

## Capítulo 2

# Análisis Forense de Imágenes Digitales

El objetivo de este capítulo es mostrar cómo se genera una imagen digital, así como describir los componentes que participan en este proceso. Estos conceptos son la base de las técnicas de análisis forense descritas en los siguientes capítulos.

### 2.1 Formación de una Imagen Digital

Para comprender el análisis forense de las imágenes digitales lo primero que se requiere conocer es cómo está compuesta una cámara fotográfica y cuál es el procedimiento que realiza para generar una imagen (a menudo llamado *pipeline*). Las cámaras fotográficas se componen de 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 o *Digital Image Processor (DIP)* [BSM08]. A pesar de que muchos de los detalles del *pipeline* se mantienen como información confidencial de los fabricantes este proceso es muy similar en la mayoría de las cámaras digitales. La estructura básica se muestra en la Figura 2.1.

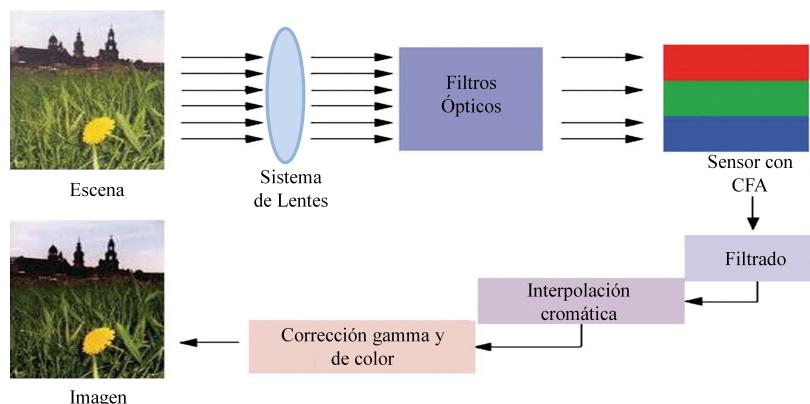


Figura 2.1: Proceso de adquisición de imágenes en cámaras digitales

Como primer paso para generar una imagen el sistema de lentes captura la luz de la escena controlando la exposición, el foco y la estabilización de la imagen. Después, la luz que entra en la cámara a través del sistema de lentes pasa por un grupo de filtros que mejora la calidad visual de la imagen. Este grupo incluye al menos un filtro infrarrojo y un filtro *anti-aliasing*. El filtro infrarrojo absorbe o refleja la luz permitiendo que sólo la parte visible del espectro pase a la siguiente fase, evitando que la radiación infrarroja ocasione pérdida de nitidez en la imagen. El filtro *anti-aliasing* se encarga de limpiar la señal produciendo imágenes con contornos más suaves.

A continuación la luz pasa al sensor de la imagen que es una matriz de elementos sensibles a la luz llamados píxeles. Cada elemento de esta matriz de píxeles integra la luz incidente y genera una señal analógica proporcional a la intensidad de la luz recibida. Esta señal se convierte en una señal digital y se transmite al procesador de imagen. Debido a que el sensor de la imagen es monocromático, para capturar una imagen a color se requieren diferentes sensores. Idealmente, un sensor para cada color. Sin embargo, debido al coste que esto implica, en la mayoría de las cámaras sólo se usa un sensor de imagen junto a una matriz de filtros de color que se coloca antes del sensor para producir los colores.

Una vez que el procesador de imagen recibe la señal digital generada por el sensor elimina el ruido y otras anomalías introducidas en las señales digitales (*artifacts*), con la finalidad de obtener una imagen visualmente agradable. Uno de los procesos que se realizan sobre la señal es la llamada *interpolación cromática* (*demosaicing*) encargada de calcular los valores de los colores faltantes debido a que el sensor únicamente proporciona información sobre una cierta cantidad de colores (los que permite pasar la matriz de filtros de color). Un proceso adicional es la corrección de píxeles defectuosos originados por imperfecciones en el sensor, que corrige estos píxeles mediante interpolación.

Otro proceso al que se somete la imagen es el *balanceo de blancos*, que permite una reproducción más fiel del color, evitando que haya colores dominantes. Por último, el *proceso de corrección gamma* ajusta los valores de intensidad de la imagen. Aunque los algoritmos para llevar a cabo estos procesos están presentes en todas las cámaras, los detalles exactos de la forma de realizarlos pueden variar entre los diferentes fabricantes e, incluso, entre los modelos de un mismo fabricante.

Finalmente, la imagen generada por el procesador de imagen se comprime. En las cámaras de dispositivos móviles normalmente se utiliza el algoritmo *Joint Photographic Experts Group* (JPEG) [Ham] para ahorrar espacio, almacenándose en la memoria del dispositivo junto con la información de la imagen en formato *Exchangeable Image File Format* (EXIF) [RSYD05].

### 2.1.1 Filtros de Color

La matriz de filtros de color es una de las partes más importantes de la cadena de procesamiento para la generación de una imagen de las cámaras de un solo sensor [APS98]. La CFA se encuentra sobre el sensor monocromo, y su función es adquirir la información del

**color de la escena.** Cada celda del filtro de color deja pasar la luz de acuerdo a un rango de longitudes de onda, de tal manera que las intensidades filtradas separadas incluyen información sobre el color de la luz. Como se ilustra en la Figura 2.2, **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** (considerando que cada píxel corresponde a un valor de intensidad).

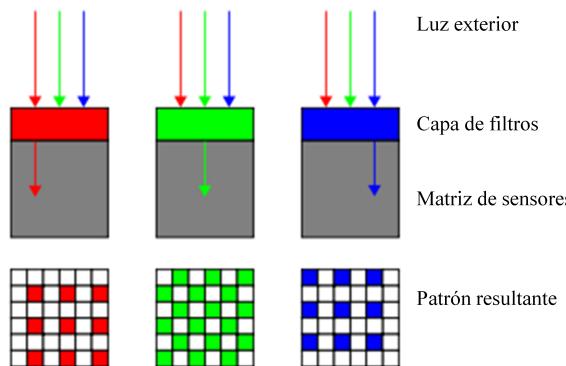


Figura 2.2: Matriz de filtros de color (CFA)

En este punto el proceso la interpolación cromática se lleva a cabo para obtener los valores que faltan para cada uno de los colores del filtro CFA. Este proceso es el más complejo en cuanto a cómputo se refiere. Su algoritmo utiliza los valores de los píxeles vecinos para obtener todos los valores que no han sido medidos.

Es posible que los tipos de filtros de color y la forma de realizar la interpolación cromática varíen entre fabricantes. Además, existe la posibilidad de que la imagen se almacene en formato CFA y el proceso de interpolación cromática se realice en un ordenador, extendiendo aún más la posibilidad de variaciones en este proceso.

El diseño de la matriz CFA utilizada influye en la imagen resultante de la cámara, tanto en la nitidez y apariencia de los bordes como en los pequeños detalles. A veces, el proceso de interpolación cromática puede generar anomalías en la imagen tales como el *aliasing* (efecto que produce el aspecto desagradable de líneas escalonadas “sierras” en los contornos de las imágenes), ruido y distorsiones en el color. El uso de otro filtro puede eliminar la presencia de estas imperfecciones en determinadas áreas de la imagen a costa de degradar la calidad en otras [LP05].

Generalmente, las cámaras usan el modelo *Green-Red-Green-Blue* (GRGB) del patrón CFA de Bayer. La salida de un sensor de este tipo es un mosaico de píxeles rojos, verdes y azules de diferentes intensidades. Como se observa en la Figura 2.2, este filtro captura el 25% de los píxeles en el canal rojo, el 50% en el canal verde y el 25% restante en el canal azul. Otras alternativas de filtros CFA son los patrones *Cyan-Yellow-Yellow-Magenta* (CYYM), *Red-Green-Blue-Emerald* (RGBE) y *Cyan-Magenta-Yellow* (CMY).

### 2.1.2 Tipos de Sensores

El sensor de la imagen es la parte más importante de las cámaras digitales. Generalmente, se considera el corazón de la cámara. Éste es una matriz de elementos sensibles a la luz llamados píxeles. Los píxeles están hechos de silicio y capturan la luz convirtiendo los fotones en electrones utilizando el efecto fotoeléctrico. Cada píxel se encarga de acumular la carga inducida por la luz durante un determinado tiempo de exposición para luego ser leído y procesado. La señal de salida del sensor es proporcional a la carga acumulada, dependiendo de la cantidad de luz que incida sobre el píxel y del tiempo de exposición a ella.

Existe una extensa literatura sobre el desarrollo y tecnologías de los sensores [HL07, Nak05, HKT07, AP13]. Sin embargo, para el propósito de este trabajo basta con tener una visión general para comprender el ruido que el sensor puede introducir en las imágenes que genera.

Los sensores de la imagen se agrupan de acuerdo a sus procesos de fabricación en CCD y CMOS [HL07]. Los dos tipos de sensores están formados esencialmente por semiconductores de metal-óxido *Metal Oxide Semiconductor (MOS)* distribuidos en forma de matriz y funcionan de una manera muy similar. Sin embargo, hay características que diferencian a estas tecnologías.

#### 2.1.2.1 Sensores CCD

La diferencia clave entre las dos tecnologías de sensores es el lugar en el que se digitalizan los píxeles y la forma en la que se lleva a cabo la lectura de las cargas.

En el caso de los sensores CCD cada una de las cargas de las celdas de la matriz se transforman en voltajes y se entrega una señal analógica como salida para que posteriormente se digitalice por la cámara. La estructura de este tipo de sensores es muy sencilla, pero tiene como inconveniente la necesidad de contar con un chip adicional que trate la información de salida del sensor (implicando equipos más grandes y costosos).

A diferencia de los sensores CMOS que soportan la lectura de la matriz de píxeles de una manera aleatoria, en los sensores CCD todos los píxeles comienzan y finalizan la integración de carga al mismo tiempo. Esto propicia una salida uniforme (resultado que se espera de un píxel sometido al mismo nivel de excitación de los demás sin que se presenten cambios notables en la señal obtenida). A este tipo de exposición se le conoce como *global shutter*. Es posible añadir circuitos en los sensores de CMOS para hacer que den un resultado similar. Sin embargo, siguen estando sobre ellos los sensores de tipo CCD.

Los sensores del tipo CCD son, por mucho, mejores que los de tipo CMOS en cuanto al rango dinámico (coeficiente entre la saturación de los píxeles y el umbral por debajo del cual no captan señal), puesto que al ser menos sensibles toleran mejor los extremos de luz. Asimismo, los sensores CCD son superiores a los CMOS en términos de ruido en la imagen, puesto que el procesado de las señales se lleva a cabo en un chip externo que puede optimizarse para el desarrollo de esta función. En contraste, los sensores CMOS

realizan el procesamiento de la señal dentro del mismo sensor dejando menos espacio para colocar los foto-diodos encargados de recolectar la luz.

### 2.1.2.2 Sensores CMOS

Los sensores CMOS son sensores con un diseño de píxeles activos e independientes. Se denominan píxeles activos debido a que la digitalización se realiza en ellos internamente en unos transistores que ofrecen mejor velocidad de procesamiento, eliminándose la necesidad de un chip externo que realice esta función, lo que reduce el coste y el tamaño de los equipos.

La característica de independencia se refiere a la flexibilidad que este tipo de sensores ofrece para la lectura de la matriz de píxeles, ya que es posible acceder a cada celda mediante la posición de su fila y columna. Generalmente, la lectura de la matriz se realiza en forma de barrido progresivo. A este esquema se le conoce como *rolling shutter* (no es necesario leer la matriz completa en un solo tiempo como en los sensores **CCD**). Además, al estar formados por celdas independientes, los sensores **CMOS** no presentan el efecto **blooming**. Este efecto se produce cuando un píxel se satura por la luz que incide sobre él y a continuación comienza a saturar a los que se encuentran a su alrededor.

Una ventaja más es que los sensores **CMOS** son más sensibles a la luz y en condiciones de poca iluminación se comportan mejor. Adicionalmente, debido a que los amplificadores de la señal se encuentran dentro de la misma celda, no se genera un consumo extra de alimentación a diferencia de los sensores **CCD**.

En sus inicios los sensores **CMOS** no eran considerados tan buenos como los sensores **CCD**. Sin embargo, la tecnología **CCD** ha llegado a su límite y ahora es cuando se está desarrollando la tecnología **CMOS** superando sus deficiencias [**CCD**]. La mayoría de las cámaras utilizan sensores **CCD**, aunque en dispositivos móviles es más común el uso de sensores **CMOS**.

### 2.1.3 Imperfecciones y Ruido de la Imagen

#### 2.1.3.1 Imperfecciones del Sensor

Durante el proceso de generación de una imagen es posible que se introduzcan algunos defectos que se vean reflejados como ruido en la imagen final.

Teniendo una noción básica del funcionamiento de los sensores, se pueden analizar los defectos que generan ruido en las imágenes finales. Estos defectos son de gran ayuda para identificar la cámara que generó una imagen determinada.

Se consideran defectos los píxeles que tienen una respuesta lo suficientemente anormal como para ser descartados y no formar parte de los datos de la imagen final. A pesar de que en la cadena de procesamiento de la imagen se realizan procedimientos para tratar de mitigar estos defectos, las correcciones pueden no ser del todo perfectas y es posible injectar defectos ocultos en la imagen; incluso algunos defectos son tolerados por los fabricantes con tal de mantener o mejorar el rendimiento de las cámaras.

De acuerdo a los factores que los ocasionan, los defectos se pueden agrupar en:

- **Defectos de fila y columna:** Pueden ser ocasionados durante el proceso de transferencia de carga. Debido a la forma en que los sensores CMOS direccionan los píxeles se pueden generar errores parciales o totales en filas o columnas de píxeles.
- **Defectos de grupo :** Este tipo de defectos afectan a un conjunto de píxeles. Pueden ser ocasionados por defectos en la superficie del sensor como suciedad o rayas. También pueden ser causados por fallos eléctricos como es el caso de algunos sensores de tipo CMOS en los que múltiples píxeles (generalmente 3 ó 4) comparten circuitería para convertir la carga en voltaje, y al haber un fallo en alguno de estos píxeles se produce un defecto en grupo.
- **Píxeles calientes:** Son los píxeles que generan altas salidas de voltaje bajo cierto tipo de condiciones, especialmente en exposiciones largas. Los puntos que se obtienen en la imagen final son siempre muy brillantes y puede ser de cualquier color, dependiendo del punto del patrón Bayer que esté precisamente frente al píxel en cuestión.
- **Píxeles muertos:** Son los píxeles que tienen una respuesta muy pobre a la luz, apareciendo como puntos negros en las imágenes finales.
- **Diferencias entre salidas múltiples:** En los sensores que tienen más de una salida pueden presentarse variaciones entre las diferentes salidas, especialmente si utilizan un convertidor analógico/digital para cada una de las salidas. Los rasgos creados en la imagen por la falta de coincidencia tendrán una textura dependiendo de la disposición geométrica de los píxeles canalizados a través de cada salida.
- **Interferencia:** Este defecto se produce cuando los fotones que deberían de ser recolectados por un píxel se recogen por un píxel vecino. La mayoría de sensores sufren este tipo de defecto que puede ser causado por problemas de reflexión en el sistema de lentes o por la difusión de la carga durante exposiciones a la luz prolongadas.
- **Saturación:** Sucede cuando un píxel acumula más carga de la que puede contener y el exceso de la carga es pasada a los píxeles vecinos generando el efecto *blooming*.
- **Rolling Shutter:** La técnica de *rolling shutter* utilizada en los sensores CMOS puede crear distorsiones en la imagen cuando la escena cambia significativamente mientras está siendo capturada como cuando hay movimientos en la escena (deformando la imagen) o cambios de iluminación (introduciendo líneas en la misma).
- **Corriente de oscuridad:** Surge de las impurezas del cristal de silicio de los sensores. Es la señal acumulada en cada píxel incluso en la ausencia de luz y que varía además con la temperatura [Nak05]. En los sensores pequeños estas variaciones son muy pequeñas, pero en los sensores de mayor tamaño suelen ser más significativas.

### 2.1.3.2 Ruido en la Imagen

Existen diversas fuentes de imperfecciones y ruido introducidas en las diferentes etapas del proceso de generación de la imagen en la cámara. Incluso si se toma una fotografía uniforme y completamente iluminada es posible observar pequeños cambios de intensidad entre los píxeles. Esto se debe al ruido de disparo que es aleatorio y, en gran parte, al patrón de ruido que es determinista y se mantiene aproximadamente igual si se toman varias fotografías de la misma escena.

El patrón de ruido en una imagen se refiere a cualquier patrón espacial que no cambia de una imagen a otra y está compuesto por el ruido espacial que es independiente de la señal o ruido de patrón fijo *Fixed Pattern Noise* (FPN) y el ruido espacial debido a la diferencia de respuesta de cada píxel a la señal incidente o ruido de respuesta no uniforme *Photo Response Non Uniformity* (PRNU) [KMC<sup>+</sup>06, LFG06, AP13]. La estructura del patrón de ruido se ilustra en la Figura 2.3.

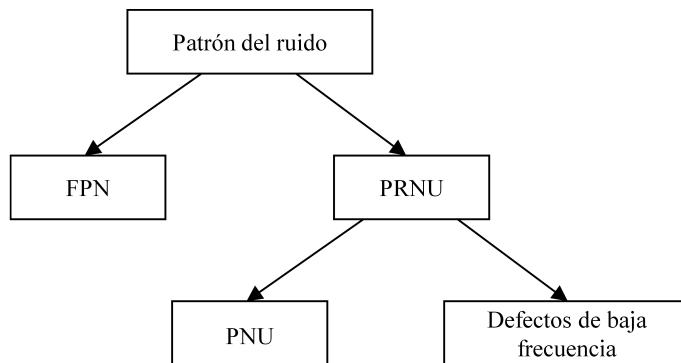


Figura 2.3: Patrón de ruido del sensor

El ruido FPN se genera por la corriente de oscuridad y también depende de la exposición y de la temperatura. Debido a que el ruido del patrón fijo es un ruido independiente aditivo, algunas cámaras lo eliminan automáticamente restando un marco oscuro a las imágenes que generan.

El ruido PRNU es la parte dominante del patrón de ruido de las imágenes y es un ruido dependiente multiplicativo. El ruido PRNU está formado principalmente por la uniformidad de pixel *Pixel Non-Uniformity* (PNU) y los defectos de baja frecuencia como la configuración del *zoom* y la refracción de la luz en las partículas de polvo y lentes.

El ruido PNU es la diferencia de sensibilidad a la luz entre los píxeles de la matriz del sensor. Se genera por la falta de homogeneidad de las obleas de silicio y las imperfecciones durante el proceso de fabricación del sensor. Debido a su naturaleza y origen es muy poco probable que incluso los sensores procedentes de la misma oblea presenten patrones PNU correlacionados. Este ruido no se ve afectado por la temperatura ambiente ni por la humedad.

El ruido PNU es normalmente más común, complejo y significativo en los sensores de

tipo CMOS debido a la complejidad de la circuitería de la matriz de píxeles.

#### 2.1.4 Diferencia entre Cámaras Digitales y Cámaras de Dispositivos Móviles

A pesar de que las cadenas de procesamiento de la imagen entre las cámaras digitales tradicionales y las cámaras de dispositivos móviles son muy parecidas, existen algunas diferencias significativas en cuanto a calidad entre las cámaras [CSA08].

Las cámaras de dispositivos móviles generan imágenes de menor calidad debido a varios factores relacionados principalmente con el *hardware* que utilizan dada la naturaleza compacta de este tipo de dispositivos. Estos factores son:

- **Apertura de la lente:** Restringida a tener valores pequeños para la apertura de la lente.
- **Resolución:** Las resoluciones son menores.
- **Distancia focal:** Tienen una distancia focal fija y restringida a valores pequeños que limita las condiciones de iluminación.
- **Flash:** Muchos de los dispositivos móviles no cuentan con *flash* y, en caso de tenerlo, no es muy robusto debido a las limitaciones de potencia. La sensibilidad a la luz del sensor de acuerdo a su tipo afecta directamente a la velocidad de obturación, y esto se relaciona con la falta de definición de la imagen.
- **Conversión Analógica Digital:** Los dispositivos móviles están limitados al uso de *Analog Digital Conversion (ADC)* de 10 bits mientras que las cámaras tradicionales típicamente usan uno de 12 bits.

Como se muestra en [CSA08] las características generadas por las cámaras digitales tradicionales y las de dispositivos móviles son diferentes. En los resultados de los experimentos realizados sólo un 49,5 % de las características resultaron ser comunes y, por lo tanto, no son intercambiables.

Las huellas CFA son más prominentes en las cámaras digitales tradicionales mientras que las cámaras de dispositivos móviles tienen una mayor contaminación de ruido debido a los factores mencionados anteriormente y a la diferencia de calidad entre los sensores CMOS y CCD. Es por ello que las técnicas de identificación de fuente basadas en el ruido del sensor y las que se basan en la transformada *wavelet* resultan ser más adecuadas en dispositivos móviles.

## Capítulo 3

# Técnicas de Análisis Forense en Imágenes

En este capítulo se describen las principales técnicas de análisis forense de imágenes digitales haciendo énfasis en las técnicas de identificación de la fuente de la imagen, ya que es la rama del análisis forense en la que se centra este trabajo.

Según [CFGL08] las tareas de análisis forense de imágenes digitales se pueden dividir en las siguientes categorías:

- **Verificación de integridad o detección de falsificaciones:** Busca descubrir procedimientos maliciosos que se hayan aplicado a las imágenes como, por ejemplo, recorte o adición de objetos a una imagen.
- **Recuperación de la historia de procesamiento:** Tiene como objetivo recuperar la cadena de procesamientos que han sido aplicados a una imagen de una manera no maliciosa como, por ejemplo, recortes, filtrados, contrastes, etc.
- **Clasificación basada en la fuente:** Tiene como objetivo clasificar las imágenes de acuerdo a su origen en cámaras digitales o escáneres.
- **Agrupación por dispositivos fuente:** Dado un grupo de imágenes se buscan los grupos de imágenes que fueron obtenidas utilizando la misma cámara.
- **Identificación de la fuente:** Busca determinar el dispositivo que generó una imagen determinada.

### 3.1 Técnicas de Identificación de la Fuente

La investigación en este campo estudia el diseño de técnicas para identificar las características, especialmente marca y modelo, de los dispositivos utilizados para la generación de imágenes digitales.

El éxito de estas técnicas depende del supuesto de que todas las imágenes adquiridas por un mismo dispositivo presentan características intrínsecas del dispositivo. Las características que se usan para identificar marca y modelo de las cámaras digitales se derivan de las diferencias que existen entre las técnicas de procesamiento de las imágenes y las tecnologías de los componentes que se utilizan. El mayor problema con este enfoque es que los diferentes modelos de las cámaras digitales usan componentes de un número reducido de fabricantes, y que los algoritmos que usan también son muy similares entre modelos de la misma marca. Es por ello que la fiabilidad de la identificación de la cámara fuente depende en gran parte de la identificación de varias características independientes del modelo. Según [VCEK07] se pueden establecer cuatro grupos de técnicas para este fin: utilización de la aberración de las lentes, interpolación de la matriz CFA, uso de las características de la imagen e imperfecciones del sensor. Esta última constituye el objeto de este trabajo. Además de las anteriores existe otro grupo de técnicas basadas en los metadatos.

### 3.1.1 Técnicas Basadas en Metadatos

Las cámaras digitales cuentan con una poderosa fuente de información que son los metadatos embebidos en los archivos de las imágenes digitales que generan. Los metadatos o “datos sobre datos” registran información relacionada con las condiciones de captura de la imagen, como fecha y hora de generación, presencia o ausencia de *flash*, distancia de los objetos, tiempo de exposición, apertura del obturador, *Global Positioning System (GPS)*, entre otros. En otras palabras, información de interés que complementa el contenido principal de un documento digital. Los metadatos pueden llegar a ser una potente ayuda para la organización y búsqueda a lo largo de librerías de imágenes.

Las imágenes digitales son almacenadas en una gran variedad de formatos como *Tagged Image File Format (TIFF)* [Ass], *JPEG* [Ham] y *Photoshop Data file (PSD)* u otros propietarios como *RAW*. Algunos de los distintos contenedores de metadatos para los distintos formatos son: *Image File Directoys (IFDs)* *EXIF/TIFF*, Adobe *eXtensible Metadata Platform (XMP)* [XMP] e *IPTC-IIM* [IPT]. La especificación *EXIF* [Comb] es la más utilizada para identificación de la fuente por ser el contenedor de metadatos más común en las cámaras digitales [Bae10]. La especificación *EXIF* incluye cientos de etiquetas, entre las que se encuentran *marca* y *modelo*. Desafortunadamente, el seguimiento del estándar no es preceptivo.

Las técnicas basadas en el análisis de los metadatos de la imagen son las más sencillas y existen gran cantidad de trabajos enfocados en los diferentes tipos de metadatos tanto para la búsqueda de información como para la clasificación de imágenes e identificación de la fuente [BL04, BL05, Tes05, RCC<sup>+</sup>08, Are11].

Sin embargo, estas técnicas dependen en gran medida de los metadatos que los fabricantes deciden insertar cuando la imagen es generada. Asimismo, este método es el más vulnerable a modificaciones malintencionadas e incluso a la eliminación total de los metadatos ya sea intencionalmente o de manera inconsciente. Ejemplo de ello son algu-

nos programas de edición fotográfica que al editar o comprimir una imagen actualizan incorrectamente los metadatos o provocan la pérdida de los mismos.

A pesar de las debilidades de este tipo de técnicas, si existe el archivo de metadatos y de alguna manera se logra comprobar que no ha sufrido modificaciones externas, su uso es de gran utilidad para los analistas forenses, ya que del contenido de la imagen no se puede inferir toda la información contenida en los metadatos como es el caso de la información de [GPS](#).

Un ejemplo del aprovechamiento de la información contenida en los metadatos es [Pla00] donde las etiquetas de tiempo han sido utilizadas satisfactoriamente para agrupar imágenes por eventos.

Un ejemplo más de la utilidad de los metadatos se presenta en [BL05] que mejora el proceso de clasificación de escenarios de imágenes con el apoyo del análisis de los metadatos. En el citado trabajo se presenta un método probabilístico para la fusión de las evidencias de los metadatos con la información proveniente de un clasificador del contenido de la imagen.

En los experimentos se consideran tres problemas para la clasificación de imágenes: interiores y exteriores, escenas de la naturaleza y de objetos creados por el hombre (esto es, naturales y artificiales) y, por último, la detección de puestas de sol. El análisis de las estadísticas de los metadatos de cada una de estas clases revela que algunas etiquetas como el tiempo de exposición, el flash y la distancia de los objetos son las más representativas para cada problema.

### 3.1.2 Técnicas Basadas en la Aberración de las Lentes

Durante el proceso de generación de la imagen en la parte del sistema de lentes se pueden introducir aberraciones. Existen diferentes tipos de aberraciones: esférica, coma, astigmatismo, curvatura de campo, distorsión radial y distorsión cromática. La distorsión radial es la que más consecuencias tiene sobre la imagen, especialmente en las cámaras que usan lentes baratas de gran angular (*wide angle*). La mayoría de cámaras digitales usan este tipo de lentes por cuestiones de coste.

En [Cho06] se propone la distorsión radial de la lente como la mejor técnica para la identificación de la fuente. La distorsión radial produce que las líneas rectas aparezcan como curvas en la imagen. Los autores concluyen que los diferentes fabricantes emplean diseños diferentes en los sistemas de lentes para compensar este efecto, dando como resultado que cada modelo de cámara exprese un único patrón de distorsión radial que ayuda a identificarla de manera única. El grado de distorsión radial de cada imagen se puede medir mediante un procedimiento que consta de tres fases: Detección de bordes, extracción de segmentos distorsionados y medición del error de la distorsión. En los experimentos se utilizaron tres cámaras diferentes y obtuvieron como resultado una precisión del 91,28 % en la identificación de la fuente.

### 3.1.3 Técnicas Basadas en la Interpolación de la Matriz CFA

Algunos autores consideran que la elección de la matriz de colores [CFA](#) y la especificación de los algoritmos de interpolación cromática generan algunas de las diferencias más marcadas entre los diferentes modelos de cámaras [[BSM06](#), [CAS<sup>+</sup>06](#), [LH06](#), [BSM08](#)].

Como se ha comentado en la sección [2.1](#), en las cámaras comerciales que tienen un solo sensor en lugar de tener sensores separados para cada componente del color

es crucial utilizar la matriz [CFA](#) y los algoritmos de interpolación cromática para capturar correctamente los detalles de la imagen. Estos algoritmos tienen un gran impacto en la calidad de los colores y en los contornos de la imagen resultante. En esencia, la interpolación cromática introduce un tipo específico de correlación entre los valores de colores de los píxeles de la imagen. La forma específica de estas dependencias (de estas “huellas dactilares”) se puede extraer de las imágenes para diferenciar los algoritmos de interpolación cromática y así determinar marca y modelo de la cámara que generó una imagen.

Dentro de este tipo de técnicas se pueden diferenciar tres grupos:

- **Huellas en la Interpolación del Color:** En [[BSM08](#)] se presenta un algoritmo para identificar y clasificar las operaciones de interpolación cromática. La propuesta se basa en dos métodos para realizar el proceso de clasificación: el primer método utiliza un algoritmo [Expectation-Maximization \(EM\)](#) para analizar la correlación del valor de cada píxel con los valores de sus vecinos; el segundo método realiza un análisis de las diferencias entre píxeles (*inter-pixel*). Los experimentos se realizaron en dos fases: la primera fase tenía como objetivo evaluar la precisión del método de identificación de marca y modelo; la segunda evaluaba la precisión del método de identificación individual de la cámara cuando estas eran de la misma marca y modelo. Los resultados obtenidos en la identificación de la fuente de una imagen entre cuatro y cinco modelos diferentes tuvieron una precisión del 88 % y 84,8 % respectivamente. En los experimentos se utilizaron imágenes con ajustes automáticos y con el más alto nivel de calidad de compresión.
- **Modelo de Correlación Cuadrática de Píxeles:** En [[LH06](#)] se utilizan las correlaciones entre píxeles en el proceso de identificación de la fuente. Definen un modelo de correlación cuadrática de los píxeles y obtienen una matriz de coeficientes para cada banda de color. Para la clasificación utilizan redes neuronales. Se probó el método para cuatro cámaras con imágenes de dibujos animados y el éxito fue de un 95 % para una cámara, del 98 % para dos cámaras y del 100 % para el resto de las cámaras. También se realizaron pruebas para imágenes modificadas (incluyendo compresión) con resultados de un 80 % de éxito para una compresión [JPEG](#) del 80 %. Dado que las cámaras del mismo fabricante utilizan el mismo algoritmo de interpolación cromática, esta técnica no es eficiente entre distintos modelos del mismo fabricante. Asimismo, como demuestran los experimentos, no obtienen buenos resultados cuando las imágenes han sido modificadas o comprimidas.

- **Medidas de Similitud Binarias:** En [CAS<sup>+</sup>06] se utiliza un conjunto de medidas de similitud binarias como métricas para estimar la semejanza entre los planos de bits de una imagen. El supuesto fundamental de este trabajo es que el algoritmo de interpolación CFA de cada fabricante deja correlaciones a lo largo de los planos de bits de una imagen y pueden ser representados por este conjunto de medidas. En este estudio se utilizaron 108 medidas de similitud binarias que se obtienen para el propósito de la clasificación de imágenes.

Los experimentos realizados con la técnica de medidas de similitud binaria para clasificar 3 grupos de cámaras obtuvieron un porcentaje de éxito entre el 81 % y el 98 %, mientras que para un grupo de 9 cámaras la precisión descendió al 62 % recolectando 200 imágenes de cada una de las cámaras. Las fotografías utilizadas en los experimentos fueron tomadas con las siguientes configuraciones: máxima resolución con un tamaño de 640 x 480 píxeles, a la luz del día y modo de enfoque automático. Claramente se puede apreciar que los resultados del método dependen del número de cámaras utilizadas en los experimentos.

### 3.1.4 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. Estas características se dividen en tres grupos: características de color, métricas de calidad de la imagen *Image Quality Metrics (IQM)* y estadísticas del dominio *wavelet*.

En [TLL07] se propone un método de identificación de la fuente utilizando las siguientes características: color, calidad de la imagen y dominio de la frecuencia. En el estudio adoptan la transformada *wavelet* como método para calcular las estadísticas del dominio *wavelet* y utilizan *SVM* [HCL03] para la clasificación. En los experimentos realizados se usaron cámaras digitales y dispositivos móviles. Los resultados obtenidos para cuatro modelos diferentes de dos fabricantes dieron una precisión cercana al 92 %.

En [MSGW08] se extiende la identificación de la fuente a diferentes dispositivos tales como teléfonos móviles con cámara integrada, cámaras digitales, escáneres y computadoras. Para ello se identifican en primer lugar las fuentes de variación entre los diferentes tipos de dispositivos y, posteriormente, entre diferentes modelos de los mismos.

En esta propuesta se usan las diferencias en el proceso de adquisición de la imagen de los dispositivos para formar dos grupos de características: coeficientes de interpolación de color y características de ruido. En los experimentos se utilizaron cinco modelos de teléfonos móviles, cinco modelos de cámaras digitales y cuatro modelos de escáneres para identificar el tipo de fuente. En los resultados globales se obtuvo un 93,75 % de precisión. En el análisis de identificación de marca y modelo de teléfonos móviles obtuvieron una precisión del 97,7 % para los cinco modelos.

En [MKY08] se propone un método que emplea las fases y las magnitudes de las estadísticas de bi-coherencia junto a las estadísticas de los coeficientes *wavelet*. Este método

captura las distorsiones únicas no lineales en el dominio *wavelet* producidas por las cámaras cuando realizan operaciones de procesamiento sobre las imágenes.

En primer lugar, para obtener la caracterización de las distorsiones no lineales se extraen las características de bi-coherencia: el bi-espectro de las señales se calcula dividiendo la señal en N segmentos (posiblemente traslapados). Posteriormente, se calcula la Transformada de Fourier para cada segmento, y se promedian las estimaciones individuales. La magnitud (la media de la magnitud de la bi-coherencia) y la fase (la entropía negativa de la fase de la bi-coherencia) se calculan como características estadísticas. Para reducir la memoria y la sobrecarga computacional implicada en el cálculo total de la bi-coherencia de cuatro dimensiones de las imágenes, restringen su análisis a filas, columnas y rodajas radiales a través del centro de la imagen de una sola dimensión.

Es interesante observar que en este trabajo no hay restricciones rigurosas en cuanto a la selección de las imágenes de muestra, debido a que con la aplicación de las estadísticas de bi-coherencia no es necesario extraer la información asociada con el contenido de la imagen (por ejemplo, segmentos de línea). A continuación, se emplea la descomposición *wavelet* en cuatro niveles para dividir el espacio de la frecuencia en cuatro escalas y orientaciones. Después se calculan cuatro estadísticas (media, varianza, asimetría y curtosis) de cada coeficiente de la sub-banda así como los errores de predicción lineal para cada orientación, nivel y canal de color.

Estas estadísticas componen el segundo grupo de vectores de características estadísticas utilizadas para la identificación de la cámara fuente. Una vez que las estadísticas de bi-coherencia y *wavelet* se calculan, el algoritmo *Sequential Forward Feature Selection (SFFS)* [PNK94] se utiliza para reducir la correlación entre las características y la carga computacional manteniendo la misma precisión de la clasificación. El método *SFFS* analiza todas las características y construye el conjunto más representativo de ellas, añadiendo y quitando características hasta que no haya más mejoras disponibles.

Por último, las características más representativas son clasificadas por una *SVM* utilizando un *kernel Radial Basis Function (RBF)*. Se realizaron experimentos bajo las siguientes condiciones: 6 modelos de cámara de 4 fabricantes, imágenes de diferentes resoluciones, formato *JPEG*, un total de 2.100 (350 de cada cámara) imágenes de tomas típicas variando la naturaleza de las escenas. Como resultado obtuvieron un notable porcentaje de precisión en la identificación de la fuente que supera el 97 % distinguiendo diferentes modelos del mismo fabricante. Caben posibles mejoras mediante la incorporación de otras características como las utilizadas en [WGKM09].

En [WGKM09] se propone un método para la identificación de la cámara fuente mediante la extracción y clasificación de las estadísticas de las características *wavelets*. Este método está compuesto por tres fases: extracción, selección y clasificación de características *wavelet*. Las características sobresalientes de dominio *wavelet* se extraen para integrar un modelo estadístico de imagen a partir de los coeficientes *wavelet*, incluyendo 216 características *wavelet* de primer orden y 135 características de co-ocurrencia de segundo orden. En este estudio las características del dominio *wavelet* se consideran más representativas

y son preferidas a las características espaciales (color de la imagen e [IQM](#)) y matrices de filtros de color [CFA](#). De manera análoga a [\[MKY08\]](#) se realiza una descomposición *wavelet* en 4 niveles basada en *Separable Quadrate Mirror Filters (QMF)* para dividir el espacio de la frecuencia, se extraen las mismas cuatro estadísticas (media, varianza, asimetría y curtosis) junto a los errores de predicción lineal. Ya que estas cuatro estadísticas no brindan información sobre la correlación de la textura se usan las características de co-ocurrencia para la extracción de características de textura de la imagen ya que según [\[RH99\]](#) son las más adecuadas para este propósito. A partir de las características de co-ocurrencia se extraen las características de segundo orden (energía, entropía, contraste, homogeneidad y correlación).

Por último, y al igual que en [\[MKY08\]](#), se seleccionan las características más representativas utilizando un algoritmo [SFFS](#) y se clasifican utilizando una [SVM](#) con un *kernel* no lineal. Bajo las mismas condiciones que en los experimentos realizados en [\[MKY08\]](#) logran distinguir entre diferentes modelos del mismo fabricante de cámaras Canon.

Este trabajo puede mejorarse evaluando la robustez del sistema de identificación propuesto por el vector de características, así como ampliando la base de datos de imágenes para realizar las pruebas incluyendo dispositivos móviles, además de cubrir más marcas, modelos, texturas y contenidos.

En [\[OA11\]](#) se plantea una técnica para diferenciar imágenes usando las transformaciones de la familia *wavelet*. Proponen modelos estadísticos para *ridgelet* y sub-bandas *contourlet*, que se describen seguidamente.

- **La Transformada Ridgelet:** Básicamente, la Transformada *Wavelet* es buena para la representación de singularidades de cero dimensiones o puntos. Sin embargo, las señales de 2-D (imágenes) por lo general contienen singularidades 1-D (bordes y esquinas). Para resolver la debilidad de la Transformada *Wavelet* en dos o más dimensiones en [\[CD99\]](#) se desarrolló un nuevo sistema de representaciones llamado *ridgelet* que puede cubrir eficazmente las singularidades de líneas en dos dimensiones. La idea es utilizar la Transformada Radón o *Radon Transform (RAT)* para mapear las singularidades de línea a singularidades punto, y así luego poder manejarlas de una manera más efectiva mediante el uso de la Transformada *Wavelet*.
- **Transformada Contourlet:** En la pintura se utilizan líneas y contornos en lugar de puntos para crear imágenes. La representación *wavelet* de una imagen es el equivalente a usar puntos en lugar de líneas; en este caso la imagen no es clara y su construcción es más difícil, como se puede observar en la Figura [3.1\(a\)](#). Del mismo modo, la representación llamada *contourlet* [\[DV05\]](#) es el equivalente a usar líneas, lo que simplifica la construcción de la imagen y le da un aspecto más realista como se aprecia en la Figura [3.1\(b\)](#).

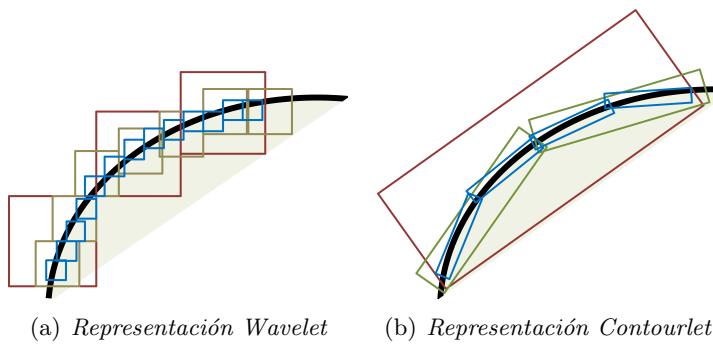


Figura 3.1: Representación *wavelet* vs representación *contourlet*

De acuerdo a los resultados de estudios previos [DV05], la representación eficiente de una imagen debería satisfacer las siguientes características:

- **Resolución Múltiple:** La representación debe ser una aproximación satisfactoria de la imagen, teniendo en cuenta las resoluciones altas y bajas.
- **Localización:** Los elementos básicos deberán estar localizados en ambos dominios, tanto el espacial como el espectral (de frecuencia).
- **El Muestreo Crítico (*Critical Sampling*):** La representación debe formar un marco o una base con bajo nivel de redundancia.
- **Direccionalidad:** Una buena representación debe tener elementos base en diferentes direcciones.
- **Anisotropía:** Para capturar contornos suaves en las imágenes, la representación debe contener elementos base usando una variedad de formas alargadas con diferentes relaciones de aspecto.

Las transformadas *wavelet* cubren las primeras tres propiedades, las transformadas *ridgelet* las primeras cuatro, y las transformadas *contourlet* las cinco, esto es, cubren todas las propiedades.

Después de definir los modelos estadísticos para los coeficientes *ridgelet* y *contourlet*, se realiza la extracción de características. Para cada sub-banda de la transformada *wavelet* se calculan ocho características estadísticas a partir de coeficientes, así como la predicción de error entre los coeficientes mediante el uso de los modelos estadísticos propuestos.

Por último, se aplica un algoritmo *Sequential Floating Search (SFS)* para selección de características y una *SVM* para la clasificación.

El método basado en *wavelets* considera 216 características útiles sólo para la representación de una dimensión, el enfoque basado en *ridgelets* toma 48 características, y la aproximación de *contourlets* contempla un total de 768 características.

La mejora de los resultados aplicando tanto las transformaciones *ridgelet* como las transformaciones *contourlet* es razonable debido al hecho de que se cuenta con las estadísticas de más de tres direcciones, teniendo en cuenta las cinco propiedades de una representación de imagen eficiente.

Los *contourlets* y *ridgelets* no sólo son efectivos para diferenciar entre modelos de cámaras, sino también para diferenciar entre imágenes producidas por diferentes cámaras o escáneres del mismo modelo. De cualquier manera los autores consideran que podrían implementar mejoras experimentando con diferentes algoritmos de selección como es el caso de [SFFS](#).

Los estudios basados en las técnicas *wavelet* tienen buenos resultados. Sin embargo, los experimentos que se han realizado se enfocan a cámaras digitales tradicionales, dejando a un lado los dispositivos móviles que es uno de los campos que está ganando más terreno cada día como ya se ha mencionado anteriormente. En el Anexo A se describen algunas generalidades de la transformada *wavelet*.

En [\[LLC<sup>+</sup>12\]](#) se propone un método que emplea la densidad marginal de los coeficientes de la Transformada Coseno Discreta o *Discrete Cosine Transform (DCT)* en las coordenadas de frecuencia baja y las características de densidad del vecindario (*neighbouring joint density*) en el domino *DCT*. Adicionalmente, se utiliza la agrupación jerárquica (*hierarchical clustering*) y una máquina de soporte vectorial *SVM* con *kernel RBF* lineal para detectar los teléfonos inteligentes fuente y los procesamientos aplicados a las imágenes. En los experimentos realizados con imágenes de diferentes factores de escala pertenecientes a cinco modelos de teléfonos inteligentes de cuatro fabricantes se obtuvo entre el 86,36 % y el 99,91 % de exactitud, alcanzando mejores resultados con un *kernel* lineal. A pesar de los resultados satisfactorios, esta propuesta puede mejorarse mediante la optimización de los parámetros del *kernel*, el aumento el tamaño del conjunto de imágenes de prueba y la adopción de un algoritmo sofisticado para la selección de las características.

### 3.1.5 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 descritos en la sección [2.1.3.2](#) pueden dejar sobre las imágenes. Estas técnicas se dividen en dos ramas: defectos de píxel y patrón de ruido del sensor *Sensor Pattern Noise (SPN)*. 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. En la segunda se construye un patrón del ruido promediando los múltiples residuos de ruido obtenidos mediante algún filtro de eliminación de ruido. La presencia del patrón se determina utilizando algún método de clasificación como correlación o máquinas *SVM*.

En [\[GBK<sup>+</sup>01\]](#) se estudian los defectos de los píxeles en los sensores de tipo *CCD*, centrándose en la evaluación de diferentes características para examinar las imágenes e identificar la fuente: defectos del sensor *CCD*, formato de los archivos usados, ruido introducido en la imagen y marcas de agua introducidas por el fabricante de la cámara.

Entre los defectos del sensor **CCD** considerados se encuentran los puntos calientes, los píxeles muertos, los defectos en grupo y los defectos de fila o columna. En sus resultados se observa que cada una de las cámaras tiene un patrón de defecto diferente. Sin embargo, también se señala que el número de defectos en los píxeles para una cámara es diferente entre fotos y varía demasiado en función del contenido de la imagen. Asimismo, se revela que el número de defectos cambia con la temperatura. Por último, el estudio encontró que las cámaras con **CCD** de alta calidad no tienen este tipo de problema. También es cierto que la mayoría de las cámaras tienen mecanismos adicionales para compensar este tipo de problemas. Al considerar únicamente los defectos de los sensores de tipo **CCD** este estudio no es aplicable al análisis de imágenes generadas por dispositivos móviles.

En [LFG06] se analiza el patrón de ruido del sensor de un conjunto de cámaras, el cual funciona como una huella dactilar, permitiendo la identificación única de cada cámara. Para obtener este patrón se realiza un promedio del ruido obtenido a partir de diferentes imágenes utilizando un filtro de eliminación de ruido. Para identificar la cámara a partir de una imagen dada, se considera el patrón de referencia como una marca de agua cuya presencia en la imagen es establecida mediante un detector de correlación. El estudio se realizó con 320 imágenes procedentes de 9 modelos distintos de cámaras. También se demuestra que este método está afectado por algoritmos de procesamiento de la imagen como la compresión **JPEG** y la corrección *gamma*. Los resultados para fotografías con diferentes tamaños y recortadas no son satisfactorios [VCEK07].

En [CESR12] se propone un enfoque para la identificación de la cámara fuente considerando escenarios abiertos, donde a diferencia de los escenarios cerrados no se da por sentado contar con acceso a todas las posibles cámaras de origen de la imagen. Esta propuesta comprende tres fases: definición de las regiones de interés, determinación de las características e identificación de la cámara fuente. Las diferentes regiones de las imágenes pueden contener información distinta sobre la huella digital de la cámara fuente. Este enfoque, en contraste con otros, considera diferentes áreas de interés *Region Of Interest (ROI)* y no sólo la región central de la imagen. Para cada imagen se definen nueve **ROIs** como se puede ver en la Figura 3.2.

Se asume que estas regiones coinciden con el eje principal de la lente y, por lo tanto, deben tener más detalles de la escena porque los fotógrafos aficionados por lo general centran el objeto de interés en el centro de la lente. Además, las regiones de la 6 a la 9 proporcionan información importante debido a que las simetrías del sistema de lentes de algunas cámaras generan un efecto de bordes oscuros en las fotografías, lo que implica una caída de la intensidad radial desde el centro de la imagen provocando una pérdida de brillo o saturación en la periferia.

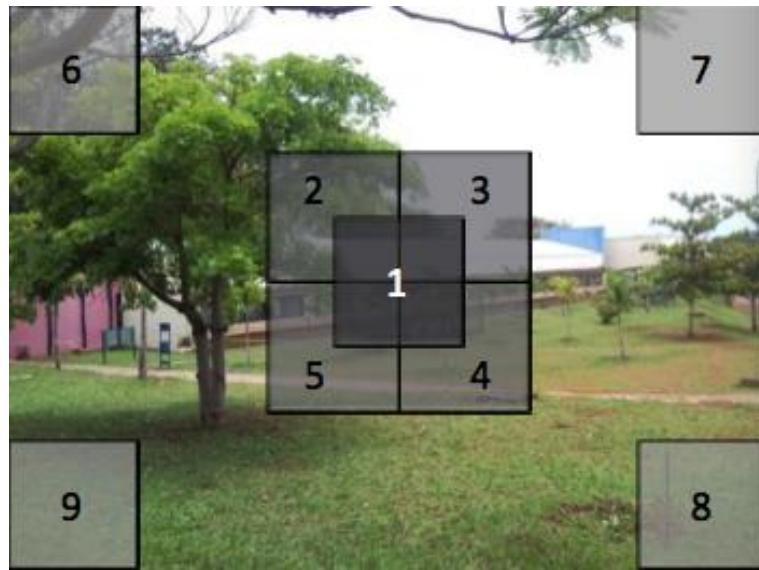


Figura 3.2: Regiones de interés

El uso de las regiones de interés permite trabajar con imágenes de diferentes resoluciones sin la necesidad de llenar con ceros las imágenes y sin el uso de artefactos de interpolación de color. Para determinar las características se calcula el **SPN** para cada uno de los canales R, G y B. Asimismo, se calcula el **SPN** para el canal Y (luminancia), que es una combinación de los tres canales ***Red-Green-Blue (RGB)*** (como una versión en escala de grises de la imagen), generándose un total de 36 características para representar cada imagen. Después, las imágenes tomadas por la cámara bajo investigación son etiquetadas como la clase positiva y las tomadas por las cámaras disponibles restantes como las clases negativas. Después de la fase de entrenamiento de la **SVM** en la que se calcula el hiper-plano que separa los casos positivos y negativos toman en cuenta las clases desconocidas del escenario abierto moviendo el hiper-plano generado por un valor dado ya sea hacia adentro (hacia las clases positivas) o hacia afuera (las clases negativas).

Mediante el movimiento del hiper-plano se puede variar que tan estricto se desea ser para determinar si una imagen pertenece a una clase o no. A este proceso se le denomina modelado de límites de decisión o ***Decision Boundary Carving (DBC)***. En los experimentos se utiliza un conjunto de 25 cámaras digitales de 9 fabricantes, 150 imágenes en formato **JPEG** de cada cámara con diferentes configuraciones de luz, *zoom* y *flash*. Los resultados de los experimentos mostraron una precisión del 94,49 %, del 96,77 % y del 98,10 %, utilizando conjuntos abiertos con 2/25, 5/25, y 15/25 cámaras, respectivamente, definiendo un conjunto abierto x/y como el conjunto de y cámaras donde x cámaras son usadas para entrenar y probar las imágenes que pueden pertenecer a cualquiera de las cámaras x conocidas, así como a las otras y-x cámaras desconocidas.

### 3.1.6 Resumen

En la Tabla 3.1 se muestran los resultados de la evaluación de las diferentes técnicas de identificación de la cámara fuente. La información que no se detallaba en los artículos correspondientes ha sido cumplimentada con las letras ND (No Detallado).

Tabla 3.1: Comparativa sobre las diferentes técnicas de identificación de la cámara fuente

Grupo Técnica	Características de la Imagen		Interpolación de la Matriz CFA		Imperfecciones del Sensor		Transformación Wavelet			
	[TLL07]	[MSGW08]	[LLC+12]	[CAS+06]	[BSM08]	[GBK+01]	[LFG06]	[CESR12]	[MKY08]	[WGKM09]
Técnica Clasificador	SVM	SVM	SVM	SVM	SVM	ND	SVM	SVM	SVM	SVM
Núcleo SVM	Lineal	Lineal	Lineal y No-lineal RBF	Lineal y No-lineal RBF	Lineal y No-lineal RBF	ND	No-lineal RBF	No-lineal RBF	No-lineal RBF	No-lineal RBF
Número de fabricantes	2	5	4	3	5	1	5	9	4	3
Número de modelos	4	5	5	9	2	2	9	25	6	6
Número de imágenes por cámara	150	100	599	200	600	ND	320	50	350	350
Resoluciones	1600 x 1200	ND	Diferentes	Diferentes	Diferentes	640 x 480	Diferentes	Diferentes	Diferentes	ND
Formato	JPEG	JPEG	JPEG	ND	JPEG	ND	JPEG	JPEG	JPEG	ND
Aplicado a móviles	Sí	Sí	Sí	Sí	No	No	Sólo 2 entre 25 cámaras	No	No	No
Aplicado a diferentes modelos del mismo fabricante	Sí	No	Sí	Sí	Sí	ND	Sí	Sí	Sí	No

## 3.2 Ataques al Análisis Forense de Imágenes

En comparación con el destacado papel de las imágenes digitales en la sociedad multimedia de hoy en día, la investigación en el campo de la autenticidad de la imagen se encuentra todavía en un estadio muy preliminar. La mayoría de las publicaciones en este campo emergente todavía carece de discusiones rigurosas y robustas contra los falsificadores estratégicos, que prevén la existencia de técnicas forenses [GKWB07].

El área que se encarga de estudiar ataques a las técnicas de análisis forense de imágenes es conocida como *counter-forensics*. Los ataques contra los algoritmos forenses de imágenes digitales son aquellas técnicas cuyo objetivo es confundir sistemáticamente a los procedimientos de identificación de la fuente de la imagen o de detección de manipulaciones maliciosas en las imágenes. Estos ataques pueden tener uno de los siguientes objetivos:

1. Camuflaje de post-procesamientos maliciosos sobre la imagen.
2. Destrucción de la identificación correcta del origen de la imagen.
3. Falsificación del origen de imagen.

El análisis forense de imágenes digitales se ha convertido en un tema de interés en los últimos años. En sus inicios la parte académica encontró utilidad del análisis forense de imágenes en ámbitos como aplicación de la ley, inteligencia, investigaciones privadas y medios de comunicación. El análisis forense surge con la idea de restablecer la confiabilidad en las imágenes digitales que de otro modo se consideraban muy fácilmente modificables. Como en la mayoría de campos de estudio existe una contracorriente, en este caso, personas como espías o estafadores hacen esfuerzos para manipular las imágenes en su propio beneficio usando el conocimiento del análisis forense de imágenes para borrar o incluso suplantar las huellas o rastros que se utilizan para determinar la identidad de las imágenes. Muchos de los algoritmos forenses existentes en la literatura no fueron diseñados teniendo en cuenta ese tipo de comportamiento y como consecuencia son fáciles de engañar.

La posibilidad de copiar las huellas digitales de una imagen se puede convertir en un ciclo infinito que puede permitir que personas inocentes sean inculpadas, también que criminales aseguren que las pruebas son resultados de una falsificación. Al final, la confianza en las técnicas forenses de imágenes se podría ver comprometida. Es por esto que surge la necesidad considerar los posibles ataques en el momento de diseñar técnicas de análisis forense en imágenes digitales.

Así como en el área de seguridad el estudio de los ataques permite mejorarla, los métodos forenses de imágenes se pueden beneficiar del estudio de las técnicas de ataque para robustecer los algoritmos de las próximas generaciones.

### 3.2.1 El Camuflaje de Post-Procesamientos

Estas técnicas tienen como objetivo ocultar la existencia de algún proceso aplicado a una imagen analizando los rasgos que éstos dejan sobre la imagen durante su aplicación

para así poder contrarrestarlos.

Entre las investigaciones realizadas sobre los rasgos de los algoritmos en las imágenes se encuentran el estudio de las dependencias introducidas durante el re-dimensionamiento o la rotación de las imágenes [PF05], el estudio de los coeficientes estadísticos de los JPEG para detectar la re-compresión [LF03], y el análisis de la fase de congruencia para detectar la composición de imágenes a través del recortado y pegado de diferentes imágenes [CSS07].

Para exemplificar este tipo de técnicas se describe a continuación la propuesta presentada en [GKWB07] para ocultar el proceso de re-muestreo (*resampling*).

El re-muestreo es el redimensionamiento con interpolación de las imágenes. Este proceso es muy común en las operaciones primitivas de imágenes como escalamiento y rotación.

Los algoritmos detectores de re-muestreo se basan en la búsqueda de las dependencias sistemáticas y periódicas entre píxeles vecinos insertadas cuando se aplica la operación de re-muestreo. Esta periodicidad se debe a la matriz de muestreo equidistante utilizada para realizar dicha operación [PF05].

Para ocultar el re-muestreo es necesario romper las equidistantias periódicas introduciendo distorsiones geométricas conocidas como ataques de marca de agua. En este caso se superpone un vector de distorsión aleatoria a las posiciones de cada píxel donde un parámetro determina el grado de distorsión introducido. Para evitar generar características visibles en la imagen como ruido se debe modular la fuerza de la distorsión empleando dos detectores de bordes: uno en dirección vertical y otro en dirección horizontal.

El ataque para generar una imagen  $\tilde{y}$  a partir de la imagen  $x$  aplicando el re-muestreo sin dejar rastro de dicha operación consiste en los siguientes pasos:

1. *Calcular el componente de baja frecuencia: A la imagen  $x$  de entrada se le aplica el re-muestreo y a este resultado se le aplica el filtro de la mediana.*
2. *Calcular el componente de alta frecuencia: Restar a la imagen  $x$  el resultado de aplicarle el filtro de la mediana, aplicar a este resultado el re-muestreo con una distorsión geométrica y una modulación de los bordes; la información de los bordes es extraída de la imagen con muestreo aplicado del paso 1 antes de aplicar el filtro de la mediana.*
3. *Obtener la imagen final  $\tilde{y}$  sumando los resultados del paso 1 y 2.*

En la Figura 3.3 se puede observar el diagrama de bloques del ataque.

Los resultados que obtuvieron en los experimentos realizados con este ataque apuntan a que es una propuesta prometedora ya que tiene una tasa de falsos positivos *False Acceptance Rate (FAR)* inferior al 1%.

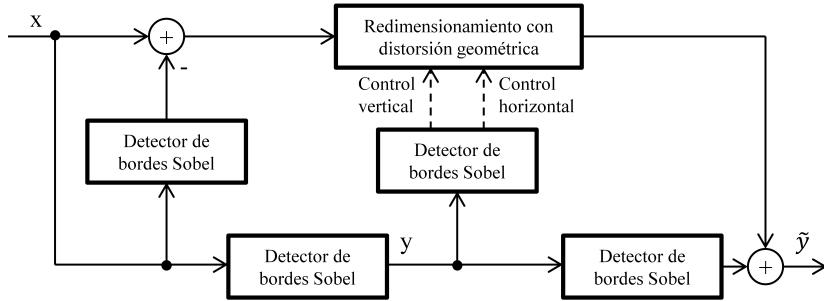


Figura 3.3: Diagrama de bloques enfoque de doble vía para ocultar re-muestreo

### 3.2.2 Manipulación de la Identificación de la Fuente

Así como para el proceso de identificación de la fuente se usa la extracción del ruido del sensor en la imagen, un contraataque lógico para esta técnica consta en la eliminación del ruido del sensor. Dando un paso más adelante se puede pensar también en la posibilidad de eliminar el ruido del sensor de la imagen y sustituirlo por el ruido del sensor que le pertenezca a otra cámara.

#### 3.2.2.1 Destrucción de la Identidad de una Imagen

En [GKWB07] se demostró que la resta de las características del dominio *wavelet* de las imágenes no es suficiente para eliminar el ruido de una imagen, además de que este procedimiento deja rastros visibles sobre la imagen. Existe otro método bastante conocido para la eliminación del ruido de una imagen llamado corrección de sensibilidad o *flatfielding*. Este método es usado típicamente en astronomía o en el proceso de escaneado de planos para mejorar la calidad de las imágenes.

La corrección de sensibilidad se realiza en base a los principales componentes del ruido de la imagen: el ruido de patrón fijo **FPN** y el ruido de respuesta no uniforme **PRNU**.

El ruido **FPN** se calcula con la ecuación 3.1 en términos de un marco oscuro  $d$  promediando  $K$  imágenes  $x_{oscura}$  capturadas en un ambiente completamente oscuro que se puede emular cubriendo completamente el lente de la cámara:

$$d = \frac{1}{K} \sum x_{oscura} \quad (3.1)$$

El ruido **PRNU** se calcula con la ecuación 3.2 en términos de un marco plano (*flatfield*)  $f$  promediando  $L$  imágenes  $x_{iluminada}$  de una escena iluminada homogéneamente. A las  $L$  imágenes se les elimina el ruido **FPN** mediante la resta del marco oscuro  $d$  antes de promediarlas.

$$f = \frac{1}{L} \sum_L (x_{iluminada} - d) \quad (3.2)$$

Como se describe en [LFG06, GKWB07], los atacantes pueden intentar evitar la identificación correcta de la fuente ya que existe la posibilidad de eliminar y extraer la huella de una imagen. La destrucción de la huella de una imagen  $x$  generada con una cámara específica se realiza con la ecuación 3.3 restando a la imagen original  $x$  el marco oscuro  $d$  y dividiendo el resultado de la resta entre el marco plano  $f$ .

$$\tilde{x} = \frac{x - d}{f} \quad (3.3)$$

A pesar que los resultados obtenidos con esta técnica son buenos, se presentan algunos inconvenientes:

- Llevar a cabo una corrección de sensibilidades perfecta en un gran número de fotos es difícil ya que los parámetros para calcular el PRNU y el FPN deben coincidir con los de la imagen a atacar.
- En la propuesta se asume que el atacante puede tener acceso a la cámara fuente de la imagen  $x$  para generar los marcos oscuros y planos y éste no es un escenario próximo a la realidad.

Existen otras posibilidades menos robustas para destruir la identidad que en ciertos casos podrían ser efectivas ya que no necesitan contar con imágenes procedentes de la cámara origen para generar el marco oscuro y el marco plano, pero a cambio de esta facilidad la calidad de la imagen puede verse reducida y podrían introducirse algunos rasgos visuales. Por ejemplo, es posible rotar la imagen unos pocos grados, escalar la imagen, o aplicar un filtro de desenfoque gaussiano.

### 3.2.2.2 Falsificación de la Identidad de una Imagen

De igual forma que se puede eliminar el ruido en una imagen haciendo uso de la técnica de corrección de sensibilidad, se puede inyectar el ruido de la imagen de otra cámara diferente mediante la corrección de sensibilidad inversa con la ecuación 3.4 [GKWB07].

$$\tilde{y} = \tilde{x} \cdot f_{falsa} + d_{falsa} \quad (3.4)$$

Donde,  $f_{falsa}$  y  $d_{falsa}$  corresponden a la cámara que se pretende plagiar y  $\tilde{x}$  es la imagen original sin ruido.

En [SLFK10] se propone un algoritmo para falsificar la identidad de una cámara, a continuación se describen los pasos a seguir:

1. *Calcular el promedio de las huellas  $F(\mathbf{C1})$  de la cámara  $\mathbf{C1}$  con la que se atacará..*
2. *Tomar una fotografía  $\mathbf{P}$  con la segunda cámara  $\mathbf{C2}$ .*
3. *Sumar  $F(\mathbf{C1})$  a la fotografía  $\mathbf{P}$ .*

En el caso de que las dimensiones de  $F(\mathbf{C}1)$  y  $\mathbf{P}$  no coincidan, es necesario aplicar un recorte o una reconstrucción para igualar el tamaño de las imágenes.

También se propone una mejora al algoritmo de falsificación anterior para enmascarar los rasgos de la cámara  $\mathbf{C}2$ . Esta técnica se presenta en el siguiente algoritmo.

1. *Calcular el promedio de las huellas  $F(\mathbf{C}1)$  de la cámara  $\mathbf{C}1$  con la que se atacará.*
2. *Calcular el promedio de las huellas  $F(\mathbf{C}2)$  de la cámara  $\mathbf{C}2$ .*
3. *Tomar una fotografía  $\mathbf{P}$  con la cámara  $\mathbf{C}2$ .*
4. *Restar  $F(\mathbf{C}2)$  a  $\mathbf{P}$ .*
5. *Sumar  $F(\mathbf{C}1)$  a la fotografía  $\mathbf{P}$ .*

Al restar  $F(\mathbf{C}2)$  se trata de eliminar la correlación entre la fotografía  $\mathbf{P}$  y la cámara  $\mathbf{C}2$ .

### 3.2.3 Detección de Falsificación de la Identidad de una Imagen

Una vez estudiada la técnica de falsificación de la huella de una imagen y la inyección de ésta en otra imagen, surge la pregunta de si es posible detectar cuándo se ha sufrido un ataque de este tipo. La respuesta a esta pregunta es sí, y se puede lograr mediante el análisis de las diferencias entre las propiedades de un patrón de ruido copiado [SLFK10, GFC11].

Para explicar la detección de la falsificación de la identidad de una imagen se presenta el siguiente escenario de ataque:

1. *Alicia, la víctima, sube algunas fotografías adquiridas con su cámara  $\mathbf{C}$  a una red social en internet.*
2. *Eva, la atacante, obtiene  $\mathbf{N}$  de esas fotografías y calcula la huella  $K'_E$  perteneciente a la cámara de **Alicia**.*
3. *Eva implanta la huella  $K'_E$  en una imagen  $\mathbf{J}$  tomada con una cámara  $\mathbf{C}'$  con el propósito de hacer parecer que **Alicia** fue la autora de esa imagen falsificada  $\mathbf{J}'$ .*

Entonces, surge la siguiente pregunta: ¿La imagen  $\mathbf{J}'$  fue falsificada? Para resolver esta pregunta Alicia en su defensa puede hacer uso de su cámara  $\mathbf{C}$  y un conjunto  $\mathbf{F}$  de fotografías compuesto por las fotografías que Eva robó más algunas extras que le pertenezcan. El escenario de defensa de Alicia propuesto en [GFC11] sería:

1. *Calcular la huella de su cámara  $K'_A$  utilizando imágenes planas inocentes que no hayan sido manipuladas por **Eva** (para obtener una mejor estimación de la huella).*
2. *Calcular el ruido residual  $PRNU$   $W_J'$  de la imagen  $\mathbf{J}'$ .*

3. Calcular el ruido residual *PRNU*  $W_I$  de una de las imágenes  $I$  utilizadas por *Eva* para realizar el ataque. Debido a que  $W_I$  participó en el cálculo de  $K'_E$  contiene una versión escalada del ruido residual total compartiendo rasgos característicos con  $W_{J'}$ .

4. *Alicia* calcula las correlaciones:

- $C_{I,J'} = \text{corr}(W_I, W_{J'})$
- $C_{I,K'A} = \text{corr}(W_I, K'_A)$
- $C_{J',K'A} = \text{corr}(W_{J'}, K'_A)$

5. Evaluar si  $J'$  fue atacada realizando la prueba del triángulo como se muestra en la Figura 3.4 con las correlaciones calculadas en el paso 4.

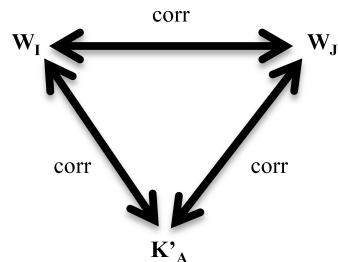


Figura 3.4: Correlaciones de la prueba del triángulo

Esta prueba se basa en el hecho de que el valor de correlación  $C_{I,J'}$  de las imágenes  $I$  que no fueron utilizadas para falsificar  $J'$  puede ser estimado de las correlaciones  $C_{I,K'A}$  y  $C_{J',K'A}$ . En caso de que la imagen  $I$  haya sido utilizada para la falsificación la correlación  $C_{I,J'}$  tendría un valor mayor que  $C_{I,K'A}$  y  $C_{J',K'A}$ .

Por supuesto este contraataque también es vulnerable ya que al poder distinguir entre las huellas originales y las que son copias los atacantes podrían comenzar a generar imágenes ilegales con características que se ajusten a las de las fotografías encontradas de la persona a la que se pretende atacar. Sin embargo, esta propuesta incrementa los esfuerzos requeridos para culpar a alguien inocente.

