línea: 15-06-2018

PACS:89.20.Mn

doi:10.17230/ingciencia.14.27.3

Resumen

En términos de evidencia digital, la imagen de la escena de un crimen es un elemento importante en un proceso legal; razón por la cual es indispensable que se garantice su cadena de custodia. Si la imagen es modificada con operaciones como eliminación, duplicidad o difuminado de objetos, debe existir un mecanismo que identifique la existencia de dicha manipulación. En este artículo, se propone una solución a la cadena de custodia basada en una técnica conocida como marcado de agua. Con el propósito de validar la sensibilidad y efectividad del sistema propuesto, se aplican seis clases de manipulaciones a imágenes marcadas; encontrando una alta sensibilidad a manipulaciones pequeñas (ej. 0.25 % del tamaño de la imagen), dado que en todos los casos fue identificada la manipulación. Con la solución propuesta, la autoridad legal puede confiar en la cadena de custodia de la evidencia digital.

Palabras clave: Marcado de agua; marcado frágil; multimedia forense; detección de manipulaciones; conjetura de Collatz.

[1] Universidad Militar Nueva Granada, hduvanortiz@gmail.com, http://orcid.org/0000-0001-7299-4700, Bogotá D.C., Colombia.
[2] Universidad Militar Nueva Granada, diego.renza@unimilitar.edu.co, http://orcid.org/0000-0001-8073-3594, Bogotá D.C., Colombia.
[3] Universidad Militar Nueva Granada, dora.ballesteros@unimilitar.edu.co, http://orcid.org/0000-0003-3864-818X, Bogotá D.C., Colombia.

Tampering Detection on Digital Evidence for Forensics Purposes

Abstract
In terms of digital evidence, images of a crime scene play an important role in a legal process; therefore it is essential to guarantee their chain of custody. If the image is tampered with operations such as cropping, duplicity or blurring of objects, there must be a mechanism that identifies the existence of such manipulation. In this paper, a solution to the chain of custody based on a technique known as watermarking is proposed. In order to validate the sensitivity and effectiveness of the proposed system, six classes of manipulations are applied to watermarked images; finding a high sensitivity to small manipulations (e.g 0.25% of the image size), since in all cases the manipulation was identified. With the proposed solution, legal authority can rely on the chain of custody of the digital evidence.

Keywords: Watermarking; fragile watermarking; digital forensics; tampering detection; Collatz conjecture.

1 Introducción

Los avances tecnológicos han permitido el acceso de la mayor parte de la población, a herramientas capaces de registrar y compartir imágenes digitales de forma instantánea y en casi cualquier lugar. Las imágenes digitales por su naturaleza son susceptibles de ser afectadas por modificaciones que pueden llegar a alterar su contenido (ej. redes sociales, portales, entre otros). Esta amenaza ha dado origen a la necesidad de buscar métodos a través de los cuales se garantice la fidelidad de la información de la imagen [1]. Como resultado aparecieron los métodos de marcado de agua y esteganografía digital, los cuales corresponden a técnicas a través de las cuales se puede incrustar un contenido sobre otro, donde dicho contenido puede ser un archivo de audio, una imagen, un video o un texto [2].

El marcado de agua en imágenes corresponde a la práctica de ocultar información dentro de una imagen, con el fin de proteger la imagen contra modificaciones que alteren su contenido [3],[4]. En la esteganografía, se utiliza la imagen como un medio de transmisión capaz de ocultar en su interior información de alta importancia y confidencialidad [5].

Entre las variantes en el área del marcado de agua en imagen se encuentra el enfoque frágil, que consiste en incrustar algún tipo de información

que se denomina marca, dentro de una imagen, de tal forma que si se manipula la imagen marcada, la marca se degrade apreciablemente [6]. Ese tipo de marcado tiene una directa aplicación en el ámbito forense, dado que permite la detección y comprobación de alteraciones en imágenes que pueden llegar a ser pruebas fehacientes en procesos legales [7]. Como consideraciones importantes, se requiere que la marca oculta no sea perceptible a simple vista, que su ocultamiento no represente un alto grado de alteración de la imagen que la contiene [8], que el método a usar soporte ataques activos enfocados a degradar la marca [9], y que permita la extracción del contenido oculto [10].

En este sentido, se requiere que los esquemas de marcado frágil sean cada vez más eficientes, cumpliendo con las características anteriormente mencionadas. Los métodos más representativos involucran el uso de técnicas como la descomposición en valores singulares (SVD) [11],[12], esquemas basados en las características del sistema visual humano (SVH), la transformada Contourlet, la transformada Discreta del Coseno (DCT) y la Transformada Wavelet Discreta (DWT) y sistemas caoticos [13]. Se destaca la DWT por facilitar el análisis de la imagen y la robustez que brinda al ser utilizada; esta transformada consiste en descomponer una imagen en nuevos elementos llamados sub-bandas utilizando análisis multi-resolución, lo que permite obtener el componente de aproximación (con el mayor porcentaje de energía), y los detalles horizontales, verticales y diagonales. Esta descomposición puede realizarse iterativamente, es decir, a partir de un componente generar más sub-bandas o niveles [14].

Actualmente, los esquemas de marcado de agua frágil no se basan únicamente en un método matemático, sino en la integración de varios de ellos, con el fin de acondicionar la información utilizada en el proceso de ocultamiento [15],[16]. Una de las características más relevantes a tener en cuenta en el diseño de métodos enfocados al ámbito forense consiste en que la marca sea altamente sensible a manipulaciones de la imagen.

Bajo este contexto, en este artículo se propone un método de marcado frágil que consiste en la inserción de contenido en la imagen a través de un método de codificación basado en la conjetura de Collatz. La imagen marcada es altamente similar a la imagen original sin distorsionar su contenido, y a la vez altamente sensible a manipulaciones intencionales como eliminación de contenido o duplicidad, evaluadas mediante la utilización de

herramientas comerciales de edición.

2 Método propuesto

El esquema presentado en la Figura 1 resume el esquema propuesto para realizar un marcado frágil sobre una imagen en escala de gris, utilizando texto como marca. Aunque el método está especialmente diseñado para el ámbito forense, se puede utilizar en otro tipo de aplicaciones de marcado frágil como la protección de derechos de autor.

Este método incluye dos procesos: ocultamiento y recuperación (Figura 1). La imagen marcada corresponde a la evidencia, la cual está protegida con una marca ante manipulaciones intencionales que busquen alterar su contenido original.

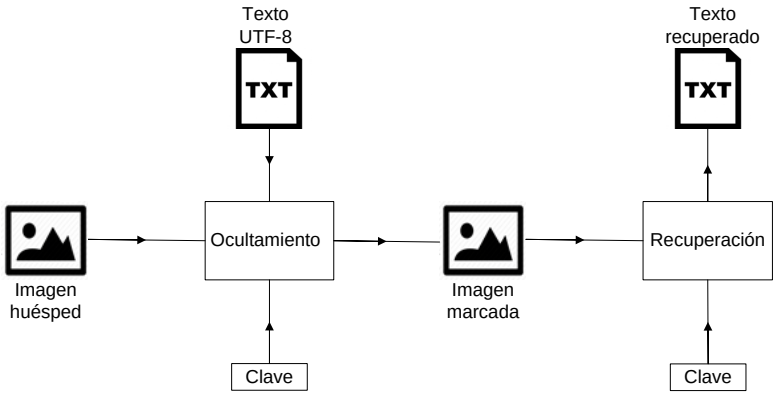


Figura 1: Esquema general propuesto de marcado frágil en imagen utilizando texto.

A continuación se describe el funcionamiento de los dos procesos principales del método propuesto, Ocultamiento y Recuperación.

2.1 Ocultamiento

En esta sección se presenta la metodología propuesta para realizar el marcado frágil utilizando texto como marca, la cual se resume en la Figura 2.

La metodología incluye los siguientes pasos: i) Aleatorización, ii) Ensanchamiento de espectro, iii) DWT, iv) Redundancia, v) QIM (Ocultar), vi) iDWT.

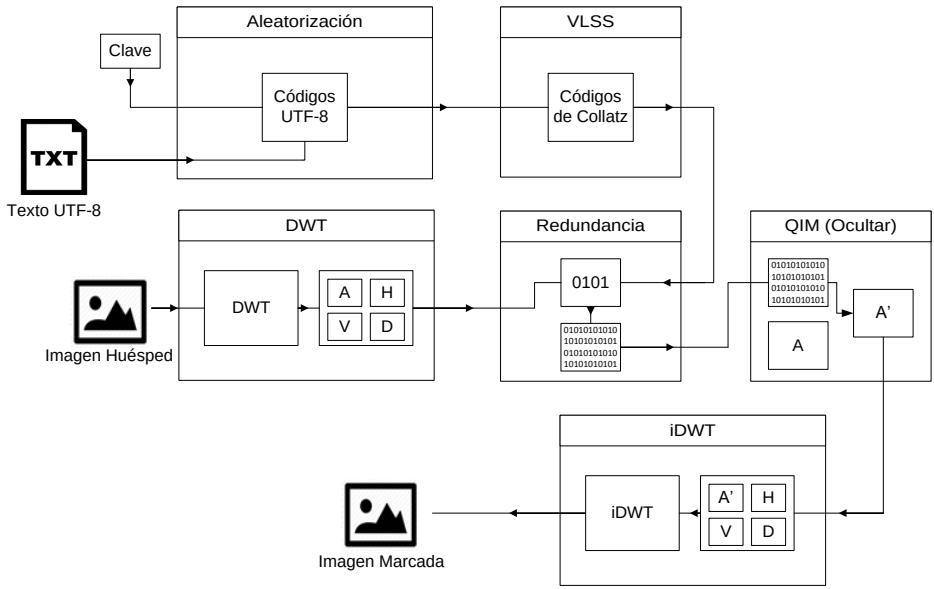


Figura 2: Esquema de ocultamiento propuesto para el marcado frágil en imagen con fines forenses.

2.1.1 Aleatorización Las entradas de este bloque corresponden a la clave y la marca (texto). En primer lugar se generan los 256 códigos UTF de 8 bits. Posteriormente, estos códigos son desordenados de acuerdo a la información contenida en la clave, la cual cumple la función de valor semilla dentro del proceso de aleatorización de los códigos. Posterior a la aleatorización, el valor decimal de cada carácter del texto de entrada permite seleccionar el respectivo código UTF-8 aleatorizado. Finalmente, cada uno de los códigos correspondiente a cada uno de los caracteres del texto de entrada se concatenan, obteniendo un vector binario de longitud 8*L, donde L es la longitud del texto.

2.1.2 Ensanchamiento de espectro con longitud variable (VLSS) Con el fin de asegurar que la marca se distribuya en toda la imagen, es

preciso realizar un proceso de ensanchamiento de espectro de los datos de entrada, lo que implica aumentar la tasa de bit de dichos datos. Para este propósito, cada código UTF-8 será reemplazado por un código proveniente del generador de códigos de longitud variable basado en la Conjetura de Collatz. Este generador permite obtener 256 códigos binarios diferentes. Por cada código UTF-8 de entrada, se obtiene un único código de salida, con la salvedad que la entrada es de longitud fija (8 bits) pero la salida es de longitud variable (entre 3 y 130 bits). De esta forma, al aumentar la tasa de bits se ensancha el espectro de los datos de entrada. Como resultado de esta fase se tiene una secuencia de bits que representa el texto de entrada mediante la utilización de códigos binarios obtenidos a través del uso de la conjetura de Collatz.

La conjetura de Collatz es un método matemático que permite reducir a la unidad un número entero ($N > 1$), a partir de la aplicación iterativa de las dos operaciones definidas en la Ecuación 1. Estas dos operaciones generan permutaciones que finalmente convergen a 1 [17]. Por ejemplo, al tener el número 10 se obtiene la siguiente secuencia: 10, 5, 16, 8, 4, 2, 1.

$$x = \begin{cases} x/2 \quad si \quad x \; mod \; 2 \equiv 0 \\ \\ 3x + 1 \quad si \quad x \; mod \; 2 \equiv 1 \end{cases} \tag{1}$$

La obtención del vector binario de longitud variable basado en la conjetura de Collatz consiste en el uso de los valores módulo, resultantes de la aplicación iterativa de la Ecuación 1 hasta alcanzar 1. Retomando el ejemplo del número 10, se obtiene la secuencia 10, 5, 16, 8, 4, 2, 1. En cada iteración, se tiene como valor módulo la secuencia: 0, 1, 0, 0, 0, 0, la cual se usa como representación binaria del número reducido; es decir, que el número 10 se representa mediante la conjetura de Collatz por el código 010000. La Figura 3 representa el método para obtener la secuencia binaria de un número entero positivo mediante la conjetura de Collatz.

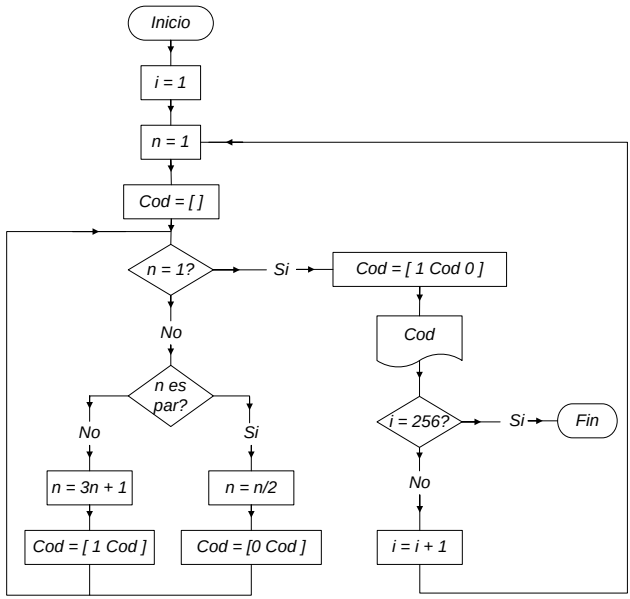


Figura 3: Diagrama de flujo utilizado para realizar la codificación en binario de $1 \leq N \leq 256$ por medio de la conjetura de Collatz.

Adicionalmente, se inserta un 1 al inicio de cada código de Collatz, como identificador (o cabecera) de cada trama. Posteriormente, los códigos provenientes de cada carácter se concatenan formando un solo vector binario de longitud $3 \cdot L \leq M \leq 130 \cdot L$, donde L corresponde a la cantidad de caracteres del texto de entrada (marca) y M es la longitud de la codificación de la marca utilizando la conjetura de Collatz.

2.1.3 Transformada wavelet discreta (DWT) La entrada de este bloque corresponde a la imagen (evidencia), a la cual se le aplica la DWT con un nivel de descomposición. El objetivo es el de seleccionar la sub-banda con el mayor nivel de energía (sub-banda de aproximación), componente que se utilizará como señal huésped de los datos a ocultar. A su vez, el método de marcado frágil debe garantizar la imperceptibilidad y no generar distorsiones significativas en el contenido original de la evidencia; esto se garantiza seleccionando un valor adecuado de cuantización (Δ) en el bloque QIM (Ocultar).

2.1.4 Redundancia Posterior a la representación binaria del texto con los códigos de Collatz, es necesario aplicar una etapa de redundancia que permita obtener una cadena binaria del mismo tamaño que el total de coeficientes de la sub-banda de aproximación.

En este caso, se calcula el valor entero de la relación entre el número de coeficientes de la sub-banda de aproximación y el número total de bits resultantes del bloque VLSS. La parte entera de esta relación corresponde a la cantidad de veces que se inserta el código ensanchado en el componente de aproximación.

2.1.5 QIM (Ocultar) Una vez se cuenta con el vector binario proveniente de la marca (texto) y la sub-banda de aproximación de la imagen (evidencia), se aplica el método QIM (Quantization Index Modulation) para el proceso de inserción.

El algoritmo QIM consiste en la utilización de una regla de cuantización basada en el valor del bit a ocultar y en el tamaño del paso de cuantización, Δ. De esta manera, el nuevo valor será un múltiplo entero de Δ (para un bit 0) o un múltiplo de Δ más un desplazamiento de $\Delta/2$ (para un bit 1) [Autor]. El proceso de ocultamiento se describe por medio la Ecuación 2.

$$S = \begin{cases} \Delta \left\lfloor \frac{h}{\Delta} \right\rfloor & si \quad w = 0 \\ \Delta \left\lfloor \frac{h}{\Delta} \right\rfloor + \frac{\Delta}{2} & si \quad w = 1 \end{cases} \tag{2}$$

Donde h es el valor del coeficiente de la sub-banda de aproximación, w es el valor de la marca (bit) y S es el nuevo valor del coeficiente cuantizado.

2.1.6 iDWT Finalmente, a partir de la sub-banda de aproximación modificada con el bloque QIM y las sub-bandas correspondientes a los detalles horizontales, verticales y diagonales, se aplica reconstrucción, obteniendo de nuevo una imagen. El resultado de esta fase corresponde a la imagen marcada, que tiene una alta similitud con la imagen original, pero a su vez incluye información que representa el texto de entrada.

2.2 Recuperación

En esta sección se presenta la metodología propuesta para realizar la recuperación del texto insertado en el proceso de ocultamiento. El esquema de recuperación se presenta en la Figura 4 y contiene los siguientes bloques: i) DWT, ii) QIM (Recuperar), iii) Ensanchamiento de espectro, iv) Des-aleatorización, v) Separación de códigos.

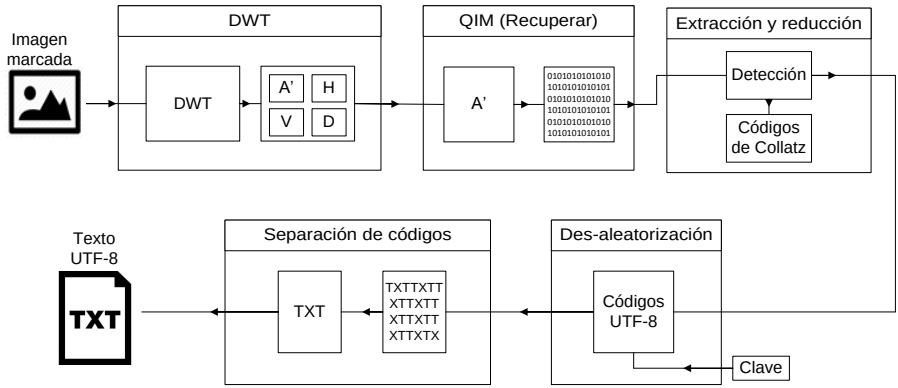


Figura 4: Esquema de recuperación propuesto para el marcado frágil en imagen utilizando texto, con fines forenses.

2.2.1 DWT El primer paso para realizar la extracción de la información consiste en aplicar la Transformada Wavelet Discreta sobre la imagen marcada con un nivel de descomposición, de forma similar a la realizada en el proceso de ocultamiento. Esto se realiza con el fin de extraer las sub-bandas de detalle y de aproximación.

2.2.2 QIM (Recuperar) Tomando como entrada la sub-banda de aproximación de la imagen marcada, se procede a realizar la extracción de la información binaria contenida. Para esto, se debe revertir el procedimiento aplicado para su ocultamiento, es decir, aplicar el algoritmo QIM inverso que permite la detección de '0's y '1's, a través de la Ecuación 3.

$$wr = \begin{cases} 1 & para \quad \frac{\Delta}{4} < \left| S - \Delta \left\lfloor \frac{h}{\Delta} \right\rfloor \right| \\ 0 & en\ otro\ caso \end{cases} \tag{3}$$

Donde wr corresponde al bit recuperado.

2.2.3 Extracción de códigos y Reducción de la tasa de bit Gracias a la inserción de la cabecera (bit '1') en cada uno de los códigos de Collatz, es posible realizar la separación de estos códigos de forma ciega (sin conocer la marca incrustada). Una vez separados los códigos, se realiza correlación entre el código de entrada y el conjunto de códigos de Collatz (todos los posibles códigos de los números entre 1 y 256). La posición del código de Collatz (valor decimal) que presente una mayor correlación con el código evaluado, indicará el valor de salida de este bloque y permitirá en la siguiente etapa obtener el carácter UTF-8 correspondiente. Lo anterior significa que la salida de este bloque es un vector que contiene los valores decimales de las posiciones (dentro del conjunto de códigos de Collatz) de los códigos ocultos.

2.2.4 Des-aleatorización A partir de las posiciones obtenidas en la fase anterior y por medio de la clave usada para aleatorizar, es posible recuperar el valor decimal de cada carácter del texto de entrada. Es decir, la clave permite des-aleatorizar los valores de entrada con el fin de obtener el valor decimal de los caracteres de entrada en formato UTF-8. Es necesario tener en cuenta que la clave debe ser la misma utilizada en el proceso de ocultamiento. El proceso anterior implica que la salida de este bloque es un vector que contiene los valores UTF-8 de los caracteres recuperados. A su vez, este vector representa n copias del texto de entrada, ya que los datos tienen la redundancia agregada en el módulo de ocultamiento.

2.2.5 Separación de códigos Debido a que en el proceso de ocultamiento se aplica redundancia para marcar la imagen completa, se requiere separar esta cadena de caracteres para obtener un único mensaje, el cual debe ser igual al texto original. Esta separación se realiza a través del cálculo de la cantidad de veces que se repiten los caracteres recuperados, estimando la posición relativa de cada carácter para posteriormente asignarlas según corresponda, obteniendo así el mensaje oculto.

3 Validación y resultados

Para la validación del método propuesto se realizaron pruebas, destinadas a verificar dos parámetros de la marca: i) Imperceptibilidad, que implica la capacidad que tiene un método para evitar que la marca insertada en la imagen sea detectada a simple vista [18] y ii) Fragilidad, que corresponde a la facilidad que tiene la marca para deteriorarse ante manipulaciones como duplicidad o eliminación de contenido [19].

3.1 Protocolo de pruebas

Para la evaluación de estos dos parámetros se usaron las 10 imágenes en escala de gris mostradas en la Figura 5, utilizando como clave para todas las pruebas la palabra 'UMNG2017', y como texto a ocultar la palabra 'Murciélago'. Las imágenes utilizadas se descargaron de bases de datos de uso libre (`http://www.unprofound.com/`, `http://www.public-domain-photos.com/`, `http://www.stockvault.net/`, `https://www.morguefile.com/`).

A partir de las 10 imágenes marcadas, se realizaron 6 tipos de modificaciones, que se detallan más adelante. Cada tipo de modificación se aplicó 5 veces para un total de 300 modificaciones, las cuales se realizaron a través de una herramienta en línea libre basada en Adobe Photoshop [20].

Adicionalmente, se evaluaron las imágenes resultantes a través de los siguientes índices de calidad: Índice de similitud estructural ($SSIM$) entre la imagen original y la imagen marcada y la distancia Hamming (HD) entre la marca original y la marca recuperada. La primera métrica se enfoca a la evaluación de imperceptibilidad de la marca, mientras que la segunda se enfoca en la fragilidad de la misma.

3.2 Validación de imperceptibilidad

Con el fin de evaluar el grado de similitud entre la imagen original y la marcada, se utilizaron las imágenes de la Figura 5. Esta comparación implicó una valoración visual de las imágenes marcadas, verificando la no aparición de artefactos que pudieran indicar la presencia de una marca en la imagen.

(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j)

Figura 5: Imágenes en escala de gris utilizadas en el proceso de validación de imperceptibilidad y fragilidad del método propuesto.

Para corroborar la verificación visual, en la Figura 6 se muestra un ejemplo de una imagen huésped y la respectiva imagen marcada con el método propuesto.

Figura 6: Ejemplo de comparación visual entre una imagen huésped y una imagen marcada a través del método propuesto.

Adicionalmente, para obtener una evaluación cuantitativa, se aplicó el índice $SSIM$ entre la imagen original y la imagen marcada, el cual indica su grado de similitud con base en las características del sistema de visión humano. El índice $SSIM$ varía en una escala de 0 a 1, donde 1 indica que las imágenes son idénticas y 0 que son totalmente diferentes [21]. El índice $SSIM$ está dado por la Ecuación 4.

$$SSIM = \frac{(2\mu_A\mu_B + C_1)(2\sigma_{AB} + C_2)}{(\mu_A^2 + \mu_B^2 + C_1)(\sigma_A^2 + \sigma_B^2 + C_2)} \tag{4}$$

Donde μ es la media, σ es la desviación estándar y C_1, C_2 son variables de estabilización.

3.3 Validación de fragilidad

De forma similar al proceso para evaluar la similitud, se marcaron las imágenes en escala de gris y sobre las imágenes resultantes se realizaron modificaciones tales como: i) quitar elementos de la imagen, ii) duplicar elementos de la imagen, iii) difuminar zonas determinadas de la imagen, iv) realizar correcciones puntuales sobre la imagen, v) aclarar zonas de la imagen y vi) oscurecer zonas de la imagen, tal como se presenta en la Figura 7.

(a) Imagen modificada quitando elementos

(b) Imagen modificada duplicando elementos

(c) Imagen modificada difuminando elementos

(d) Imagen modificada realizando correcciones puntuales

(e) Imagen modificada aclarando elementos

(f) Imagen modificada oscureciendo elementos

Figura 7: Ejemplos de modificaciones realizadas a las imágenes en escala de gris utilizadas en el proceso de validación de fragilidad del método propuesto.

Las manipulaciones se aplicaron sobre una zona muy pequeña de la imagen, que corresponde máximo al 0.25 % de los pixeles. Por ejemplo, para una imagen de 512×512 pixeles, se modificaron como máximo 655 pixeles de los 262144 que contiene la imagen. Posterior a la aplicación de las modificaciones, se utiliza el algoritmo de recuperación con el ánimo de verificar, comparar y evaluar las diferencias entre la marca original y la marca recuperada. En este sentido, se calcula la Distancia Hamming (HD) entre las marcas para cuantificar el porcentaje de bits que cambiaron en la marca luego de la manipulación. Para el cálculo de HD se utilizó la Ecuación 5.

$$HD_{ij} = (w_i \oplus w_j) \tag{5}$$

Donde w_i y w_j corresponden a la marca original y la marca recuperada después de la manipulación, y $\oplus$ es la suma modular entre las dos secuencias [22]. Una vez se obtiene el valor total de bits que difieren entre las marcas, se calcula su porcentaje, expresando su valor en una escala de 0 a 100 %. El valor máximo, 100 %, implica que las dos secuencias son completamente diferentes.

3.4 Resultados consolidados

En esta sección los resultados se agrupan utilizando gráficas de rango de confianza. Este tipo de gráficas muestra el valor mínimo y máximo de los resultados, y representa mediante un rectángulo el rango de valores en donde se ubica el 95 % de los resultados.

Para la presentación de los resultados, primero se consolidan las 300 simulaciones en términos del valor de $SSIM$ entre la imagen original y la imagen marcada (Figura 8), con el propósito de determinar si al insertar la marca se introduce distorsión significativa en la imagen (evidencia). Como se mencionó anteriormente, entre mayor sea el valor de $SSIM$ (acercándose a 1), mejores son los resultados en términos de imperceptibilidad de la marca.

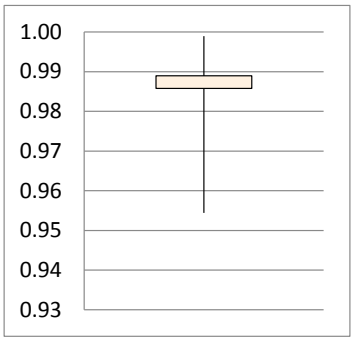


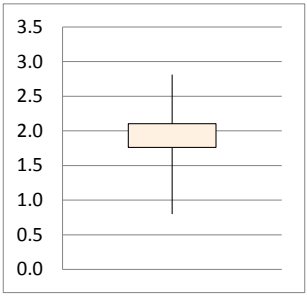
Figura 8: SSIM entre la imagen original y la imagen marcada: consolidado de 300 simulaciones.

Segundo, se consolidan los resultados de HD (en porcentaje) para cada una de las manipulaciones realizadas a la imagen marcada, correspondiendo a 50 pruebas por manipulación (Figura 9). Un valor de $HD > 0$ significa que el método es capaz de identificar la manipulación. Es importante aclarar que en un escenario real de manipulación forense, el tamaño de la zona manipulada puede ser muy pequeño en comparación con la imagen total, ocasionando que una pequeña manipulación pueda pasar inadvertida. Por lo tanto, las 300 pruebas realizadas se limitaron a un área máxima de manipulación del 0.25 % del tamaño de la imagen.

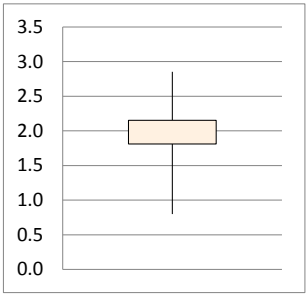
Finalmente, con el propósito de determinar la homogeneidad de la detección de manipulación (medida a través del parámetro HD), se consolidan los resultados de las 300 pruebas realizadas (Figura 10).

De acuerdo a los resultados de la Figura 8, en el peor de los casos la similitud entre las dos imágenes es del 95 %, y en la mayoría de los casos es aproximadamente igual al 99 %. Existen imágenes marcadas que llegan a ser casi exactamente igual a la imagen original. Estos altos valores de $SSIM$ se interpretan como una muy baja perceptibilidad de la marca, es decir, que con el método propuesto no se insertan distorsiones en la imagen que sean visualmente perceptibles y se puede considerar a la imagen marcada como una fiel copia (en términos de contenido) de la imagen original.

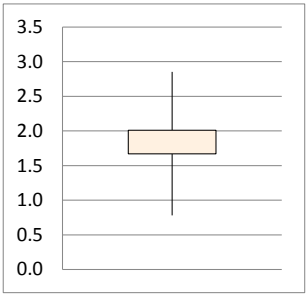
Por otro lado, con el propósito de determinar si el método permite la identificación de manipulaciones en la imagen marcada, aun cuando estas manipulaciones se apliquen en zonas extremadamente pequeñas de la



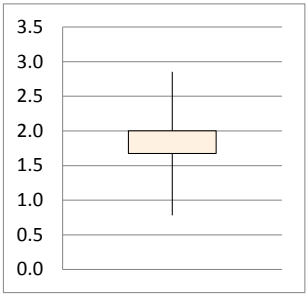
(a) HD (%) manipulación eliminación de objeto



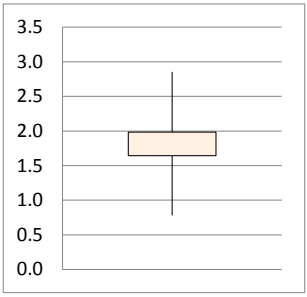
(b) HD (%) manipulación duplicidad de objeto



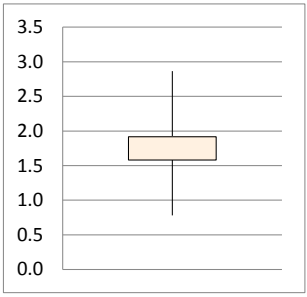
(c) HD (%) manipulación difuminar objeto



(d) HD (%) manipulación corrección de objeto



(e) HD (%) manipulación aclarar objeto



(f) HD (%) manipulación oscurecer objeto

Figura 9: Consolidado de resultados HD (%) entre la marca original y la recuperada, 50 resultados por manipulación.

imagen, se comparan la marca original y la recuperada mediante HD; estos resultados se muestran en la Figura 9. De estos resultados se tienen las siguientes observaciones: en todos los casos, los rangos de confianza se ubican alrededor del 2 %; el valor mínimo está por encima del 0.5 % y el valor máximo cercano al 3 %. Se resalta que el valor de HD es mayor al porcentaje de la zona manipulada, es decir, aún con una zona manipulada extremadamente pequeña ($\approx$0.25 %), las diferencias entre la marca original y la marca recuperada son cercanas a 2 % y en algunos casos a 3 %. Así mismo, en todos los casos, el valor de HD es diferente de cero, garantizando que siempre se detecten las manipulaciones. Como conclusión de lo anterior, el valor de HD es alrededor de 8 a 10 veces el porcentaje de la zona manipulada, y de esta forma se garantiza que cualquier cambio mínimo en la imagen se refleje en un cambio importante en la marca recuperada.

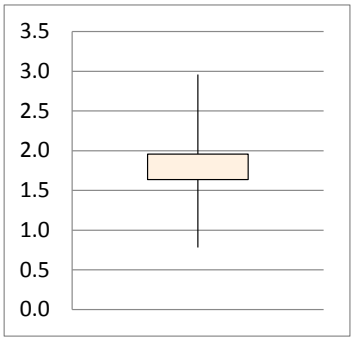


Figura 10: Consolidado de resultados HD (%), 300 simulaciones.

Adicionalmente, se muestran los resultados consolidados de la métrica HD (Figura 10). Como consecuencia de una respuesta similar en términos de HD a los diferentes tipos de manipulación, esta gráfica resume el comportamiento del HD ante las diferentes manipulaciones realizadas. Este comportamiento es positivo en términos de homogeneidad del método, dado que su respuesta es independiente del tipo de manipulación.

4 Conclusiones

En este documento se presenta un esquema de marcado frágil sobre una imagen en escala de gris utilizando texto como marca, con las siguientes características:

La marca no es perceptible al ojo humano, gracias a que el proceso de inserción se realiza en el dominio wavelet de la imagen, específicamente en la sub-banda de aproximación.

La marca corresponde a un texto, el cual se dispersa en toda la imagen, a través de varios procesos: códigos UTF de 8 bits, aleatorización de códigos, ensanchamiento de espectro a través de la codificación basada en la conjetura de Collatz y redundancia.

La codificación propuesta basada en la conjetura de Collatz permite obtener códigos de longitud variable, disminuyendo la predictibilidad de la inserción y dificultando la falsificación de la marca.

Gracias al uso del método QIM, el proceso de inserción de la marca es completamente reversible, es decir, que si no hay manipulación en la imagen, la marca recuperada es exactamente igual a la marca insertada.

Con las anteriores características, el método se puede aplicar en el ámbito forense, por las siguientes razones:

La marca insertada en la imagen que se presenta como evidencia, no genera distorsiones de contenido. Al ojo humano, estas imágenes son perceptualmente iguales.

Una pequeña modificación en algunos pixeles de la imagen, representa una mayor modificación en la marca recuperada, en relación con la marca original. Es decir, la marca es altamente sensible a manipulaciones, aun cuando estas se realicen en una zona concentrada y pequeña de la imagen y visualmente no se observen distorsiones de la misma.

Finalmente, proponemos como trabajo futuro:

Utilizar el método propuesto en una validación amplia de detección de manipulaciones en imágenes forenses, midiendo los falsos positivos, falsos negativos, verdaderos positivos y verdaderos negativos.

Complementar el método propuesto para poder identificar no solamente si la imagen ha sido manipulada, sino también la zona en la cual se realizó la manipulación.

Referencias

[1] L. M. Vargas, E. Vera de Payer, and A. Di Gianantonio, "Marcas de agua: una contribución a la seguridad de archivos digitales," *Revista de la Facultad de Ciencias Exactas, Físicas y Naturales*, vol. 3, no. 1, pp. 49–54, 2016. 54

[2] M. Cedillo Hernández, M. Nakano Miyatake, and H. Pérez Meana, "A robust watermarking technique based on image normalization," *Revista Facultad de Ingeniería Universidad de Antioquia*, no. 52, pp. 147–160, 2010. 54

[3] L. F. Huallpa Vargas and L. P. Yapu Quispe, "Watermark resistente en el dominio de las frecuencias de imágenes digitales para su autenticación segura mediante autómatas celulares," *Revista Investigación & Desarrollo*, vol. 1, no. 11, 2011. 54

[4] D. Renza, D. M. Ballesteros L, and H. D. Ortiz, "Text hiding in images based on qim and ovsf," *IEEE Latin America Transactions*, vol. 14, no. 3, pp. 1206–1212, 2016. 54

[5] A. Soria Lorente, R. A. Cumbrera González, and Y. Fonseca Reyna, "Algoritmo esteganográfico de clave privada en el dominio de la transformada discreta del coseno," *Revista Cubana de Ciencias Informáticas*, vol. 10, no. 2, pp. 116–131, 2016. 54

[6] X. Qi and X. Xin, "A singular-value-based semi-fragile watermarking scheme for image content authentication with tamper localization," *Journal of Visual Communication and Image Representation*, vol. 30, pp. 312–327, 2015. 55

[7] P. Singh and R. Chadha, "A survey of digital watermarking techniques, applications and attacks," *International Journal of Engineering and Innovative Technology (IJEIT)*, vol. 2, no. 9, pp. 165–175, 2013. 55

[8] N. Boujemaa, E. Yousef, L. Rachid, B. M. Aziz *et al.*, "Fragile watermarking of medical image for content authentication and security," *IJCSN-International Journal of Computer Science and Network*, vol. 5, no. 5, 2016. 55

[9] M. Botta, D. Cavagnino, and V. Pomponiu, "A successful attack and revision of a chaotic system based fragile watermarking scheme for image tamper detection," *AEU-International Journal of Electronics and Communications*, vol. 69, no. 1, pp. 242–245, 2015. 55

[10] Y.-C. Fan and Y.-Y. Hsu, "Novel fragile watermarking scheme using an artificial neural network for image authentication," *Applied Mathematics & Information Sciences*, vol. 9, no. 5, p. 2681, 2015. 55

[11] S. Dadkhah, A. Abd Manaf, Y. Hori, A. Ella Hassanien, and S. Sadeghi, "An effective svd-based image tampering detection and self-recovery using active watermarking," *Signal Processing: Image Communication*, vol. 29, no. 10, pp. 1197–1210, 2014. 55

[12] O. Benrhouma, H. Hermassi, and S. Belghith, "Security analysis and improvement of an active watermarking system for image tampering detection using a self-recovery scheme," *Multimedia Tools and Applications*, vol. 76, no. 20, pp. 21 133–21 156, 2017. 55

[13] X. Tong, Y. Liu, M. Zhang, and Y. Chen, "A novel chaos-based fragile watermarking for image tampering detection and self-recovery," *Signal Processing: Image Communication*, vol. 28, no. 3, pp. 301–308, 2013. 55

[14] A. Joshy and N. Suresh, "A dual security approach for image watermarking using aes and dwt," *International Journal of Digital Application & Contemporary research*, vol. 3, no. 1, 2014. 55

[15] D. Renza, D. M. Ballesteros L, and C. Lemus, "Authenticity verification of audio signals based on fragile watermarking for audio forensics," *Expert Systems with Applications*, vol. 91, pp. 211–222, 2018. 55

[16] D. Renza, C. Lemus, and D. M. Ballesteros L., "Audio authenticity and tampering detection based on information hiding and collatz p-bit code," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 8, pp. 1294–1304, Nov. 2017. 55

[17] T. Tarver, "The collatz conjecture: Determining an infinite convergent sequence," *Asian Journal of Mathematical Sciences (AJMS)*, vol. 1, no. 02, pp. 102–104, 2017. 58

[18] V. Gupta and A. Barve, "A review on image watermarking and its techniques," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 4, no. 1, pp. 92–97, 2014. 63

[19] S. M. Mousavi, A. Naghsh, and S. Abu-Bakar, "Watermarking techniques used in medical images: a survey," *Journal of digital imaging*, vol. 27, no. 6, pp. 714–729, 2014. 63

[20] P. EDITOR, "Photoshop online en español (en linea)," 2017, http://photoshopen.blogspot.com/. 63

[21] A. Abbasi, C. S. Woo, R. W. Ibrahim, and S. Islam, "Invariant domain watermarking using heaviside function of order alpha and fractional gaussian field," *PloS one*, vol. 10, no. 4, p. e0123427, 2015. 65

[22] J. Zhou, W. Sun, L. Dong, X. Liu, O. C. Au, and Y. Y. Tang, "Secure reversible image data hiding over encrypted domain via key modulation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 3, pp. 441–452, 2016. 67