TRABAJO FIN DE GRADO

GRADO EN INGENIERÍA INFORMÁTICA

CURSO 2016-2017

AUTENTICACIÓN DE IMÁGENES DIGITALES MEDIANTE PATRONES LOCALES DE TEXTURAS

Pablo Blanco Peris

María Solana González

Directores:

Luis Javier García Villalba

Ana Lucila Sandoval Orozco

Departamento de Ingeniería del Software e Inteligencia Artificial

---

FACULTAD DE INFORMÁTICA

UNIVERSIDAD COMPLUTENSE DE MADRID

Los abajo firmantes autorizan a la Universidad Complutense de Madrid (UCM) a difundir y utilizar con fines académicos, no comerciales y mencionando expresamente a su autor el presente Trabajo Fin de Grado: "Autenticación de Imágenes Digitales mediante Patrones Locales de Texturas", realizado durante el curso académico 2016-2017 bajo la dirección de Luis Javier García Villalba y Ana Lucila Sandoval Orozco en el Departamento de Ingeniería del Software e Inteligencia Artificial, y a la Biblioteca de la UCM a depositarlo en el Archivo Institucional E-Prints Complutense con el objeto de incrementar la difusión, uso e impacto del trabajo en Internet y garantizar su preservación y acceso a largo plazo.

Periodo de embargo:

■ 12 meses

Pablo Blanco Peris	María Solana González
Luis Javier García Villalba	Ana Lucila Sandoval Orozco

Agradecimientos

A nuestros directores Luis Javier García Villalba y Ana Lucila Sandoval Orozco y también a Esteban Armas Vega por darnos la oportunidad de poder trabajar con ellos, y por toda su dedicación durante todo este curso para poder llevar a cabo este trabajo.

A todos nuestros compañeros de la Facultad por su apoyo diario.

Y por último a toda nuestra familia que gracias a su esfuerzo nos han permitido realizar nuestros estudios.

Resumen

La autenticidad de una imagen digital sufre graves amenazas debido a la existencia de poderosas herramientas para la edición de imágenes digitales que facilitan la modificación del contenido de las mismas sin dejar huellas visibles de tales cambios. Este problema unido a la facilidad de distribución de la información a través de plataformas digitales como blogs, Internet o redes sociales, ha provocado que la sociedad tienda a aceptar como cierto todo lo que ve sin cuestionar su veracidad. En este trabajo se propone un método de autenticación de imágenes digitales mediante el análisis de patrones locales de textura. El sistema propuesto combina el patrón binario local con la transformada discreta wavelet y la transformada discreta del coseno para extraer las características de cada uno de los bloques de la imagen investigada. Posteriormente, se utiliza la máquina de soporte vectorial para crear el modelo que permita la verificación de la autenticidad de una imagen. Para la evaluación del método propuesto se realizaron experimentos con bases de datos públicas de imágenes falsificadas que son ampliamente utilizadas en la literatura.

Palabras clave

Dispositivos Móviles, Falsificación, Patrón binario local, Transformada Discreta del Coseno, Transformada Wavelet.

Abstract

The authenticity of a digital image suffers serious threats due to the existence of powerful digital image editing tools that make easier the modification of the contents of the image without leaving visible traces of such changes. This problem, joined with the ease of distribution of information through digital platforms such as blogs, the Internet or social networks, has caused society to tend to accept as true all that it sees without questioning its truthfulness. This work proposes a method of authentication of digital images through the analysis of local texture patterns. The proposed system combines the local binary pattern with the discrete wavelet transform and the discrete cosine transform to extract the characteristics of each blocks of the investigated image. Subsequently, the vector support machine is used to create the model that allows verification of the authenticity of an image. For the evaluation of the proposed method experiments were carried out with public databases of tampered images that are widely used in the literature.

Keywords

Mobile Devices, Falsification, Local Binary Pattern, Discrete Cosine Transform, Wavelet Transform.

.

Lista de Acrónimos

1NN	First Nearest Neighbor
BPPM	Mapa de Probabilidad Posterior en Bloque
CCD	Charged-Coupled-Device
CDR	Correct Detection Ratio
CFA	Color Filter Array
CMOS	Complementary Metal-Oxide-Semiconductor
CMY	Cyan-Magenta-Yellow
CYYM	Cyan-Yellow-Yellow-Magenta
DBC	Decision Boundary Carving
DCT	Discrete Cosine Transform
DIP	Digital Image Processor
DWT	Discrete Wavelet Transform
EC	Error Cumulants
EM	Expectation-Maximation
ERE	Eigenfeature Regularization
EXIF	Exchangeable Image File Format

FDR	False Detection Ratio
GA	Algoritmo Genético
GPS	Global Positioning System
GRGB	Green-Red-Green-Blue
HOGM	Histogram Of Orientated Gabor Magnitude
IQM	Image Quality Metrics
JPEG	Join Photograph Expert Group
LBP	Local Binary Pattern
NASA	National Aeronautics and Space Administration
NGS	Normalized Group
PNU	Pixel Non-Uniformity
PRNU	Photo Response Non-Uniformity
PSD	Photoshop Document
PSVM	Probabilistic Support Vector Machine
QF	Qualify Factor
QMF	Quadrature Mirror Filter
RGBE	Red-Green-Blue-Emerland

ROI	Region Of Interest
SFFS	Sequential Floating Forward Selection
SIFT	Scale Invariant Feature Transform
SPN	Sensor Pattern Noise
SVM	Support Vector Machine
TIFF	Tagged Image File Format

ÍNDICE

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

1. INTRODUCCIÓN

1.1. Motivación

Las imágenes digitales están presentes a día de hoy en todos sitios, en medios de comunicación, en Internet, etc. A pesar de que no todo el mundo es consciente de ello, la mayoría de estas imágenes han sido manipuladas, aunque sea, simplemente, para lograr un mejor aspecto. Sin embargo, hay cierto tipo de imágenes que debido a su importancia debe garantizar su autenticidad e integridad. Debido a la facilidad que existe de manipular las imágenes digitales muchas veces no se puede confirmar que la imagen que se está viendo sea real o manipulada. En los últimos años, ciertos sectores de la política, los medios de comunicación, e incluso la ciencia se han visto afectados por la falsificación de imágenes. Es por ello que imágenes que sean de gran importancia deben ser una garantía de integridad y autenticidad, para así, poder tomarlas como verdaderas. Un ejemplo de esto es la polémica generada por una imagen que dejaba ver la silueta de un hombre en el planeta Marte (ver Figura 1.1).

Figura 1.1: Foto captada por la NASA que deja ver una silueta humana en el planeta Marte.

La foto fue captada por una sonda de la NASA, Administración Nacional de la Aeronáutica y del Espacio, esta imagen estuvo en el foco de los noticieros en 2008, pero la autenticidad de dicha imagen sigue en duda a día de hoy[1].

El crecimiento exponencial en los últimos años de herramientas para manipular imágenes con una facilidad asombrosa, permite a cualquier persona manipular una imagen a su antojo sin necesidad de ser un experto y sin dejar rastro a la vista de la manipulación. Por otro lado, las herramientas con intención de detectar las imágenes alteradas han abundado por su escasez. [1]

La falsificación de imágenes nació con la propia fotografía. En los primeros años la fotografía se convirtió en un método para tener un retrato, por lo que los fotógrafos de la época aprendieron que manipulando ciertas fotografías incrementaban sus ventas. Algunos ejemplos de falsificación y manipulación de imágenes en situaciones importantes de la historia.

En la Figura 1.2 se muestra un claro ejemplo de manipulación de fotografías desde los principios de la propia fotografía. En la Figura 1.2.a se muestra la fotografía de Lenin y Trotsky, de la Unión Soviética, mientras que en la Figura 1.2.b se muestra la misma imagen manipulada, donde Trotsky y otra persona fueron removidas por motivos políticos, ya que, Trotsky, al ser un personaje no grato para la vida política de dicho país fue removido de muchas de las imágenes.

La falsificación de imágenes afecta también al sector político y, es por ello, que cada uno muestra lo que le interesa manipulando a su antojo las fotografías tomadas para conseguir sus objetivos a pesar de mostrar información que no sea auténtica.

(a) Imagen real (b) Imagen manipulada

Figura 1.2: Lenin y Trotsky

Asimismo, en la Figura 1.3 se aprecia como en la imagen (Figura 1.3.b) han sido eliminadas de la escena las personas que aparecen detrás del hombre que está en primer plano en la imagen original (Figura 1.3.a).

(a)Imagen Original (b)Imagen manipulada

Figura 1.3: Ejemplo de manipulación de imágenes con objetivo de ocultar contenido.

Otro caso se presenta en una de las imágenes más impactantes del 2005, en la que se aprecia la técnica de composición de imágenes a partir de dos imágenes diferentes. La fotografía mostrada en la Figura 1.4 fue tomada en una maniobra del ejército inglés en las costas africanas, pero posteriormente se descubrió que fue un montaje y se descubrió la imagen donde aparece el tiburón. La imagen de Figura 1.4.a está manipulada, ya que en realidad es la mezcla de las imágenes de las Figuras 1.4.b y 1.4.c.

(a) Imagen compuesta

(b) Imagen (1) (c) Imagen (2)

Figura 1.4: Ejemplo de empalme de imágenes.

En los ejemplos se observa la dificultad de identificar imágenes manipuladas o falsificadas. ya que, en muchos casos el acabado final tiene una calidad muy alta que hace prácticamente imposible que el ojo humano identifique la manipulación.

Con la llegada de algunas poderosas herramientas de edición de imágenes, manipularlas y cambiar sus contenidos se ha convertido en un hecho trivial, que se puede añadir, cambiar o borrar información significativa de una imagen, sin dejar huella de las alteraciones. Esto ha convertido en una prioridad el desarrollo de métodos para identificar las operaciones de falsificación y validación de la credibilidad de las imágenes digitales. Asimismo, el crecimiento exponencial que experimenta la tecnología y los poderosos

algoritmos de manipulación de imágenes, incluyendo software como Photoshop, Corel Draw y otros, se está complicando el hecho de distinguir entre una imagen auténtica y su versión modificada [2].

La manipulación de imágenes usando técnicas informáticas está ganando mucha popularidad y aceptación en áreas como investigación forense, tecnologías de la información, servicios de inteligencia, escáneres médicos, periodismo, efectos especiales y películas. Por todas estas razones el análisis forense de imágenes digitales de dispositivos móviles está teniendo mucho auge. El estudio debe ser concreto para este tipo de dispositivos, ya que poseen características específicas que permiten obtener mejores resultados.

1.2. Contexto

El presente Trabajo Fin de Grado se enmarca dentro de un proyecto de investigación titulado RAMSES aprobado por la Comisión Europea dentro del Programa Marco de Investigación e Innovación Horizonte 2020 (Convocatoria H2020-FCT-2015, Acción de Innovación, Número de Propuesta: 700326)y en el que participa el Grupo GASS del Departamento de Ingeniería del Software e Inteligencia Artificial de la Facultad de Informática de la Universidad Complutense de Madrid (Grupo de Análisis, Seguridad y Sistemas, http://gass.ucm.es, grupo 910623 del catálogo de grupos de investigación reconocidos por la UCM).

Además de la Universidad Complutense de Madrid participan las siguientes entidades:

- Treelogic Telemática y Lógica Racional para la Empresa Europea SL (España)
- Ministério da Justiça (Portugal)
- University of Kent (Reino Unido)
- Centro Ricerche e Studi su Sicurezza e Criminalità (Italia)

- Fachhochschule fur Offentliche Verwaltung und Rechtspflege in Bayern (Alemania)
- Trilateral Research & Consulting LLP (Reino Unido)
- Politecnico di Milano (Italia)
- Service Public Federal Interieur (Bélgica)
- Universitaet des Saarlandes (Alemania)
- Dirección General de Policía - Ministerio del Interior (España)

1.3. Objetivos

El presente Trabajo Fin de Grado (TFG) tiene los siguientes objetivos:

- Realizar un estudio de los trabajos e investigaciones existentes hasta el momento relacionados con los tipos de modificaciones existentes en las imágenes para realizar una clasificación de las técnicas más relevantes.

- Analizar las técnicas de análisis forense de autenticación de imágenes digitales existentes con objeto de analizar y comprender las más relevantes.

- Implementar un método que permita determinar si una imagen digital ha sido manipulada mediante el análisis del contenido de la imagen.

1.4. Plan de Trabajo

La planificación del desarrollo del proyecto fue la siguiente: Primero, se realizaron reuniones semanales con los tutores para definir los objetivos y el alcance del Trabajo de Fin de Grado. Una vez definido el trabajo a realizar, se iniciaron las actividades de estudio de la literatura relacionada con la investigación para tener una visión del estado actual de la misma. Posteriormente, se llevaron a cabo actividades relacionadas con diseño y posterior implementación del algoritmo definido en la fase anterior. En paralelo, se realizaron actividades de seguimiento y control del avance del

proyecto, revisando cada una de las actividades acordadas. Las principales tareas realizadas en esta etapa son: Especificación de los requisitos, diseño e implementación del algoritmo y por último las pruebas del algoritmo con varios bases de datos de fotografías públicas y generados durante el desarrollo del proyecto. Finalmente, en paralelo a las etapas descritas, se generó toda la documentación requerida para el Trabajo Fin de Grado. Todas las actividades han sido realizadas de forma conjunta por los integrantes del grupo de trabajo.

1.5. Estructura de la Memoria

El trabajo está organizado en 8 capítulos, siendo el presente la introducción que describe la motivación que nos ha llevado a la realización de este proyecto, los objetivos, la planificación y estructura del trabajo.

En el capítulo 2, se presentan los conceptos generales sobre imágenes digitales. En él se describe como está compuesta una cámara, la formación de una imagen y las diferentes técnicas de identificación de la fuente de imágenes que existen.

En el capítulo 3 se describen las diferentes técnicas de falsificación de imágenes digitales existentes como retoque, copia-pega, empalme, eliminación de la huella digital y doble compresión JPG.

En el capítulo 4 se analizan las diferentes técnicas de identificación de manipulación de imágenes digitales mencionadas.

En el capítulo 5 se presenta la contribución de este trabajo. En él se detalla el funcionamiento del método propuesto para detectar imágenes manipuladas. Asimismo, se describen los experimentos realizados para evaluar su funcionamiento y se analizan los resultados obtenidos.

En el capítulo 6 se presentan las conclusiones a las que se ha podido llegar en

este trabajo, y el trabajo futuro que se desprende del mismo.

Finalmente, en los capítulos 7 y 8 se encuentran la introducción y conclusiones en inglés.

2. IMÁGENES DIGITALES

Las cámaras fotográficas están formadas por varios componentes: un sistema de lentes, un grupo de filtros, una matriz de filtro de colores o CFA, un sensor de imagen y un procesador de imagen, al cual se le llama DIP.

El proceso de generación de una imagen digital se detalla en estos pasos:

1. El sistema de lentes capta la luz de la escena controlando la exposición, el foco y la estabilización de la imagen.
2. Posteriormente, la luz pasa por un grupo de filtros que mejora la calidad visual de la imagen (incluye al menos un filtro infrarrojo y un filtro "anti-aliasing").
3. El filtro infrarrojo absorbe o refleja la luz para que sólo pase la parte visible del espectro a la siguiente fase, evitando perder cierta nitidez en la imagen. El filtro "anti-aliasing" limpia la señal produciendo contornos más suaves.
4. La luz pasa al sensor de la imagen, que genera una señal analógica proporcional a la intensidad de la luz recibida, a través de unos píxeles sensibles a la luz.
5. La señal analógica se convierte en digital y se transmite al procesador de imagen, que elimina el ruido y otras anomalías.
6. Se realiza la interpolación cromática, que consiste en calcular los valores de los colores faltantes debido a que el sensor proporciona información sobre ciertos colores dependiendo de la matriz CFA.
7. Se ejecutan algunos procesos de mejora, como la corrección de píxeles defectuosos y balanceo de blancos.
8. Por último, el proceso de corrección gamma ajusta los valores de intensidad de la imagen (suele variar entre fabricantes).
9. Finalmente, la imagen generada por el procesador se comprime.

Los pasos descritos anteriormente sobre la estructura básica de la formación de la imagen se describen en la Figura 2.1 manera sencilla y gráfica, dónde se puede apreciar de manera notable el cambio desde la escena inicial hasta la imagen final.

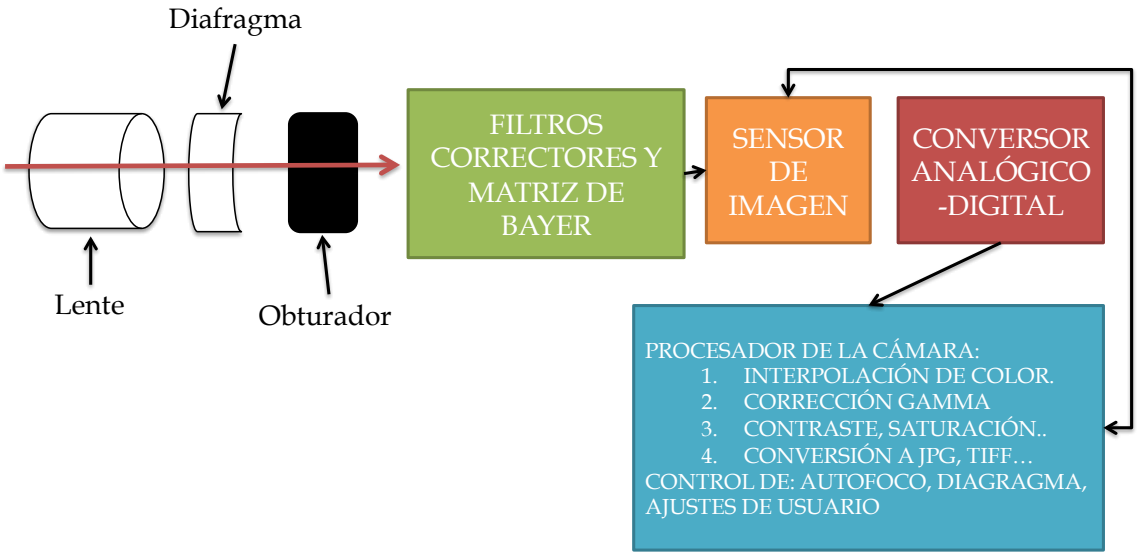


Figura 2.1: Proceso de adquisición de imágenes en cámaras digitales.

2.1. Matriz de Filtros de Color

Respecto a la matriz de filtros de color, o, CFA, es un componente que se encuentra sobre el sensor monocromo, y su función es adquirir la información del color de la escena.

La intensidad de la luz que pasa por cada una de las celdas forma una imagen en escala de grises y, dependiendo de la configuración del filtro CFA, se interpreta como una imagen a color. En este punto se realiza el proceso de la interpolación cromática para obtener los valores de los colores restantes. Generalmente, las cámaras usan el modelo GRGB, donde en la Figura 2.2 podemos ver un ejemplo. Pero hay varias alternativas de filtros CFA: CYYM, RGBE o CMY.

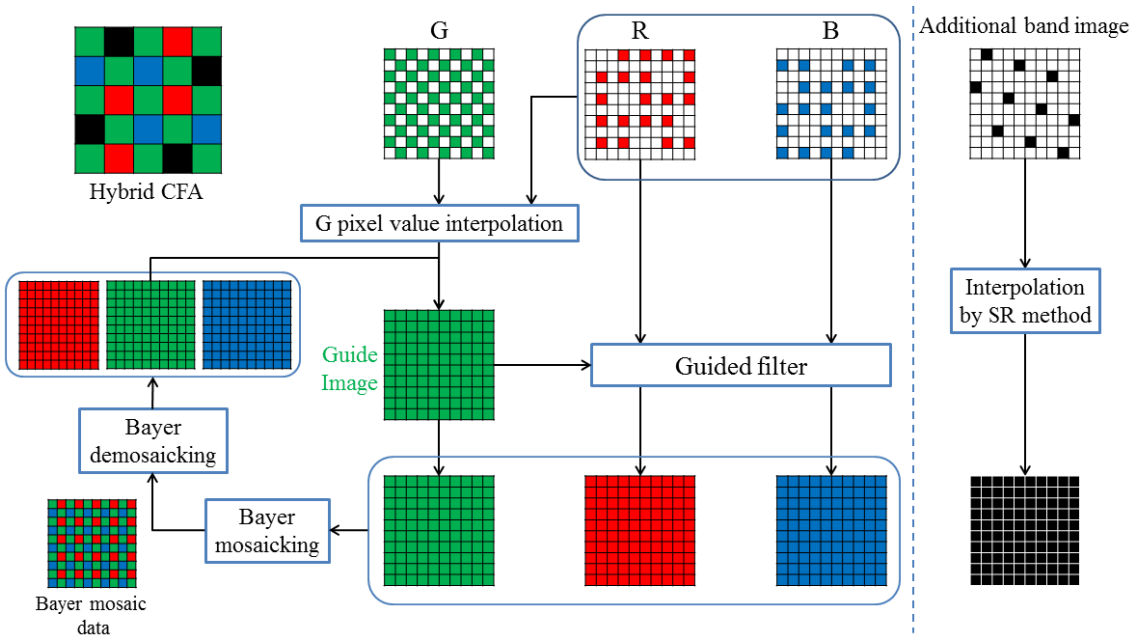


Figura 2.2: Matriz de color de filtros (CFA)

2.2. Sensor de la Imagen

El sensor de la imagen es la parte más importante de las cámaras digitales, se compone de unos píxeles hechos de silicio, sensibles a la luz que capturan la luz. Los sensores de la imagen pueden ser de tipo CCD (*Charged-Coupled-Device)*, y CMOS (*Complementary Metal-Oxide-Semiconductor*), dependiendo al proceso de fabricación [3].

Los sensores CCD están presentes en la mayoría de las cámaras digitales. Este tipo de sensor es menos sensible a la luz que el CMOS, por lo que captura un rango más amplio de tonos en las fotografías. Son más costosos en términos económicos y por norma general más grandes debido a sus componentes. Respecto al nivel de ruido generado los sensores CCD son superiores a los CMOS, ya que éstos generan imágenes de alta calidad con poco ruido digital, mientras que los sensores CMOS son más sensibles a ello.

Los sensores CMOS se encuentran por regla general en la mayoría de dispositivos móviles, es decir, smartphones, tabletas y demás. Son más sensibles a la luz que los CCD debido a que están compuestos por menos

componentes, ya que se busca un acabado minimalista, es por ello que suelen ser más pequeños que los CCD y también más económicos. Sin embargo, el proceso no es tan costoso y tan perfeccionista, por lo que el ruido es más apreciable en este tipo de sensores.

Durante la generación de una imagen es posible que se produzcan defectos que se vean reflejados como ruido en la imagen final. Estos defectos característicos pueden determinar la cámara que generó cierta imagen. Los defectos se pueden agrupar en: Defectos de fila y columna, defectos de grupo, píxeles calientes, Píxeles muertos, diferencias entre salidas múltiples, interferencia, saturación, corriente de oscuridad.

2.3. Ruido en la Imagen

En el proceso de generación de la imagen en la cámara se puede producir gran cantidad de imperfecciones y ruido. El patrón de ruido consiste en cualquier patrón espacial que no cambia de una imagen a otra, compuesto por el ruido espacial, totalmente independiente del ruido de patrón fijo (FPN). El ruido FPN se genera en función de: la oscuridad, la exposición y la temperatura. El ruido PRNU (*Photo Response Non-Uniformity*), es la parte dominante del patrón de ruido de las imágenes, compuesto por el ruido PNU (*Pixel Non-Uniformity*), y los defectos de baja frecuencia, tales como el zoom. El ruido PNU es la diferencia de sensibilidad a la luz entre los píxeles de la matriz del sensor. Se encuentra con mayor frecuencia en los sensores de tipo CMOS.

2.4. Compresión JPEG

Es un estándar muy popular para la codificación de imágenes digitales. Utiliza varias técnicas para garantizar tasas de compresión muy altas a expensas de una pequeña degradación de la calidad de la imagen. En pocas palabras, esta técnica supone que una imagen de entrada está codificada en el

formato YCbCr, donde el canal Y contiene la componente de luminancia de la imagen, y los canales de crominancia Cb y Cr mantienen, respectivamente, el componente luminoso menos azul de la imagen y la componente de luminancia menos roja de la imagen. Una vez en el dominio de la frecuencia, estos coeficientes se comprimen reduciendo drásticamente la cantidad de información proporcionada por las altas frecuencias y, a continuación, redondeando los valores resultantes. El factor de calidad de una imagen JPEG puede variar en el rango [1, 100], donde valores más pequeños resultan en una calidad inferior de la imagen comprimida y un mayor grado de compresión.

2.5. Diferencias entre Cámaras Digitales y Cámaras de Dispositivos Móviles

Si bien es cierto que las imágenes se procesan de manera similar tanto en cámaras digitales como en dispositivos móviles, hay algunas diferencias importantes relativas a la calidad. Las cámaras integradas en dispositivos móviles son de menor calidad debido al hardware requerido para que el tamaño sea lo más reducido posible. Algunos de los factores que causan estas diferencias son:

- **Apertura de la lente**: limitada, valores pequeños.
- **Resolución**: menor.
- **Distancia focal**: fija y restringida a valores pequeños lo que supone que las condiciones de iluminación sean limitadas.
- **Flash**: no muy potente.
- **Conversión Analógica Digital**: conversión de 10 bits en comparación a las cámaras tradicionales que poseen 12 bits.

2.6. Técnicas de Análisis Forense

Entre las técnicas de análisis forense en imágenes digitales, las dos ramas más importantes son:

1. **Verificación de integridad o detección de falsificaciones**: trata de encontrar técnicas que se hayan incorporado a la foto mediante recortes o adición de objetos.
2. **Identificación de la fuente**: como su nombre indica, encontrar el dispositivo que tomó una imagen o grabó un vídeo.

Este trabajo se centra en la primera, verificación de Integridad o detección de falsificaciones en imágenes digitales.

2.6.1. Técnicas de Identificación de la Fuente

El propósito de las técnicas de identificación de la fuente se centra en la identificación de marca, modelo y dispositivo empleado para la adquisición de una imagen digital. Hay que tener en cuenta que el éxito de estas técnicas depende en gran medida de la suposición de que todas las imágenes adquiridas por un dispositivo presentan una serie de características que lo hacen único. Dichas características que son empleadas para poder identificar tanto la marca como el modelo de las cámaras digitales, vienen derivadas de la presencia de diferencias entre las técnicas de procesamiento de imágenes y las tecnologías de los componentes que se emplean. A día de hoy el problema es que los modelos de las cámaras digitales existentes usan componentes de un número reducido de fabricantes y que los algoritmos llevados a cabo para la generación de las imágenes también son muy similares entre los modelos de una misma marca.

Según [4] hay 5 grupos de técnicas para la identificación de la fuente que se basa en cada uno de los procesos que intervienen en la generación de una imagen:

1. **Técnicas basadas en metadatos:** Consiste en la extracción de los datos de marca y modelo del objeto de estudio. Las cámaras digitales cuentan con una poderosa fuente de información que son los metadatos integrados en los archivos de las imágenes digitales. Estos registran información relacionada con las condiciones de captura de la imagen, como pueden ser la versión de software empleado, geolocalización, presencia o ausencia de flash o fecha y hora de adquisición. Los metadatos, entre otros usos, también pueden llegar a ser una gran ayuda a la hora de organizar y buscar en librerías de imágenes. Estas técnicas son las más sencillas, pero tienen mucha dependencia de la información que los fabricantes deciden insertar cuando la imagen es generada. [5] [6] realizan un estudio a fondo, donde se demuestra que los fabricantes no siguen fielmente la especificación EXIF, lo que puede tener como consecuencia la extracción errónea o inválida para fines forenses. Este método es uno de los más vulnerables en cuanto a modificaciones malintencionadas, ya que incluso puede darse el caso de la eliminación completa de los metadatos, intencionadamente o de manera inconsciente. Un claro ejemplo son aquellas aplicaciones o programas para editar imágenes, que como consecuencia de su uso actualizan incorrectamente los metadatos o provocan la pérdida de los mismos. Un ejemplo de utilidad de los metadatos se presenta en [7] donde se propone un método para fusionar la información de un clasificador del contenido de la imagen con las evidencias de los metadatos.
2. **Técnicas basadas en la aberración de las lentes:** Dependiendo del tipo de lente que utilice la cámara se pueden introducir diferentes tipos de aberraciones en la imagen. [8] La distorsión radial es la más común en cámaras de dispositivos móviles ya que usan lentes más baratas por cuestiones de coste. El grado de la distorsión radial en cada imagen se puede medir siguiendo un proceso que consta de tres fases: Detección de bordes, extracción de segmentos distorsionados y medición del error de la distorsión. En [9] se propone la aberración cromática lateral como técnica

para la identificación de la fuente.

3. **Técnicas basadas en la interpolación de la matriz CFA**: Los algoritmos de interpolación cromática generan algunas de las diferencias más marcadas entre los distintos modelos de cámaras. Dentro de este tipo de técnicas se engloban varios grupos: Huellas en la Interpolación del Color [10], modelo de correlación cuadrática de píxeles [11], medidas de similitud binarias [12], Información del proceso de interpolación [13], correlación inter-channel [14].
4. **Técnicas basadas en las características de las imágenes:** Estas técnicas utilizan un conjunto de características extraídas del contenido de la imagen para hacer la identificación de la fuente. [15] [7] [16] [17] combinan diferentes grupos de características (de color, métricas de calidad de la imagen (IQM) y estadísticas del dominio wavelet). En [18] se utiliza un algoritmo genético (GA) para buscar automáticamente las características óptimas y una Máquina de Soporte Vectorial (SVM del inglés Support Vector Machine) como clasificador. [19] [20] usan las características del dominio wavelet para integrar un modelo estadístico a partir de los coeficientes wavelet haciendo una descomposición wavelet en 4 niveles basada en Quadrature Mirror Filter (QMF). En [21] se plantea una técnica basada en los modelos estadísticos para ridgelet y sub-bandas contourlet que posteriormente aplica un algoritmo SFFS, Sequential Floating Forward Selection, para selección de características y SVM para la clasificación. En [22] se propone un método que emplea la densidad marginal de los coeficientes de la transformada discreta del coseno (DCT) en las coordenadas de baja frecuencia y las características de densidad conjunta de vecindad en el domino DCT.
5. **Técnicas basadas en el uso de las imperfecciones del sensor**: Estas técnicas se basan en el estudio de las huellas que los defectos del sensor pueden dejar sobre las imágenes. Se dividen en dos ramas: defectos de píxel y patrón de ruido del sensor (SPN del inglés Sensor Pattern Noise). En la primera se estudian los defectos de píxel, los píxeles calientes, los píxeles muertos, los

defectos de fila o columna, y los defectos de grupo [23]. En la segunda se construye un patrón del ruido dosificando los residuos de ruido obtenidos aplicando algún filtro de eliminación de ruido [24] [14]. Algunas investigaciones analizan la huella presente en diferentes regiones de la imagen [25] ya que las diferentes regiones de las imágenes pueden contener información distinta sobre la huella digital de la cámara fuente. Asimismo, debido a la propiedad determinista del patrón de ruido del sensor presente todas las imágenes, éste se puede usar como huella para identificar el dispositivo que generó la imagen [26].

En la Tabla 2.1 se presenta una comparación de los trabajos de identificación de la fuente de acuerdo a la técnica utilizada para la identificación.

Tabla 2.1. Técnicas de identificación de la fuente

Técnica	Propuesta	Clasificador	Nº Marcas	Nº Modelos	Formato	Resolución	Nº Imágenes	Móviles	Modelos	% Acierto
Aberración de las Lentes	[8]	SVM	3	3	JPEG	Variado	300	-	No	87'38%- 91'53%
	[9]	SVM	3	3	JPEG	Variado	1800	No	Sí	72'75% - 92'22%
Interpolación Matriz CFA	[10]	SVM	5	2	JPEG	Variado	600	No	Sí	84'8% - 88%
	[11]	Red Neuronal	4	4	S/C	-	800	No	Sí	98'25%
	[12]	SVM	3	9	-	Variado	200	Sí	Sí	81%-98%
	[13]	PSVM 1NN	4	11	JPEG	Variado	-	No	Sí	94'8%-99'4%
	[14]	1NN	3	4	JPEG	-	200	No	Sí	94'5%
Características de las Imágenes	[16]	SVN	2	4	JPEG	1600 x 1200	150	Sí	Sí	61'7% 99'72%
	[27]	SVM	5	5	JPEG	-	100	Sí	Sí	97'7%
	[22]	SVM	4	5	JPEG	Variado	599	Sí	Sí	86'3% 99'91%
	[20]	SVM	4	6	JPEG	-	350	-	Sí	98%
	[17]	SVM	4	10	JPEG	1024 x 1024	3000	No	Sí	47%92%
	[21]	SVM	3	3	-	-	2000	No	No	93'3% 99'7%
Imperfecciones del Sensor	[25]	-	1	2	-	640 x 480	-	No	-	-
	[24]	-	5	9	JPEG	Variado	2880	No	Sí	-
	[23]	SVM	9	25	JPEG	Variado	1250	2 de 25	Sí	94'49% 96'77%98'10%

3. FALSIFICACIÓN DE IMÁGENES DIGITALES

Con la cantidad de herramientas de edición de imágenes que existen hoy en día no es difícil modificar una imagen para cambiar el contenido que se muestra en ella. Asimismo, no es demasiado complicado insertar, cambiar o eliminar cierta información contenida en una imagen sin dejar rastro apreciable para el ojo humano. Esto plantea un problema de confianza en el ámbito de la autenticidad del contenido digital actual, especialmente cuando se presenta como evidencia en un tribunal para verificar algo en concreto.

El objetivo básico de la falsificación o manipulación de una imagen es la distorsión de la información. La historia de la falsificación de imágenes comenzó en 1840 cuando Hippolyte produjo una imagen falsa suya suicidándose. La primera imagen falsificada de la historia se muestra en la Figura 3.1.

Figura 3.1: Primera imagen falsificada

Más recientemente, en la Figura 3.2 se muestra a John Kerry (secretario de Estado estadounidense) con Jane Fonda (actriz de Hollywood) hablando a una multitud en un evento de paz. Esta foto fue manipulada durante la campaña

presidencial electoral de 2004, pero, en realidad, el evento nunca tuvo lugar. La foto original de John Kerry (Figura 3.2.b) fue tomada en 1971 en una concentración de la paz en Long Island, mientras que la foto de Jane Fonda (Figura 3.2.c) fue tomada en 1972.

(a) Manipulada

(b) Original 1

(c) Original 2

Figura 3.2: (a) representa una falsificación usando combinando las imágenes (b) y (c).

Existen varios tipos de manipulación de imágenes, a continuación, se presentan los más comunes: retoque de imágenes, copia-pega, falsificación mediante empalme de imágenes y modificación o eliminación de la huella digital [2]. Todas estas técnicas generan una doble compresión en las imágenes JPG por lo que también es objeto de estudio.

3.1. Retoque de Imágenes

Esta técnica comúnmente es usada en la industria de los medios de comunicación. Está aceptada y es un método de manipulación de imágenes muy usado, ya que, en cuanto a apariencia suele mejorar los resultados de las imágenes originales, dándoles un toque más atractivo. Es popularmente usada en revistas de fotos y películas. La imagen es alterada para proporcionar un mejor acabado. No está considerada una técnica de falsificación, pero está incluida porque incluye manipulación de la imagen original.

Es muy habitual ver imágenes de estas características teniendo en cuenta que existe una cantidad importante de aplicaciones que se dedican a ello, además de los cursos que se proporcionan hoy en día en este ámbito. Incluso los propios dispositivos móviles de hoy en día contienen software de fábrica para llevar a cabo esta edición.

Esta técnica se emplea mayoritariamente para crear una admirable representación de la belleza. Se basa en el copia-pega de la mayoría de las partes similares de la imagen con una regla bien definida. Es diferente de la técnica copia-pega en el sentido de que las regiones copiadas proceden de la misma área en lugar de diferentes regiones como en el copia-pega. Sus aplicaciones suelen ser para quitar texto, sellos, para eliminar arrugas o modificar de alguna forma ciertas siluetas o colores.

En la Figura 3.3 se puede observar como la imagen (b) ha sido retocada sin reflejar ningún tipo de marca que llame la atención o que haga sospechar que la imagen ha sido manipulada.

(a) Original (b) Modificada

Figura 3.3: Ejemplo de retoque de imágenes (1).

En la Figura 3.4 se observa la portada de un periódico español (ABC) que manipuló una fotografía cuando se produjo una agresión al presidente del gobierno español Mariano Rajoy durante un paseo electoral en Pontevedra en el año 2015. Las imágenes se hicieron virales debido al abuso de la manipulación en la imagen (b) que provocó que la gente se diera cuenta y aparecieran las comparaciones entre la original y la editada por el periódico.

(a) Original (b) Modificada

Figura 3.4: Ejemplo de retoque de imágenes (2).

3.2. Copiar y Pegar Regiones de una Imagen

Las falsificaciones de copia-pega es una de las técnicas más comunes empleadas, debido a su facilidad gracias a las herramientas típicas de procesamiento de imágenes. En este tipo de falsificación una región de la imagen es copiada y pegada en otra región de la misma foto, ya sea para eliminar cierto contenido o duplicar ciertos elementos de la imagen. Sin embargo, el proceso de detectar manipulaciones se vuelve más costoso debido al post-procesamiento de estas imágenes, ya que, para mejorar estas ediciones, el área copiada puede ser escalada, rotada o incluso difuminada.

Un ejemplo de esta técnica se muestra en la Figura 3.5 en la cual se aprecia en la primera imagen un solo pájaro, mientras que en la segunda aparecen dos pájaros debido a la clonación del contenido [7]. En este caso se lleva a cabo una duplicación del contenido, quizás para una imagen como esta, que no tenga ningún tipo de influencia no es relevante, pero este tipo de manipulación puede conllevar diferentes consecuencias si se trata de alguna imagen que contenga

números o letras y estos caracteres sean duplicados con intención de manipular algún tipo de identificador, ya sea números de teléfono o de una cuenta de banco en algún tipo de documento, o una matrícula de un coche.

(a) Original (b) Modificada

Figura 3.5: Ejemplo de duplicación por técnica de copia-pega.

Esta técnica también puede ser utilizada con el objetivo de ocultar información o contenido en una imagen digital. En la Figura 3.6 se muestran dos imágenes, al comparar ambas se nota como en la imagen (Figura 3.6.b) se ha ocultado una parte de la información de la imagen original (Figura 3.6.a) [28].

(a) Original (b) Modificada

Figura 3.6: Ejemplo de ocultación de información mediante copia-pega

3.3. Empalme

La técnica de manipulación de imágenes mediante empalme es una de las más empleadas hoy en día, este método consiste en cortar cierto contenido de una imagen y pegarlo en otra. Con intención de mezclar dos imágenes para manipular una escena. Actualmente existe una cantidad abrumadora de imágenes creadas mediante empalme, a pesar de ello, muchos medios de información las intentan vender como contenido original y verdadero, tanto en revistas o televisión, por ejemplo. La gran variedad de aplicaciones de edición de imágenes que existe en esta época facilita de manera exponencial la creación de este tipo de imágenes, ya que, cualquier persona que sepa manejar una aplicación de este estilo, no necesariamente a nivel experto, puede crear contenido de estas características sin ninguna complicación sin que la manipulación sea apreciable. En la Figura 3.7 aparecen 5 ejemplos de imágenes manipuladas mediante empalme. Las imágenes de la Figura3.7.c son las resultantes de la mezcla de las imágenes de las figuras Figura3.7.a y Figura.3.7.b [29].

(a) Imagen 1

(b) Imagen 2

(c) Imagen manipulada

Figura 3.7: Ejemplo de empalme de imágenes obtenidas del Dataset CASIA TIDE V2.0.

3.4. Modificación o Eliminación de la Huella Digital

La técnica de manipulación de imágenes mediante eliminación de huella consiste en modificar el origen de la imagen, es decir, la información sobre el modelo y la marca del dispositivo que generó la imagen.

Los atacantes intentan evitar la identificación de la fuente ya que existe la posibilidad de eliminar y extraer la huella de una imagen. Una técnica común consiste en tratar de añadirle un origen distinto al original tratando de simular que una fotografía ha sido generada por un dispositivo que realmente no lo ha hecho.

Del mismo modo que se puede eliminar ruido de una imagen mediante la corrección de sensibilidad también se puede inyectar el ruido de la imagen de una cámara en otra diferente a través de la corrección de sensibilidad inversa. La detección de adición de ruido también es importante, en [30] se propone un método que a través del cálculo de unas características en bloques de la imagen se detecta el ruido global. Posteriormente mediante la distribución de valores de pixeles en bloques se detecta la adición de ruido.

3.5. Doble Compresión JPG

La mayoría de las contribuciones desarrolladas en el campo de las falsificaciones de imágenes se centran en el formato de datos de imágenes digitales JPEG, ya que es el estándar por defecto para la adquisición e intercambio de imágenes digitales. El éxito del formato JPEG se debe principalmente a la posibilidad de lograr un equilibrio óptimo entre la tasa de compresión de una imagen y su calidad resultante (según un factor de calidad elegido por el usuario).

El factor de calidad (QF) de una imagen JPEG, puede variar en un intervalo [1; 100], donde valores más pequeños resultan en una menor calidad de la

imagen comprimida y un mayor grado de compresión. Por un lado, detectar si una imagen JPEG es alterada o no puede ser más difícil que para otros formatos porque la compresión empleada por esta codificación puede borrar los rastros de falsificación dejados en una foto. Por otro lado, es posible diseñar un algoritmo que detecte características de compresión y los utilice para rastrear posibles falsificaciones. Las características de bloqueo JPEG introducidas por la compresión JPEG pueden considerarse como una "marca de agua" inherente a las imágenes comprimidas. Cuando una imagen JPEG es manipulada, estas características deben ser siempre alteradas por las operaciones de falsificación. Por lo tanto, una imagen obtenida mediante el empalme de una imagen secundaria leída de un archivo JPEG con un factor de calidad sobre otra imagen con un factor de calidad diferente es fácilmente detectable como manipulado.

Muchos algoritmos de detección de imágenes manipuladas funcionan utilizando trazas de características producidas a partir de compresión JPEG. En general, estos algoritmos utilizan algunas de las propiedades estadísticas de los coeficientes de transformada discreta de coseno (DCT) para detectar inconsistencias en las características de bloqueo de una imagen. Una de las primeras contribuciones en esta área se describe en [29]. Los autores estiman la tabla de cuantificación primaria a partir de una imagen JPEG doblemente comprimida usando los histogramas de los coeficientes de DCT individuales. Aquí se utiliza el histograma de los coeficientes de DCT para estimar el tamaño del paso de cuantificación y, a continuación, el algoritmo mide la inconsistencia de los errores de cuantificación entre diferentes regiones de la imagen. Lamentablemente, requiere una intervención humana preliminar para seleccionar una región sospechosa de la imagen a analizar.

4. IDENTIFICACIÓN DE MANIPULACIONES DE IMÁGENES DIGITALES

Como se ha expuesto en el capítulo 3, la facilidad de manipular contenido multimedia hace que imágenes, documentos y archivos sean muy vulnerables. Por tanto, es necesario contar con métodos eficaces para verificar si una imagen ha sido manipulada y validar de esta manera su credibilidad, respecto al contenido de su escena.

El campo de investigación de la integridad de la imagen digital se refiere al problema de evaluar si una imagen digital es el resultado de alguna operación de falsificación. Detectar la falsificación de una imagen JPEG puede ser más difícil que para otros formatos porque los pasos de compresión empleados por esta codificación pueden borrar los trazos de falsificación dejados en una imagen alterada. Las señales introducidas por la compresión JPEG pueden verse como una "marca de agua" inherente para las imágenes comprimidas. Estos artefactos resultan modificados cuando se modifica una imagen JPEG mediante operaciones de falsificación. Estos algoritmos utilizan algunas de las propiedades estadísticas de los coeficientes DCT para detectar inconsistencias en los artefactos de bloqueo de una imagen JPEG objetivo.

Debido a las diferentes técnicas de manipulación de imágenes, se han desarrollado diferentes métodos de detección. Los actuales enfoques forenses de imágenes digitales se clasifican en técnicas activas y pasivas (también conocidas como técnicas ciegas).

Las técnicas activas se basan en determinar la autenticidad de la imagen a través de la huella digital o marca de agua de la imagen, suelen investigar el PRNU de las imágenes, por lo que no tiene nada que ver con la parte gráfica de la imagen. Por el contrario, las técnicas pasivas o ciegas se basan en la parte

visual de la imagen por lo que tratan de encontrar incoherencias en ciertas regiones que determinen la integridad de la imagen analizada. En la Figura 4.1 se muestra una sencilla taxonomía de las técnicas de identificación de manipulaciones de imágenes digitales, dividiéndolas en activas y pasivas.

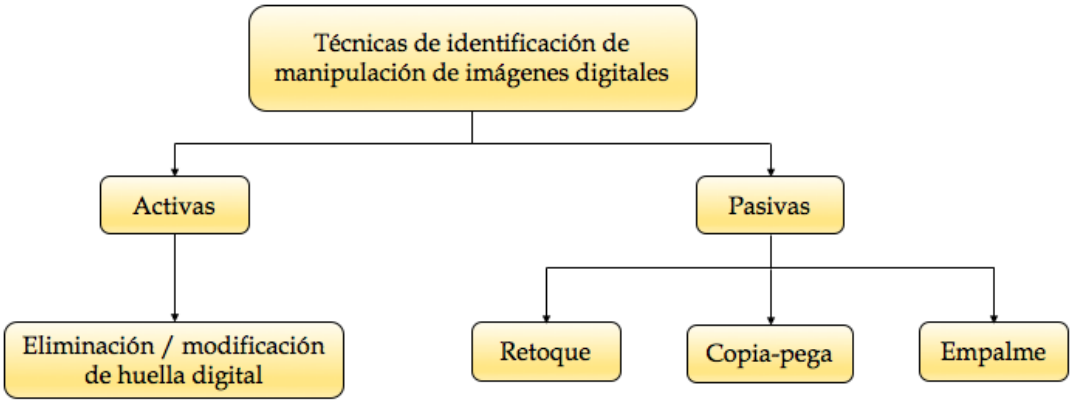


Figura 4.1: Taxonomía de métodos de identificación de edición de imágenes digitales.

Con el paso de los años la cantidad de publicaciones de artículos que contienen técnicas de detección de imágenes de estas características ha ido aumentando proporcionalmente. En la Figura 4.2 se aprecia el crecimiento del número de publicaciones por año entre 2002 y 2014 de artículos basados en técnicas de identificación de imágenes manipuladas según [2].

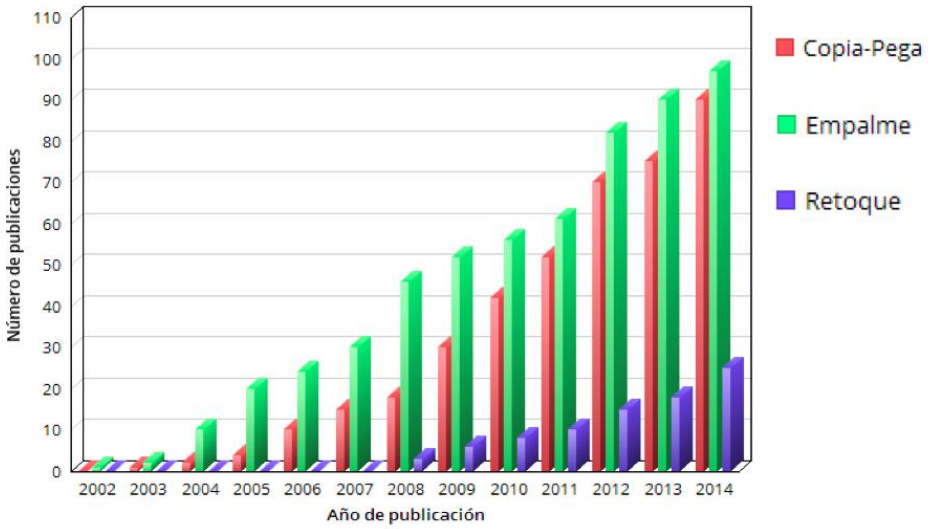


Figura 4.2: Gráfico de publicaciones de los últimos años.

4.1. Identificación de Retoque de Imágenes

Estos métodos se basan en la detección de operaciones de mejora en una imagen como por ejemplo la relación entre los bloques adyacentes o bien modificaciones geométricas y fotométricas. Como se aprecia en la Figura 4.2 las publicaciones de técnicas de identificación para imágenes manipuladas con la técnica de retoque no son de las más abundantes.

En [31] se propuso un algoritmo eficiente para la detección de estas manipulaciones. Este método comienza con la búsqueda de bloques similares para detectar regiones sospechosas y los bloques pertenecientes a áreas adyacentes se filtran utilizando similitud vectorial. Con la adición de la técnica MRR (Multi-region relation) se mejora la localización de la falsificación para eliminar bloques sospechosos pertenecientes a regiones uniformes.

En [32] se describe un algoritmo que devuelve como resultado una métrica con un rango de 1 a 5 para cuantificar las alteraciones de la fotografía digital de una cara humana mediante técnicas digitales de edición. Una puntuación de 1 representa un retoque mínimo mientras que 5 representa un cambio drástico respecto a la original. Se utilizaron diferentes modelos de retoque para cuantificar la mejora perceptual debido a modificaciones fotométricas y geométricas. La modificación geométrica y la fotométrica se cuantifican utilizando varias estadísticas como medias y desviaciones estándar.

4.2. Identificación de Copiar y Pegar Regiones de una Imagen

En [33] se propone un método que consiste en convertir la imagen a escala de grises y dividirla en bloques de tamaño fijo superpuestos. Posteriormente se aplica el HOGM, Histogram Of Orientated Magnitude, a cada bloque para la extracción de características locales y reducir las dimensiones para facilitar la

comparación de las medidas. Finalmente, esas características son clasificadas y las regiones de falsificación se detectan a través de la identificación de pares de bloques similares y las partes falsificadas son detectadas. También se aplica un detector de ruido, para reducir la posibilidad de coincidencias falsas. La técnica propuesta es capaz de detectar de manera precisa regiones duplicadas a pesar de haber sido sometidas a técnicas de post-procesamiento comunes. Este algoritmo alcanza un mejor rendimiento que otros enfoques conocidos. Por lo que proporciona una valiosa contribución en el campo del análisis forense de imágenes digitales.

Un método diferente al mencionado anteriormente se presenta en [28], que se basa primero en la extracción de los descriptores SIFT (*Scale Invariant Feature Transform*), de una imagen, que son invariantes a la rotación, ruido, cambios de iluminación, etc. Posteriormente los descriptores se comparan entre sí para comparar similitudes y así buscar cualquier posible falsificación. Para aumentar la robustez, sólo se aceptan coincidencias si la relación de los vecinos más próximos a los segundos vecinos es menor que un umbral ω. Las primeras etapas del algoritmo SIFT encuentran las coordenadas de los puntos claves en una determinada escala y a cada punto se le asigna una orientación. Los resultados de estas etapas garantizan invariabilidad a localización en la imagen, escala y rotación. Luego se calcula un descriptor para cada punto clave. Este descriptor debe ser altamente distintivo y parcialmente robusto a otro tipo de variaciones como iluminación y perspectiva 3D.

Para crear su descriptor se propone obtener un arreglo de histogramas. Estos histogramas se calculan a partir de los valores de orientación y magnitud en una región de pixeles alrededor del punto de modo que cada histograma se forma de una subregión del histograma mayor. El descriptor consiste en un vector resultado de la concatenación de estos histogramas. Este vector es normalizado con el fin de lograr invariabilidad a cambios de iluminación [34]. Los resultados después de haber aplicado el algoritmo propuesto con tres

umbrales diferentes sobre la Figura 3.6 son los siguientes [28]:

(a) ω = 0.3 (b) ω = 0.45 (b) ω = 0.45

Figura 4.3: Imágenes resultantes de aplicar el algoritmo SIFT.

4.3. Identificación de Empalme de Imágenes

Como se puede observar en la Figura 4.2 las investigaciones relacionadas con técnicas para detectar manipulaciones basadas en empalme de imágenes son las más destacadas en cuanto a cantidad. Esto se debe a que la técnica de editar imágenes uniendo dos o varias imágenes es una de las más utilizadas por excelencia.

La detección de estas manipulaciones requiere el análisis de todo el contenido de la imagen con métodos estadísticos robustos. La mayoría de trabajos de investigación se centran en el desarrollo de algoritmos robustos para detectar operaciones de manipulación, pero, a veces, es más importante detectar que regiones de la imagen han sido manipuladas. Por esta razón, algunos investigadores han adoptado la perspectiva de clasificación binaria.

En las figuras Figura 4.4 y Figura 4.5 se muestra el flujo de trabajo de esta técnica, que se basa en dos fases: fase de entrenamiento en la cual se entrena al sistema a través de imágenes auténticas y manipuladas y las clasifica, y la segunda fase, la fase de pruebas en la que el sistema predice si la imagen es manipulada o auténtica.

En ambas fases se ejecuta un algoritmo con cada imagen para obtener una serie de características específicas de la imagen y se tratan. No todos los

métodos de detección de empalme de imágenes realizan las primeras tres fases iguales, el pre-procesamiento, y la extracción y selección de características difiere dependiendo de cada algoritmo, pero se suele trabajar siempre con bloques o regiones de la imagen.

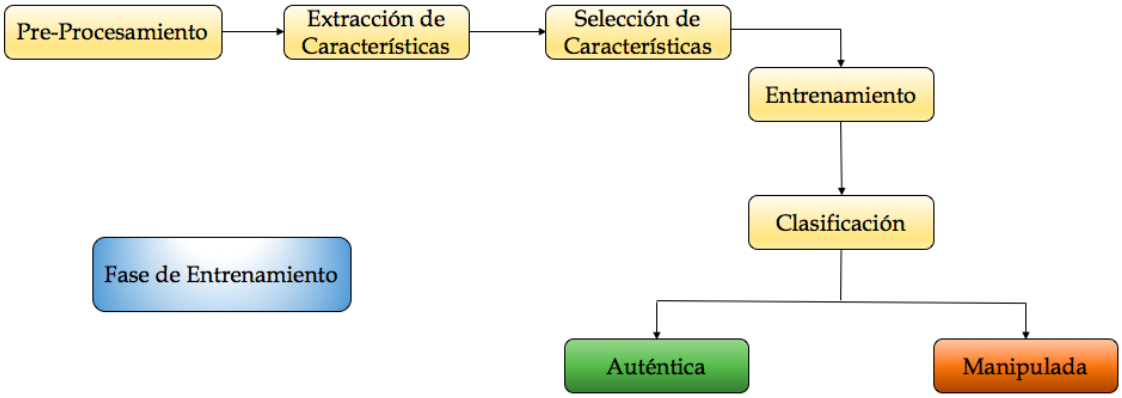


Figura 4.4: Ejemplo de fase de entrenamiento.

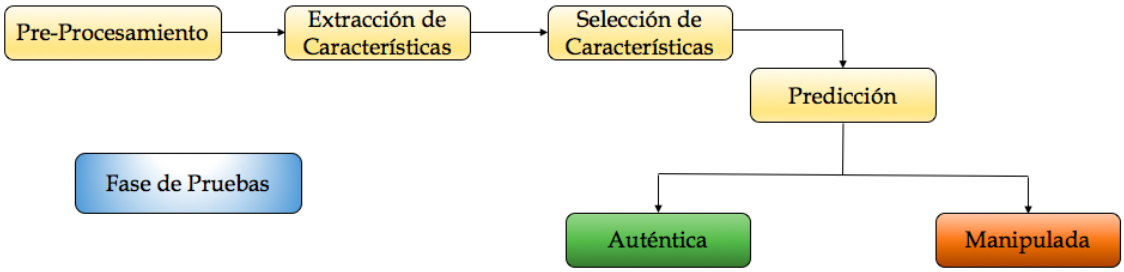


Figura 4.5: Ejemplo de fase de predicción.

Las características más tratadas, entre otras, en la mayoría de las investigaciones suelen ser: las transformadas de wavelet [18] o las transformadas discretas del coseno (DCT) [35], imágenes transformadas a escala de grises [36], uso de la técnica LBP [37] y características de Markov [38]. En los últimos años se ha superado, sin mucha dificultad, el 90% de acierto en las predicciones a través de la SVM llegando como máximo al 99.14% con [39]. La mayoría de estos trabajos utilizan los mismos datasets con intención de poder comparar los porcentajes de manera real respecto al resto de algoritmos. Los datasets más utilizados son Columbia y Casia v1.0 y v2.0. Estas cifras

corroboran que, a pesar, de las técnicas de post-procesamiento, siempre permanece una incoherencia entre los píxeles de una región que ha sido manipulada, permitiendo así la detección de la edición realizada en la imagen.

4.4. Identificación de Modificación o Eliminación de la Huella Digital

La autenticidad de una imagen puede ser verificada a través de las características del ruido del sensor que genera la imagen. Concretamente, la falta de uniformidad en el ruido se utiliza para la identificación del dispositivo origen. Algunas de las técnicas que tratan esta característica de identificación han comenzado a utilizarse más recientemente como evidencia en un tribunal de justicia, con objetivos de ser más útiles e influentes con el paso del tiempo. Hasta la fecha, el método de obtención de la fuente más eficaz se basa en el patrón de ruido de no-uniformidad, también conocido como ruido PRNU. El ruido PRNU genera un patrón único que está presente en cada imagen capturado por el sensor de formación de imágenes, por lo que actúa como huella digital del dispositivo. Esta huella se puede obtener a través de unas técnicas de "denoising" las cuales extraen el ruido. Este tipo de técnicas se basan en el estudio de las huellas digitales que dejan los dispositivos al generar una imagen, por lo que se basa en el rastro que los defectos del sensor pueden dejar sobre las imágenes.

La autenticidad de una imagen digital se puede verificar utilizando las características de ruido del sensor de la imagen [40], los defectos físicos [41] e incluso las distorsiones [42].

Ciertos estudios muestran que la identificación de la cámara a través del PRNU es bastante robusta a las manipulaciones de imágenes como la compresión JPEG, recorte, impresión [43], y la reducción de tamaño.

Sin embargo, un patrón PRNU puede ser transferido de una imagen a otra para uso malicioso en un ámbito judicial. Aunque hay algunos métodos forenses para detectar esa transferencia de ruido.

Por ejemplo, en [44] se desarrollan unos métodos para revelar actividades contra-forenses en las que un atacante estima la huella digital de una cámara a través de un conjunto de imágenes y lo pega en una imagen de una cámara diferente con la intención de introducir información falsa.

En [45] se propone un método fiable para detectar tales huellas digitales y detectar la adición de ruido.

En [8] investigan la robustez de la huella digital de PRNU contra las operaciones de procesamiento comunes, como el "denoising", la re-compresión y el desmontaje fuera de cámara. Sus resultados muestran que incluso después de ocho rondas de eliminación de ruido todavía hay una correlación significativa entre el patrón de ruido de la imagen y la huella digital PRNU de la cámara. De manera similar, se muestra que la re-compresión no es capaz de eliminar el ruido PRNU dentro de límites tolerables de pérdida de calidad de imagen.

4.5. Detección de Doble Compresión JPEG

En [46] Farid, propuso una técnica basada en la detección de fantasmas JPEG, para establecer si una región de una imagen fue originalmente comprimida con un factor de calidad diferente de otras regiones de la misma imagen. La desventaja de esta técnica es que solo funciona cuando la región alterada tiene una calidad inferior a la imagen circundante.

Jen-Chun Lee en [33] expone el hecho de que si la mayoría de las imágenes manipuladas se almacena en formato JPEG significa que el algoritmo de detección de falsificaciones de imágenes propuesto debe ser capaz de hacer

frente a las falsificaciones comunes asociadas a este tipo de compresión.

Para ello emplearon dos conjuntos (CoMoFoD, Manipulación de imágenes) de datos para evaluar la robustez del algoritmo propuesto a la compresión JPEG con distintos factores de calidad (entre 20 y 100). La resolución de la imagen no se alteró cuando se guardaron las imágenes. Para el primer conjunto de datos se obtuvieron 360 imágenes falsificadas y para el segundo 432. Obviamente, se realizaron experimentos para evaluar la efectividad del algoritmo propuesto para resistir los efectos de la compresión JPEG. Los resultados mostraron que el método propuesto funciona independientemente del factor de calidad aplicado durante la compresión e incluso cuando las imágenes manipuladas fueron comprimidas empleando un factor de calidad de 60, el valor CDR superó 0,9. En la mayoría de los casos, los FDR se mantuvieron por debajo de 0,1, lo que indica que los resultados son prometedores. Asimismo, el método propuesto es capaz de detectar la compresión JPEG con un factor de calidad superior a 40, resultando en valores CDR y FDR aceptables.

En [47] comentan que Zuo et al. en [48] propuso un método de detección de imágenes compuestas sobre la base de los rastros de re-muestreo y compresión JPEG. El algoritmo implica dividir una imagen en bloques superpuestos y a continuación se define y evalúa un factor de medida de los bloques. Dicho factor contiene las características de re-muestreo y de compresión JPEG de cada uno de los bloques superpuestos. Finalmente, el factor de medida de los bloques es aplicado para poder detectar las regiones alteradas. A diferencia de otros métodos propuestos en el artículo para la detección de falsificaciones de imágenes JPEG, cuando el factor de calidad de doble compresión es menor que el factor de calidad primario, el método propuesto en [48] puede funcionar bien. Sin embargo, basándose en las sugerencias de los autores en [48] deja de ser eficaz cuando la imagen está compuesta de dos imágenes sin comprimir y las regiones manipuladas no se someten a operaciones geométricas. No obstante, en [47], se construyó un dataset de imágenes compuesto por 50 imágenes con

doble compresión JPEG. En este conjunto de datos, hay una imagen destino comprimida en JPEG con factor de calidad QF1, y la imagen compuesta comprimida con factor de calidad QF2, donde QF1 = [50,60,70,80,90] y QF2 = [50,60,70,80]. Los resultados experimentales mostraron que la detección de manipulación propuesta tiene un buen desempeño en la factibilidad de la compresión JPEG. No obstante, se utilizaron dos métodos de última generación, el método de detección de manipulación basado en la compresión JPEG [49] y el método de detección de alteraciones basadas en el mapa de probabilidad posterior en bloque (BPPM) [50] para comparar con el método propuesto para evaluar el rendimiento en JPEG de datos de compresión. Los resultados experimentales muestran que las regiones manipuladas no se detectan con precisión. Además, la identificación de las imágenes manipuladas debe ser decidida manualmente por el usuario porque la regla de identificación no fue dada.

El método de detección de manipulación basado en BPPM, Mapa de Probabilidad Posterior en Bloque, [51] hace que la detección de manipulación no sea correcta para todas las imágenes manipuladas debido a que la región alterada no está dentro del 30% -70% de la imagen compuesta en el dataset, por lo que la capacidad de detección del método de detección de manipulación basado en BPPM es ineficaz para el dataset de compresión JPEG.

Para finalizar, en comparación con los métodos de detección de manipulación basada en la compresión JPEG [48] [49] [51], los resultados experimentales muestran que la propuesta de [47] puede funcionar bien sin la limitación de los factores de calidad en la compresión JPEG para las imágenes objetivo y compuestas. Los resultados experimentales muestran que el método propuesto supera los métodos de [49] [51].

5. CONTRIBUCIÓN

En este trabajo se ha desarrollado una técnica de detección de imágenes digitales falsificadas basada en los patrones locales de textura. La técnica combina el patrón binario local (LBP, el inglés *Local Binary Pattern*) con la transformada wavelet y la transformada discreta del coseno para extraer las características. Posteriormente, las características son procesadas por una máquina de soporte vectorial para clasificar las imágenes falsificadas de las auténticas.

5.1. Consideraciones Generales

El método propuesto trabaja sobre un enfoque híbrido usando las técnicas DWT, DCT y LBP. A continuación, se presenta una breve descripción de las técnicas que hacen parte del método propuesto para la detección de la falsificación de imágenes digitales.

5.1.1. Patrones Binarios Locales

LBP es una poderosa característica para la clasificación de texturas. Consiste en asignar un código a cada píxel de la imagen. Se calcula utilizando la ecuación (1).

$$LBP_{p,r} = \sum_{i=1}^{p-1} S(p_i - p_c)2^i \tag{1}$$

$$S(p_i - p_c) = \begin{cases} 1, p_i - p_c \geq 0 \\ 0, p_i - p_c < 0 \end{cases} \tag{2}$$

donde,

p es el número de píxeles vecinos, r es el radio de vecindad y p_c es el valor del píxel central. Este operador calcula el código LBP mediante el valor de 8

vecinos. Si el valor del píxel del vecino es menor que el del centro, se le asigna el dígito binario 0, de lo contrario un 1. Después los dígitos binarios del vecino se juntan para construir un código binario. El código LPB será el valor decimal de ese código binario.

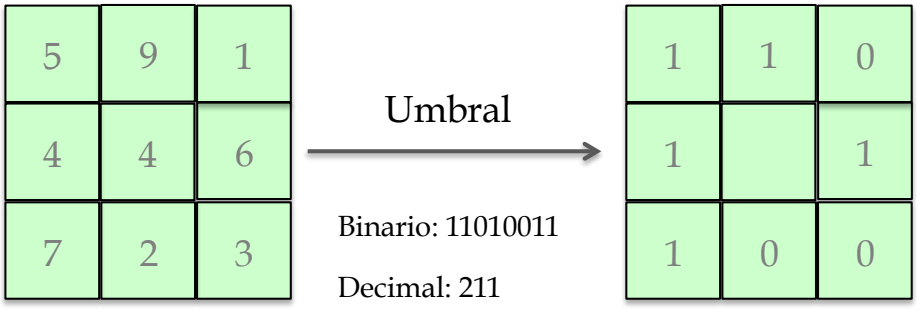
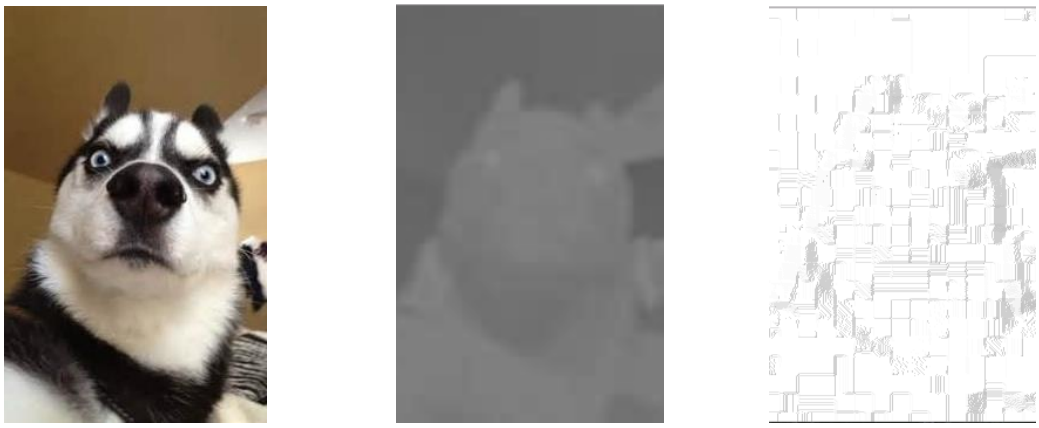


Figura 5.1: Ejemplo conversión a LBP.

Las características más importantes de LBP son el alto poder discriminante, la tolerancia frente a cambios de iluminación y su simplicidad computacional. Es por esto que se ha convertido en una de las soluciones más empleadas en la detección de texturas. Al método para calcular el LBP se le han incorporado circulantes vecinos circulares, donde dichos vecinos se ubican a la misma distancia, consiguiendo una interpolación binaria en caso de que los vecinos no coincidan con el centro de un píxel de la imagen. La Figura 5.1 muestra el resultado de aplicar LBP y Crominancia roja sobre una imagen.



(a)Original (b) Crominancia roja (c) LBP

Figura 5.1: Ejemplo de aplicación del Patrón binario local y crominancia roja sobre una imagen.

En [52] se empleó LBP para realizar segmentación no supervisada basada en características de textura de la imagen. Para ello se divide la imagen en un número de regiones escogido por el usuario. A continuación, se realiza una fusión de aquellas texturas contiguas que tengan un descriptor similar dejando para el final el refinamiento de las regiones, eliminando las aristas producidas en el paso de la división jerárquica, obteniendo una segmentación más ajustada y precisa.

En [53] se realiza un estudio para evaluar distintas técnicas basadas en LBP para el reconocimiento de rostros demostrando, que si se divide el rostro en subregiones para describirlas usando LBP, se incrementa notablemente la precisión de los sistemas de clasificación frente a la descripción del rostro por completo.

En [54] se propone detectar los ojos de las personas estimando la posición de la cara asumiendo que el ojo es una región de la imagen que tiene una alta variabilidad.

En [55] se presenta una técnica basada en la construcción de un grafo adaptativo que utiliza está LBP para la descripción de las regiones. La investigación se basa en que el descriptor de una imagen está formado por una combinación de los descriptores restantes de la base datos a los que se les asigna unos pesos concretos. Así, si una imagen contiene un objeto concreto, el peso asociado a ese objeto será alto, lo cual permite su detección.

5.1.2. Transformada Wavelet

Una función wavelet es una pequeña onda cuya energía se concentra en el tiempo y sirve como herramienta para analizar en términos de tiempo y frecuencia fenómenos estacionarios, no estacionarios y variables en tiempo. Por esta razón, las transformadas wavelets han comenzado a jugar un papel más significativo en la compresión de imágenes. Actualmente, son consideradas

como la herramienta de representación de señal más potente aplicada en el procesamiento de señales e imágenes, la compresión de datos, la detección de características en las imágenes, la eliminación de ruido, entre otras técnicas.

La transformada wavelet (WT, del inglés Wavelet Transform), generan bloques de información en escala y tiempo de una señal. El proceso de transformación wavelet de una señal recibe el nombre de análisis, y el proceso inverso para reconstruir una señal recibe el nombre de síntesis.

Actualmente hay distintas familias de wavelets, para las cuales aún no se ha definido un criterio que evalúe su calidad puesto que depende en gran parte de la aplicación y características que se requieran. Las familias más reconocidas son: Haar, Daubechies, Coiflets, Symlets, Biortogonales, Meyer, Mexican hat, Shannon y Morlet.

En el método de identificación de imágenes manipuladas se emplea la DWT en la reducción del ruido de imágenes. Esta técnica permite reducir la magnitud de cada coeficiente wavelet a un valor concreto en función del nivel del ruido que se estime de la señal. Uno de los métodos más empleados para la reducción del ruido es la designación de un umbral (λ). Una vez definido el umbral, los valores que quedan por debajo se eliminan (umbral duro o "hard-threshold") o se reducen en magnitud (umbral suave o "soft-threshold").

En primer lugar, para reducir el ruido se calcula la DWT de la señal. Después, se selecciona un umbral para poder reducir los coeficientes de menor energía. Se pueden diferenciar dos métodos para la reducción del ruido por selección de umbral: umbrales globales que son aquellos que definen un valor fijo para ser aplicados a todos los coeficientes wavelet, mientras que los umbrales recursivos escogen un valor distinto para cada resolución.

5.1.3. Transformada Discreta del Coseno

La transformada discreta del coseno (en inglés *Discrete Cosine Transform)* es una variación de la transformada de Fourier discreta, donde la imagen se descompone en sumas de cosenos, pero utilizando números reales únicamente. Sólo actúa sobre funciones periódicas con simetría par y el resultado es una secuencia de números reales.

Forma una secuencia finita de puntos como resultado del sumatorio de varias señales con distintas frecuencias y amplitudes. Se suele usar para representar este registro empleando las componentes espectrales más representativas de tal forma que la señal reconstruida aún tenga semejanza con la señal original.

También se puede entender multidimensionalmente, de manera equivalente para una matriz de N x N calculada por cada fila y columna.

La DCT de una secuencia u [m, n], 0 <= m, n <= N-1 se define como:

$$v[k,l] = \alpha(k)\alpha(l) \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} u[m,n] \cos\left[\frac{(2m+1)k\pi}{2N}\right] \cos\left[\frac{(2n+1)l\pi}{2N}\right]$$

Donde el coeficiente α(ψ) (ψ = k, l) toma los valores:

$$\alpha(\psi) = \begin{cases} \sqrt{\frac{1}{N}} & \psi = 0 \\ \sqrt{\frac{2}{N}} & \psi = 1, \ldots, N-1 \end{cases}$$

Tal y cómo se muestra es una transformación lineal, de núcleo de transformación separable.

Esta transformada cuenta con algunas variantes, las más usadas son la DCT-I

$$f_j = \frac{1}{2}(x_0 + (-1)^j x_{n-1}) + \sum_{k=1}^{n-2} x_k \cos\left[\frac{\pi}{n-1} kj\right]$$

y la DCT-II.

$$f_j = \sum_{k=0}^{n-1} x_k \cos\left[\frac{\pi}{n} j \left(k + \frac{1}{2}\right)\right]$$

La DCT-III se conoce popularmente como la IDCT, es decir, la transformada inversa discreta del coseno.

Esta operación se utiliza de manera muy habitual en el ámbito de la compresión debido a sus características, de hecho, está presente en varios algoritmos cómo por ejemplo en la compresión JPEG.

La DCT consigue concentrar la mayor parte de la información en pocos coeficientes transformados tal y como muestra la figura 5.2, en la que aparece una comparación de la concentración de potencia de una DCT-II bidimensional con la concentración de potencia de una DFT también bidimensional.

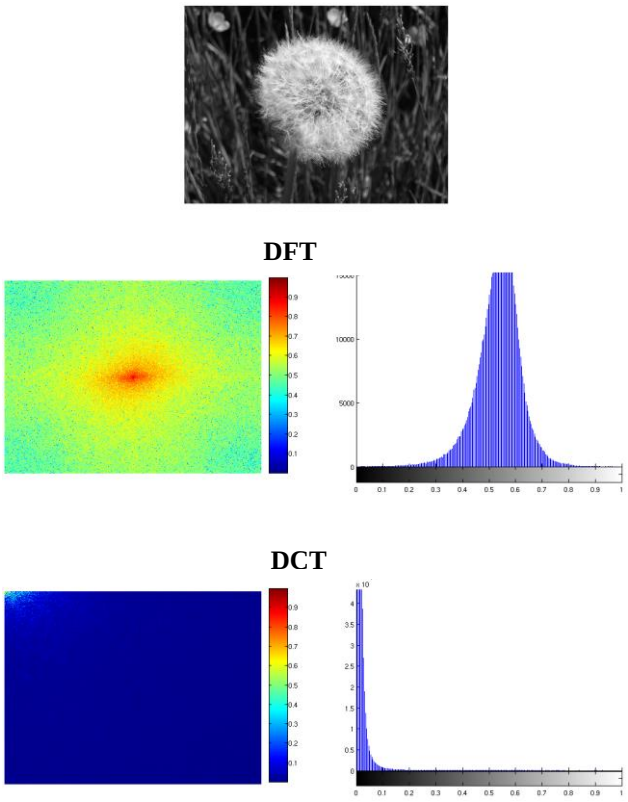


Figura 5.2: Comparación DFT y DCT.

La transformada no depende de los datos, es decir, el algoritmo no varía según los datos que recibe, tal y como sucede en otros algoritmos de compresión. Además, produce pocos errores en los límites de los bloques de la imagen. La minimización de los errores permite reducir el efecto de bloque en las imágenes reconstruidas. La capacidad de interpretar los coeficientes desde el ámbito de la frecuencia permite aprovechar al máximo la capacidad de compresión.

5.1.4. Histograma

Un histograma es una representación gráfica de una variable en forma de barras, donde la superficie de cada barra es proporcional a la frecuencia de los valores representados. Los histogramas se utilizan para variables continuas o para variables discretas, con un gran número de datos.

Sirven para obtener una vista general de la distribución de la muestra, respecto a una característica. De esta manera ofrece una visión de grupo permitiendo observar tendencias hacia una determinada región de valores dentro del espectro de valores posibles. Gracias a esto, se pueden determinar comportamientos, y así, observar el grado de homogeneidad, el grado de variabilidad, y la dispersión de valores.

5.2. Funcionamiento

Para poder detectar si una imagen digital ha sido modificada es necesario extraer las características de cada una de las regiones de la imagen. Los algoritmos de extracción de las características combinan 3 técnicas: Uso de la transformada wavelet y de la transformada discreta del coseno en conjunción con el patrón binario local presente en cada uno de los bloques de la imagen. Este algoritmo está basado en las principales ideas presentadas en [56].

5.2.1. Extracción de Características Wavelets de cada Bloque LBP

El primer paso es convertir la imagen a escala de grises YCbCr para obtener las crominancias "Cb" y "Cr", obteniendo dos matrices del mismo tamaño que la imagen con las características de las crominancias roja y azul.

A continuación, para cada matriz se hace una subdivisión en bloques de 8x8 píxeles, no superpuestos. Cabe mencionar que se pueden utilizar bloques de mayor tamaño múltiplos de ocho (16x16 o de 32x32) para reducir el tiempo de ejecución. Sin embargo, entre más pequeños son los bloques, más eficiente es el algoritmo. Si la resolución de la imagen es múltiplo de ocho se descartan los pixeles de los extremos para tomar siempre bloques de 8x8.

El tercer paso es extraer el patrón binario local a cada bloque de la imagen.

Posteriormente, se aplica la transformada de Wavelet (DWT), obteniendo así las bandas vertical, horizontal y diagonal. Se aplica a cada banda el filtro espejo en cuadratura (QMF) y se calcula la media aritmética de cada una de ellas.

Se genera un vector con un total de 6 características por cada bloque: 1 característica x 3 bandas x dos matrices de crominancia. El vector de característica es ordenado primero por las banda horizontal, vertical y diagonal y luego por la crominancia, primero la roja y luego la azul. Los vectores generados se concatenan de manera ordenada tal y como se recorren los bloques a través de la imagen. El número total de características que se obtienen de cada imagen depende de la cantidad de bloques obtenidos, por lo que forman un total de 6 características x el número de bloques. En la Figura 5.1 se muestra un esquema conceptual del algoritmo descrito.

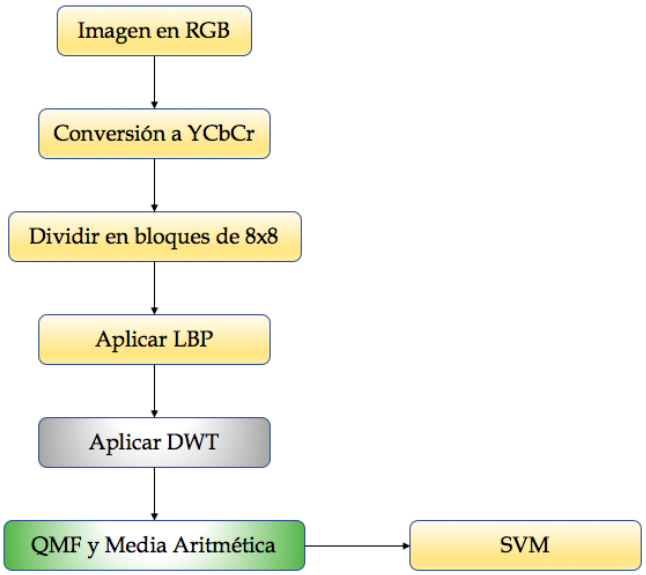


Figura 5.1: Descripción conceptual del algoritmo 1.

5.2.2. Extracción de Características de la Transformada Discreta del Coseno de cada Bloque LBP

Para extracción de las características que se desprenden de la transformada discreta del coseno se siguen los 3 primeros pasos de la sección 5.2.1

A continuación, se aplica la transformada discreta del coseno a cada bloque LBP y se calcula la desviación estándar de los valores DCT obtenidos.

Por cada bloque se genera un vector con un total de 2 características: 1 característica x dos matrices de crominancia. Los vectores generados se concatenan de manera ordenada tal y como se recorren los bloques a través de la imagen. El número total de características que se obtienen de cada imagen depende de la cantidad de bloques obtenidos, por lo que forman un total de 2 características x el número de bloques. En la Figura 5.2 se muestra un esquema conceptual del algoritmo descrito.

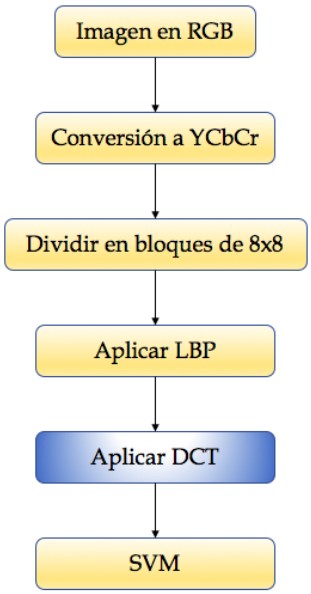


Figura 5.2: Descripción conceptual del algoritmo 2.

5.2.3. Extracción de Características del Histograma de cada Bloque LBP

En la Figura 5.3 se muestra un esquema conceptual del proceso de extracción de características del histograma de cada bloque.

El primer paso es convertir la imagen a escala de grises YCbCr para obtener la crominancia "Cr", obteniendo una matriz del mismo tamaño que la imagen con las características de las crominancias roja.

Luego, a la matriz resultante se le aplica la transformada wavelet para obtener 4 bandas (LL, LH, HL y HH). Posteriormente se dividen las matrices en bloques iguales de 8x8 píxeles, no superpuestos.

A continuación, para cada matriz se hace una subdivisión en bloques de 8x8 píxeles, no superpuestos.

El cuarto paso es extraer el patrón binario local a cada bloque de la imagen.

Finalmente, se calcula el histograma a cada banda de cada bloque. Con los

cuatro histogramas por bloque se forma un vector. Por cada bloque se genera un vector con un total de 4 características: 1 característica x cada banda (LL, LH, HL y HH). El número total de características que se obtienen de cada imagen depende de la cantidad de bloques obtenidos, por lo que forman un total de 4 características x el número de bloques.

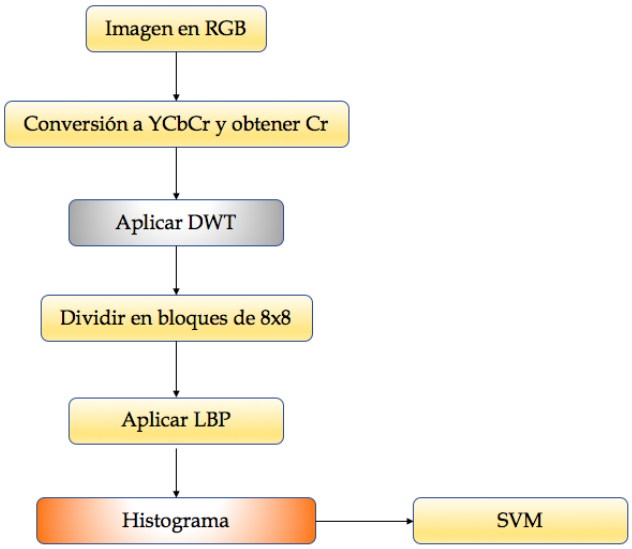


Figura 5.3: Descripción conceptual del algoritmo 4.

5.3. Evaluación

La efectividad del método de autenticación de imágenes propuesto se ha evaluado utilizando 3 datasets de imágenes con diversas modificaciones: CASIA v1.0 [57], CASIA v2.0 [57] y un dataset personal. La Tabla 5.1 se presentan las principales características de los datasets utilizados.

Tabla 5.1. Características de los datasets utilizados en la evaluación.

Dataset	Formato	Resolución	Número de Imágenes		
			Originales	Modificadas	Total
CASIA TIDE v1.0	JPEG	374×256	800	921	1725
CASIA TIDE v2.0	JPEG, BMP, TIFF	240×160 to 900×600	7491	5123	12614
Propio	JPEG	3264×2448 a 4160×3120	100	100	200

El dataset denominado “propio” fue creado siguiendo las siguientes pautas:

1. Se seleccionaron 5 dispositivos móviles de diferentes fabricantes. Con cada dispositivo se tomaron 10 fotografías.

2. Se utilizaron las 5 herramientas de edición de imágenes digitales consideradas como las más populares en Internet para realizar modificaciones sobre cada una de las fotografías del dataset: DarkRoom, Layout, Picsart, Que_Doodle, VSCOcam. Con cada herramienta se realizaron las siguientes modificaciones:

    - Copiar y pegar regiones de una imagen en otra de la misma imagen. A las regiones cortadas se les aplicó operaciones de recorte, modificación del tamaño, rotación y otras distorsiones para posteriormente pegarlas en otra región de la imagen.

    - Filtros de retoque propios de cada una de las herramientas utilizadas, como: Filtro granulado con valor +12, filtro reflejo, filtro “Galaxy” (que modifica las tonalidades y texturas de las imágenes).

Una vez extraídas las características con los algoritmos diseñados se utilizó LIBSVM [59] como implementación de la máquina de soporte vectorial para la clasificación. Los valores óptimos para los parámetros del núcleo (γ and C) se calculan automáticamente mediante un proceso de correlación cruzada. La tasa de acierto resultante de los experimentos se calcula promediando la tasa de acierto obtenida en las 10 ejecuciones de cada experimento con diferentes muestras de cada dataset.

El primer experimento evaluó la efectividad del método propuesto en este trabajo con un subconjunto de imágenes auténticas y modificadas de los datasets de la Tabla 5.1 con las características extraídas con cada uno de los algoritmos presentados en la sección 5.2:

- LBP-YCbCr-WT: Características wavelets extraídas de cada bloque LBP.
- LBP-YCbCr-DCT: Características DCT extraídas de cada bloque LBP
- LBP-YCbCr-Hist: Características del histograma de cada bloque LBP
- Todas: Combinación de las características LBP-YCbCr-WT + LBP-YCbCr-DCT + LBP-YCbCr-Hist.

En el primer experimento se analizó los resultados obtenidos con las características de cada uno de los algoritmos de extracción evaluadas de forma independiente. Los resultados de estos experimentos se muestran en la Tabla 5.2. Como se puede observar, la tasa de acierto es mejor en los casos donde se utiliza el método de autenticación que combina las características de todos los algoritmos. Asimismo, se observa que la mayor tasa de acierto se obtiene con el dataset "propio" alcanzando un 95.74%. Este resultado puede verse afectado por la resolución de las imágenes del dataset. En los datasets CASIA v1.0 la resolución es de 374×256, mientras que el resto de datasets tiene resoluciones superiores a 512×512.

Tabla 5.2. Experimento comparativo con diferentes datasets.

Dataset	LBP-YCbCr-WT	LBP-YCbCr-DCT	LBP-YCbCr-Hist	Todas
CASIA v1.0	50,50%	50,00%	50,00%	51,50%
CASIA v2.0	55,00%	61,00%	55,00%	96,50%
Propio	76,90%	85,66%	87,32%	95,74%

El siguiente experimento se realizó con el dataset "Propio" para analizar las trazas que deja la aplicación con la que se realiza la modificación. Los resultados de este experimento se presentan en la Tabla 5.3.

Tabla 5.2. Experimento comparativo con diferentes datasets.

Dataset	LBP-YCbCr-WT	LBP-YCbCr-DCT	LBP-YCbCr-Hist	Todas
LAYOUT	100,00%	100,00%	100,00%	100,00%
APM	56,41%	58,97%	56,41%	76,92%
PicsArt	100,00%	100,00%	100,00%	100,00%
Doodle	72,50%	80,00%	87,50%	100,00%
DarkRoom	75,00%	85,00%	92,50%	97,50%

6. CONCLUSIONES Y TRABAJO FUTURO

6.1. Conclusiones

Hoy en día es complicado poder demostrar que una imagen es auténtica, ya que existen infinidad de aplicaciones y métodos para modificarlas tanto con buenas intenciones como con malas. Además del buen acabado que tienen la mayoría de este tipo de imágenes en las que no es posible percibir para el ojo humano que la imagen ha sido modificada.

Sin embargo, las técnicas de falsificación de imágenes digitales están teniendo un crecimiento exponencial en los últimos años, con intención de asegurar la integridad y autenticidad de las imágenes digitales hoy en día.

Con la presencia de internet y todos los dispositivos que están apareciendo en el mercado, surge la necesidad de desarrollar técnicas para poder identificar dichas falsificaciones y detener este problema.

En este trabajo se ha desarrollado un método de detección de modificaciones mediante los patrones locales de texturas. En el método propuesto se extraen las características de cada bloque LBP a través de tres algoritmos: Características wavelets, características de la transformada discreta del coseno y características del histograma. Para poder llevar a cabo la identificación de las manipulaciones se han llevado a cabo una serie de experimentos con diferentes dataset: 3 dataset públicos (CASIA v1.0, CASIA v2.0) y un dataset propio.

Los mejores resultados se obtienen cuando se combinan todas las características extraídas. Los algoritmos presentados como regla principal emplean el método del LBP (Local Binary Pattern) en todos ellos, incluyendo variantes en cada uno.

6.2. Trabajo Futuro

Como trabajo futuro pueden señalarse las siguientes líneas de investigación:

- Evaluar el uso de otros métodos de identificación de manipulaciones.
- Combinar partes de los algoritmos desarrollados para tratar de conseguir mejores resultados.
- Analizar resultados con distintos tipos de Datasets.
- Ampliar los algoritmos de identificación de manipulación a vídeos.

RESUMEN EN INGLÉS

7. INTRODUCTION

7.1. Motivation

Nowadays, digital images are present everywhere, in the media, on the Internet, etc. Even though not everyone is aware of it, most of these images have been manipulated, only just to acquire better results. However, there are certain types of images that due to their importance must guarantee their authenticity and integrity. Due to the ease of manipulating digital images many times it can't be proved that the image being viewed is real or tampered. In the last few years, certain sectors of politics, media, and even science have been affected by the falsification of images. That's why images which are of great importance must be a guarantee of integrity and authenticity, in order to be able to take them as true. An example of this is the controversy generated by an image that showed the silhouette of a man on the planet Mars (see Figure 7.1).

Figure 7.1: Photo taken by NASA that shows a human silhouette on Mars.

The photo was captured by a NASA probe, this image was in the focus of the news in 2008, but the authenticity of this image still in doubt today [1]. The exponential growth in the last few years of manipulating images with astonishing ease allows anyone to manipulate an image without the need of being an expert and without a trace on naked eye. On the other hand, the tools with intention of detecting tampered images have shone by its absence [1].

Image tampering was born with the photograph itself. In the early years photography became a method to have a portrait, so first period photographers learned that manipulating certain photographs increased their sales. Some examples of forgery and manipulation of images in important situations in history.

Figure 7.2 shows a clear example of manipulation of photographs from the beginning of the photograph itself. Figure 1.2.a shows the photograph of Lenin and Trotsky from the Soviet Union, while Figure 1.2.b shows the same manipulated image, where Trotsky and another person were removed for political reasons, as Trotsky, being a character not grateful for the political life of that country was removed from many of the images. The falsification of images also affects the political sector, and that's why each shows what interests him manipulating the photographs taken to achieve their goals despite of showing information that is not authentic.

(a) Original (b) Tampered

Figure 1.2: Lenin and Trotsky

Also, Figure 7.3 shows how people who appear behind the man in the foreground in the original image (Figure 1.3.a) have been removed from the scene in Figure 1.3.b.

(a) Original (b) Tampered

Figure 7.3: Example of image manipulation with the purpose of hiding content.

Another case is presented in one of the most striking images of 2005, in which we can appreciate the technique of composition of images from two different images. The photograph shown in Figure 7.4 was taken in an English Army maneuver on the African coasts, but later it was discovered that it was a montage with the image where the shark appears. The image in Figure 7.4.a is manipulated, as it is actually a mixture of images in Figures 7.4.b and 7.4.c.

(a) Tampered (b) Image (1) (c) Image (2)

Figur3 7.4: Example of image splicing.

Examples show the difficulty of identifying manipulated or falsified images, considering that, in many cases the result has a very high quality that makes it practically impossible for human eye to identify the manipulation.

With the advent of some powerful image editing tools, manipulating them and changing their contents has become a trivial fact, because you can add, change or delete meaningful information from an image without leaving a trace of the alterations. This has made it a priority to develop methods for identifying counterfeiting operations and validating the credibility of digital images. Also, the exponential growth experienced by technology and powerful image manipulation algorithms, including software such as Photoshop, Corel Draw and others, is becoming complicated the distinction between an authentic image and its modified version [2].

Image manipulation using computer techniques is gaining much popularity and acceptance in areas such as forensic research, information technology, intelligence services, medical scanners, journalism, special effects and movies. For all these reasons forensic analysis of digital images of mobile devices is booming. The study must be concrete for this type of devices, since they have specific characteristics that allow better results.

7.2. Project goal

The objectives of the project are listed below:

- To carry out a study of previous works and investigations existing so far related to the types of modifications existing in the images to find out a classification of the most relevant techniques.
- Analyze forensic analysis techniques for existing digital image authentication in order to detail and understand the most relevant ones.
- Implement an algorithm to determine if a digital image has been manipulated by analyzing the image content.

7.3. Work Planification

The project development planning was as follows: First, weekly meetings were held with the tutors to define the objectives and range of the End-of-Grade Work. Once the work to be carried out was defined, the activities of study of the literature related to the investigation began to have a vision of the current state of it. Afterwards, activities related to design and subsequent implementation of the algorithm defined in the previous phase were carried out. At the same time, activities were carried out to monitor and control the progress of the project, reviewing each of the agreed activities. The main tasks performed in this stage are: Specification of the requirements, design and implementation of the algorithm and lastly the tests of the algorithm with several databases of public photographs and generated during the development of the project. Finally, in parallel to the stages described, all the documentation required for the End of Grade Work was generated.

8. CONCLUSIONS AND FUTURE WORK

8.1. Conclusions

Nowadays it's doubtful to be able to prove that an image is authentic, since there are infinite of applications and methods to modify them with good intentions as well as with bad ones. In addition, this type of images obtain good results in which it is not possible to perceive on naked eye that the image has been modified.

However, digital image forgery techniques are having an exponential growth in the last few years, with the intention of ensuring the integrity and authenticity of digital images today.

Due to the presence of the internet and all devices that are appearing on the market, the need arises to develop techniques to identify such counterfeits and solve this problem.

In this work four algorithms of detection of manipulation of digital images by splicing have been developed. In order to carry out the identification of the manipulations, a series of experiments with different datasets have been done.

In short, the best way to deal with the detection of manipulations by splicing is to treat the image by dividing it into blocks to try to locate the inconsistencies between them. This has been the way in which the proposed algorithms have been developed as well as those of the investigated articles.

The evaluation of the proposal is presented with photos of the dataset mentioned in the previous paragraph. With regard to, algorithms presented, as main rule use the LBP (Local Binary Pattern) method in all of them, including variants in each of the experiments.

8.2. Future Work

As future work may be noted the following investigation items:

- Assess the use of other kind of methods of identifying manipulations beyond splice.
- Combine fractions of the algorithms developed to try to achieve better results.
- Improve algorithms proposed or implement new ones with the aim of detecting more than image splicing modifications.
- Analyze results with different types of Dataset.
- Expand the manipulation identification algorithms to videos.

BIBLIOGRAFÍA

[1] M. A. Rosales García, «Análisis Forense en Imágenes Digitales», jun. 2009.

[2] M. A. Qureshi y M. Deriche, «A bibliography of Pixel-Based Blind Image Forgery Detection Techniques», *Signal Process. Image Commun.*, vol. 39, Part A, pp. 46–74, nov. 2015.

[3] G. C. Holst y T. S. Lomheim, *CMOS/CCD Sensors and Camera Systems*, vol. 172. JCD Publishing, 2007.

[4] T. Van Lanh, K. S. Chong, S. Emmanuel, y M. S. Kankanhalli, «A Survey on Digital Camera Image Forensic Methods», en *Proceedings of the 2007 IEEE International Conference on Multimedia and Expo*, Beijing, China, 2007, pp. 16–19.

[5] A. L. Sandoval Orozco, D. M. Arenas González, L. J. Garcıa Villalba, y J. C. Hernández Castro, «Anomalıas en el Seguimiento de Exif en el Análisis Forense de Metadatos de Imágenes de Moviles», en *Actas del XII Reunión Espanola sobre Criptologıa y Seguridad de la Información*, Donostia-San Sebastián, Spain, 2012.

[6] A. L. Sandoval Orozco, D. M. Arenas González, L. J. García Villalba, y J. Hernández-Castro, «Analysis of Errors in Exif Metadata on Mobile Devices», *Multimed. Tools Appl.*, vol. 74, n.º 13, pp. 4735–4763, jul. 2015.

[7] M. Boutell y J. Luo, «Beyond Pixels: Exploiting Camera Metadata for Photo Classification», *Pattern Recogn*, vol. 38, n.º 6, pp. 935–946, jun. 2005.

[8] K. San Choi, E. Y. Lam, y K. K. Y. Wong, «Source Camera Identification Using Footprints from Lens Aberration», en *Proceedings of the SPIE 6069, Digital Photography*, San Jose, CA, USA, 2006, p. 60690J–60690J.

[9] L. Tran Van, S. Emmanuel, y M. S. Kankanhalli, «Identifying Source Cell Phone Using Chromatic Aberration», en *Proceedings of the 2007 IEEE International Conference on Multimedia and Expo*, Beijing, China, 2007, pp. 883–886.

[10] S. Bayram, H. T. Sencar, y N. Memon, «Classification of Digital Camera-Models Based on Demosaicing Artifacts», *Digit. Investig.*, vol. 5, n.º 1, pp. 49–59, sep. 2008.

[11] Y. Long y Y. Huang, «Image Based Source Camera Identification Using Demosaicking», en *Proceedings of the 2006 IEEE 8th Workshop on Multimedia Signal Processing*, Victoria, BC, Canada, 2006, pp. 419–424.

[12] O. Celiktutan, I. Avcibas, B. Sankur, y N. Memon, «Source Cell-Phone Identification», *IEEE Signal Process. Commun. Appl.*, pp. 1–3, 2005.

[13] H. Cao y A. C. Kot, «Mobile Camera Identification Using Demosaicing Features», en *Proceedings of the 2010 IEEE international symposium on Circuits and systems (ISCAS)*, Paris, France, pp. 1683–1686.

[14] J. S. Ho, O. C. Au, J. Zhou, y Y. Guo, «Inter-Channel Demosaicking Traces for Digital Image Forensics», en *Proceedings of the 2010 IEEE international conference on Multimedia and expo (ICME)*, Suntec City, Singapore, 2010, pp. 1475–1480.

[15] Q. Liu *et al.*, «Detection of JPEG Double Compression and Identification of Smartphone Image Source and Post-Capture Manipulation», *Appl. Intell.*, vol. 39, n.º 4, pp. 705–726, dic. 2013.

[16] M. J. Tsai, C. L. Lai, y J. Liu, «Camera/Mobile Phone Source Identification for Digital Forensics», en *Proceedings of the ICASSP 2007. IEEE International Conference on Acoustics, Speech and Signal Processing*, Honolulu, HI, USA, 2007, p. II–221.

[17] Y. Hu, C. T. Li, y C. Zhou, «Selecting Forensic Features for Robust Source Camera Identification», en *Proceedings of the 2010 International Computer Symposium (ICS)*, Tainan, Taiwan, Taiwan, 2010, pp. 506–511.

[18] W. Lu, W. Sun, F.-L. Chung, y H. Lu, «Revealing digital fakery using multiresolution decomposition and higher order statistics», *Eng. Appl. Artif. Intell.*, vol. 24, n.º 4, pp. 666-672, 2011.

[19] F. J. Meng, X. W. Kong, y X. G. You, «Source Camera Identification Based on Image Bi-Coherence and Wavelet Features», en *Proceedings of the 4th Annual IFIP WG 11.9 International Conference on Digital Forensics*, Kyoto, Japan, 2008, vol. 285, pp. 207–218.

[20] F. J. Meng, X. W. Kong, y X. G. You, «Source Camera Identification Based on Image Bi-Coherence and Wavelet Features», en *Proceedings of the 4th Annual IFIP WG 11.9 International Conference on Digital Forensics*, Kyoto, Japan, 2008, vol. 285, pp. 207–218.

[21] L. Ozparlak y I. Avcibas, «Differentiating Between Images Using Wavelet-Based Transforms: A Comparative Study», *IEEE Trans. Inf. Forensics Secur.*, vol. 6, n.º 4, pp. 1418–1431, jul. 2011.

[22] Q. Liu *et al.*, «Identification of Smartphone-Image Source and Manipulation», en *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, Berlin, Germany, 2012, pp. 262–271.

[23] Z. J. Geradts, J. Bijhold, M. Kieft, K. Kurosawa, K. Kuroki, y N. Saitoh, «Methods for Identification of Images Acquired with Digital Cameras», en *Proceedings of the SPIE 4232, Enabling Technologies for Law Enforcement and Security*, Boston, USA, 2001, pp. 505–512.

[24] J. Lukas, J. Fridrich, y M. Goljan, «Digital Camera Identification from Sensor Pattern Noise», *IEEE Trans. Inf. Forensics Secur.*, vol. 1, n.º 2, pp. 205–214, jun. 2006.

[25] F. Costa, M. Eckmann, W. J. Scheirer, y A. Rocha, «Open Set Source Camera Attribution», en *Proceedings of the 2012 25th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Ouro Preto, Brazil, 2012, pp. 71–78.

[26] J. Rosales Corripio, D. M. Arenas González, A. L. Sandoval Orozco, y L. J. García Villalba, «Identificación de la Fuente de Imágenes de Dispositivos Móviles Basada en el Ruido del Sensor», en *Actas del XIV Reunión Espanola sobre Criptologıa y Seguridad de la Información* Alicante, Spain, 2014.

[27] C. McKay, A. Swaminathan, H. Gou, y M. Wu, «Image Acquisition Forensics: Forensic Analysis to Identify Imaging Source», en *Proceedings of the ICASSP 2008. IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, NV, USA, 2008, pp. 1657–1660.

[28] H. Huang, W. Guo, y Y. Zhang, «Detection of Copy-Move Forgery in Digital Images Using SIFT Algorithm», en *Proceedings of the 2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application*, Wuhan, China, 2008, pp. 272–276.

[29] C. Li, Q. Ma, L. Xiao, M. Li, y A. Zhang, «Image Splicing Detection Based on Markov Features in {QDCT} Domain», *Neurocomputing*, vol. 228, pp. 29–36, mar. 2017.

[30] C. L. Lai y Y. S. Chen, «The Application of Intelligent System to Digital Image Forensics», en *Proceedings of the IEEE 2009 International Conference on Machine Learning and Cybernetics*, Hebei, China, 2009, vol. 5, pp. 2991–2998.

[31] Z. Liang, G. Yang, X. Ding, y L. Li, «An Efficient Forgery Detection Algorithm for Object Removal by Exemplar-Based Image Inpainting», *J. Vis. Commun. Image Represent.*, vol. 30, pp. 75–85, jul. 2015.

[32] E. Kee y H. Farid, «A Perceptual Metric for Photo Retouching», *Proc. Natl. Acad. Sci.*, vol. 108, n.º 50, pp. 19907–19912, oct. 2011.

[33] J. C. Lee, «Copy-Move Image Forgery Detection Based on Gabor Magnitude», *J. Vis. Commun. Image Represent.*, vol. 31, pp. 320–334, ago. 2015.

[34] M. M. Duarte Villasenor y L. Chang Fernández, «Clasificación de Objetos en Imágenes Usando SIFT», sep. 2015.

[35] A. A. Alahmadi, M. Hussain, H. Aboalsamh, G. Muhammad, y G. Bebis, «Splicing Image Forgery Detection Based on DCT and Local Binary Pattern», en *Proceedings of the 2013 IEEE Global Conference on Signal and Information Processing*, Austin, TX, USA, 2013, pp. 253–256.

[36] W. Wang, J. Dong, y T. Tan, «Effective Image Splicing Detection Based on Image Chroma», en *Proceedings of 2009 16th IEEE International Conference on Image Processing (ICIP)*, Cairo, Egypt, 2009, pp. 1257–1260.

[37] G. Muhammad, M. H. Al-Hammadi, M. Hussain, y G. Bebis, «Image Forgery Detection Using Steerable Pyramid Transform and Local Binary Pattern», *Mach. Vis. Appl.*, vol. 25, n.º 4, pp. 985–995, may 2014.

[38] P. Sutthiwan, Y. Q. Shi, J. Dong, T. Tan, y T. T. Ng, «New Developments in Color Image Tampering Detection», en *Proceedings of the 2010 IEEE International Symposium on Circuits and Systems*, Paris, France, 2010, pp. 3064–3067.

[39] P. Sutthiwan, Y. Q. Shi, W. Su, y T. T. Ng, «Rake Transform and Edge Statistics for Image Forgery Detection», en *Proceedings of the 2010 IEEE International Conference on Multimedia and Expo*, Suntec City, Singapore, 2010, pp. 1463–1468.

[40] J. Lukas, J. Fridrich, y M. Goljan, «Digital "Bullet Scratches" for Images», en *Proceedings of the IEEE International Conference on Image Processing 2005*, Genova, Italy, 2005, p. III-65-8.

[41] A. E. Dirik, H. T. Sencar, y N. Memon, «Digital Single Lens Reflex Camera Identification From Traces of Sensor Dust», *IEEE Trans. Inf. Forensics Secur.*, vol. 3, n.º 3, pp. 539–552, sep. 2008.

[42] K. S. Choi, Y. E. Lam, y K. K. Y. Wong, «Automatic Source Camera Identification Using the Intrinsic Lens Radial Distortion», *Opt Express*, vol. 14, n.º 24, pp. 11551–11565, nov. 2006.

[43] M. Goljan y J. Fridrich, «Camera Identification from Cropped and Scaled Images», en *Proceedings of the SPIE 6819, Security, Forensics, Steganography, and Watermarking of Multimedia Contents*, San Jose, CA, USA, 2008, p. 68190E–68190E–13.

[44] M. Goljan, J. Fridrich, y M. Chen, «Sensor Noise Camera Identification: Countering Counter-Forensics», en *Proceedings of the SPIE 7541, Media Forensics and Security II*, San Jose, CA, USA, 2010, p. 75410S-75410S-12.

[45] M. Goljan, J. Fridrich, y M. Chen, «Defending Against Fingerprint-Copy Attack in Sensor-Based Camera Identification», *IEEE Trans. Inf. Forensics Secur.*, vol. 6, n.º 1, pp. 227–236, mar. 2011.

[46] H. Farid, «Exposing Digital Forgeries from JPEG Ghosts», *IEEE Trans. Inf. Forensics Secur.*, vol. 4, n.º 1, pp. 154–160, feb. 2009.

[47] W. C. Hu, J. S. Dai, y J. S. Jian, «Effective Composite Image Detection Method Based on Feature Inconsistency of Image Components», *Digit. Signal Process.*, vol. 39, pp. 50–62, abr. 2015.

[48] J. Zuo, S. Pan, B. Liu, y X. Liao, «Tampering Detection for Composite Images Based on Re-Sampling and JPEG Compression», en *Proceedings of the 2011 First Asian Conference on Pattern Recognition (ACPR)*, Beijing, China, 2011, pp. 169–173.

[49] T. Bianchi y A. Piva, «Image forgery localization via block-grained analysis of JPEG artifacts», *IEEE Trans. Inf. Forensics Secur.*, vol. 7, n.º 3, pp. 1003–1017, 2012.

[50] Z. Lin, J. He, X. Tang, y C. K. Tang, «Fast, Automatic and Fine-Grained Tampered JPEG Image Detection Via DCT Coefficient Analysis», *Pattern Recognit.*, vol. 42, n.º 11, pp. 2492–2501, 2009.

[51] Z. Lin, J. He, X. Tang, y C.-K. Tang, «Fast, automatic and fine-grained tampered JPEG image detection via DCT coefficient analysis», *Pattern Recognit.*, vol. 42, n.º 11, pp. 2492–2501, 2009.

[52] T. Ojala, M. Pietikainen, y T. Maenpaa, «Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns», *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, n.º 7, pp. 971–987, 2002.

[53] B. Yang y S. Chen, «A comparative Study on Local Binary Pattern (LBP) Based Face Recognition: LBP Histogram Versus LBP Image», *Neurocomputing*, vol. 120, pp. 365–379, 2013.

[54] J. Gu y C. Liu, «Feature Local Binary Patterns with Application to Eye Detection», *Neurocomputing*, vol. 113, pp. 138–152, 2013.

[55] F. Dornaika, A. Bosaghzadeh, H. Salmane, y Y. Ruichek, «Graph-Based Semi-Supervised Learning with Local Binary Patterns for Holistic Object Categorization», *Expert Syst. Appl.*, vol. 41, n.º 17, pp. 7744–7753, 2014.

[56] F. Hakimi, M. Hariri, y F. GharehBaghi, «Image Splicing Forgery Detection Using Local Binary Pattern and Discrete Wavelet Transform», en *Proceeding of the 2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI)*, Tehran, Iran, 2015, pp. 1074-1077.

[57] Dong, J. and Wang, W., *CASIA TIDEv2.0*.

[58] Y.-F. Hsu y S.-F. Chang, «Detecting Image Splicing Using Geometry Invariants and Camera Characteristics Consistency», en *International Conference on Multimedia and Expo*, Toronto, Canada, 2006, pp. 549–552.

[59] C.-C. Chang y C.-J. Lin, «LIBSVM: A library for support vector machines», *ACM Trans. Intell. Syst. Technol.*, vol. 2, n.º 3, p. 27:1–27:27, 2011.