4. ANÁLISIS FORENSE EN VÍDEOS DIGITALES DE DISPOSITIVOS MÓVILES

Este capítulo presenta el estado del arte de las técnicas de análisis forense de vídeos clasificando los trabajos relacionados en los dos grandes grupos: técnicas de análisis forense de vídeos clasificados según su objetivo y técnicas anti-forenses en vídeos. Se comienza con la exposición de los trabajos relacionados referentes a la identificación de la fuente de adquisición de vídeos digitales. Seguidamente se muestran los trabajos relacionados referentes al análisis de compresión de un vídeo la identificación de la fuente de adquisición de vídeos digitales. Posteriormente se exponen los trabajos referentes al análisis de manipulación de un vídeo. Finalmente, se presentan las técnicas anti-forenses en vídeos digitales. Cabe destacar que aunque este trabajo este centrado en los dispositivos móviles, en el estado del arte se aúnan referencias a técnicas sobre vídeos de todo tipo de dispositivos, ya que su conocimiento puede ser válido para la aplicación o adaptación a vídeos de dispositivos móviles.

4.1. Importancia del Análisis Forense

La necesidad de realizar un análisis forense en dispositivos móviles surge a partir de las bondades y características tecnológicas que pueden ofrecer este tipo de dispositivos. Por ejemplo, la facilidad de implementar aplicaciones para estos dispositivos en poco tiempo [AZ06], la capacidad de almacenar, editar, eliminar e imprimir documentos electrónicos, el uso de mensajes de texto, mensajes multimedia y conversaciones a través de aplicaciones de redes sociales (lo más requerido en la actualidad), el uso de plataformas en línea para realizar operaciones bancarias, compras a través de la web y todas aquellas operaciones realizadas con datos sensibles. Por tanto, hoy en día, los dispositivos móviles son elementos que proveen información útil que deberá tratarse con cautela y precisión.

En cuanto al análisis forense en vídeos de dispositivos móviles no hay duda de la importancia que puede tener su aplicación en casos judiciales. Los dispositivos móviles pueden contener vídeos almacenados de carácter personal o con contenidos de delitos flagrantes. Todos estos tipos de vídeos pueden ser evidencias de un hecho y elementos potenciales de uso en procesos judiciales y, consecuentemente, elementos de estudio del análisis forense.

4.2. Técnicas de Análisis Forense

Según [BFM+12], las técnicas forenses en vídeos se agrupan en herramientas forenses según el objetivo a cumplir: herramientas forenses para el análisis de adquisición, compresión y falsificación de un vídeo.

4.2.1. Herramientas Forenses para el Análisis de Adquisición

El análisis de adquisición de imágenes es uno de los primeros problemas que surgieron en la ciencia forense multimedia, que tiene como objetivo identificar la fuente u origen de una imagen. Este análisis captura información como tipo de dispositivo (cámara, escáner, ordenador, etc.), marca y modelo del dispositivo usado. A la fecha las investigaciones sobre técnicas de análisis de adquisición en imágenes se han realizado a más profundidad en comparación que las de vídeo digital que no alcanza aún un estado de madurez.

La mayoría de las técnicas que se pueden aplicar a una imagen se pueden emplear con los diferentes fotogramas de un vídeo. En [SOAGRC+13] se realiza una comparación detallada de los principales grupos de técnicas de identificación de fuente de adquisición. Estas se dividen en cinco grupos y están basadas en: metadatos, características de la imagen, defectos de la matriz CFA e interpolación cromática, imperfecciones del sensor y las transformadas *wavelet*.

En cuanto a la técnica basada en metadatos, es la más sencilla de analizar, aunque depende en gran medida de los datos que inserta el fabricante.

Asimismo la agregación de metadatos a la imagen no es obligatoria. En [RCC+08][BL05][Tes05][BL04] se utilizan los metadatos con fines de clasificación de imágenes digitales.

Estas técnicas utilizan un conjunto de características extraídas del contenido de la imagen para hacer identificar la fuente. Estas características se dividen en tres grupos: características de color, métricas de calidad de la imagen o *Image Quality Metrics* (IQM) y estadísticas del dominio *wavelet*. En [TLL07] se identifica la fuente utilizando tres tipos de características de la imagen: características de color, características de calidad o IQM y características de la imagen en el dominio de la frecuencia. La clasificación de las imágenes se realiza a través de una Maquina de Soporte Vectorial o SVM. El resultado obtenido para una clasificación de cuatro cámaras de dos fabricantes distintos con contenidos similares en la imágenes fue del 100%, mientras que para la clasificación de imágenes con contenidos distintos entre sí fue 93.05%.

En cuanto a la técnica de defectos de la matriz CFA y la especificación de los algoritmos de interpolación cromática algunos autores precisan que generan ciertas diferencias marcadas entre los diferentes modelos de cámaras [BSM06] [CAS+06] [LH06] [BSM08a]. En [CAS+06] [CAS+06] se utiliza una técnica que se basa en los algoritmos propietarios de interpolación cromática, los cuales dejan correlaciones a través de los planos de bits adyacentes de una imagen. Estos pueden ser representados mediante un conjunto de 108 métricas de similitud binarias y 10 métricas de calidad de la imagen IQM. Con un clasificador *k-Nearest Neighbors* (KNN) se realizan experimentos utilizando 9 cámaras de teléfonos móviles y 200 fotos de cada una. Para el entrenamiento se utilizaron 100 fotos de cada cámara y las 100 restantes para las pruebas. Se obtuvo un rendimiento promedio del 93,4% de 16 experimentos que se realizaron. Hay diversos grupos de investigación que han aportado en esta área, en donde se presentan buenos resultados, por ejemplo en [LH06] [BSM08a].

Dentro de los métodos existentes que se basan en las imperfecciones del sensor, hay dos grandes ramas de las cuales se pueden trabajar: defectos del pixel o patrón de ruido del sensor. En [LFG06] se demostró que los sensores de las cámaras generan un patrón de ruido *Sensor Pattern Noise* (SPN) que podría ser utilizado como método único de identificación. En [Li10] se demostró que el ruido del sensor extraído de las imágenes podría ser severamente contaminado por los detalles de las escenas concretas. Para lidiar con ese problema, se propuso un nuevo enfoque para la atenuar la influencia del detalle de las escenas en el ruido del sensor mejorando la tasa de acierto. En los experimentos se tomaron 9 cámaras y 320 fotos de cada una, variando las escenas al aire libre e interiores. En [GBK+01][VCEK07][CESR12] se presentan otros métodos de identificación de fuente basados en las imperfecciones del sensor.

Por último, en el área de la transformada *wavelet* existen diversos enfoques. Por ejemplo en [WHL12] se propone una nueva técnica de identificación basada en las características de probabilidad condicional. Este tipo de características fueron propuestas inicialmente para detectar mensajes ocultos en imágenes [WBSH09]. Se obtuvieron unos resultados del 98,6%, 97,8% y 92,5% de acierto en la clasificación de 2, 3 y 4 IPhones respectivamente con un recorte de imagen de 800x600. En [RCAGSO+13] se determina que el uso del patrón de ruido del sensor conjuntamente con la transformada *wavelet* es un método efectivo para la identificación de fuente, alcanzando una tasa de éxito promedio del 87,21%.

En el caso del desarrollo de técnicas para la identificación de fuente de vídeo, existen pocas referencias al respecto. Algunas se basan directamente en la secuencia de codificación y otras en la extracción de fotogramas aplicando algún método de clasificación para imágenes fijas.

En [SXD09], se propone un algoritmo en base a la información del vector de movimiento en el flujo codificado. En los experimentos realizados se utilizaron 100 secuencias de vídeo, de las cuales 20 de ellas procedentes de *Video Quality*

Experts Group (VQEG) y 80 de disco versátil digital o *Digital Versatile Disc* (DVD). Todos los vídeos fueron codificados por diferentes aplicaciones de edición de vídeo conocidos. Mediante un experimento se obtuvo un 74,63% de precisión en la identificación del software que se utilizó en la codificación.

En [YHW12] propone un método de identificación utilizando los fotogramas extraídos de vídeos. Las características de probabilidad condicional se extraen directamente de los fotogramas del vídeo. En las pruebas realizadas se utilizaron 4 modelos diferentes de cámaras y un clasificador SVM, obteniendo, en un primer experimento aplicado en el dominio del espacio con los valores de luminancia, un 82,6% de precisión. En un segundo experimento usando el mismo conjunto de vídeos, tomando el valor de luminancia, el promedio de clasificación fue de 100%. En un tercer experimento en donde se utilizaron un conjunto de vídeos con mayores cambios en las escenas se obtuvo un 97,2% de acierto.

Un problema importante en la protección de los derechos de autor es la proliferación de vídeos piratas. En muchos casos las copias ilegales de películas se publican en Internet incluso antes de su lanzamiento oficial. Una gran parte de estas copias falsas se producen mediante la grabación de películas con cámaras de vídeo en salas de cine. Las técnicas que forenses que contribuyen a hacer frente a estos problemas son: las técnicas de detección de re-adquisición y las técnicas de detección de copia.

Las técnicas de detección de re-adquisición se usan en vídeos capturados a partir de una secuencia de vídeo que se reproduce en una proyección. Para hacer frente a la re-adquisición se proponen diversos enfoques basados en marca de agua, tanto para la identificación del vídeo pirata [LKL+08] como para localizar la posición del pirata en salas de cine [LKL12]. En [WF08] se muestran muchos experimentos realizados dando buenos resultados; estos vídeos re-adquiridos se detectaron con un 88% de precisión y con un 0,4% de

probabilidad de falsos positivos.

En cuanto a la de detección de copia del vídeo, es común extraer características más destacadas del contenido visual que no dependen del dispositivo utilizado para capturar el vídeo. Sin embargo, en [BSM08b] se señala que las firmas sólidas basadas en contenidos pueden obstaculizar la capacidad de distinguir entre los vídeos que son similares, aunque no son copias de uno al otro. Por esta razón, se propone utilizar las características del dispositivo fuente extraídas de los vídeos para la construcción de una técnica de detección de copia. En [BSM08b], una firma de vídeo se obtiene mediante la estimación de las huellas digitales PRNU de videocámaras involucradas en la generación del vídeo. Como consecuencia, esto produce automáticamente una media ponderada de los diferentes patrones PRNU, en el que el mayor número de fotogramas tomados con la misma cámara se traducirá en un mayor peso asignado a la misma. Los autores también muestran que la huella digital es robusta frente a un conjunto de operaciones de tratamiento común, es decir, la mejora del contraste, desenfoque, irregularidades de la imagen, subtítulos, ajuste de brillo, de compresión, etc. Los experimentos realizados en vídeos descargados de YouTube muestran una tasa de detección del 96% y una probabilidad de falsa alarma de 5%.

4.2.2. Herramientas Forenses para el Análisis de Compresión

El contenido del vídeo suele estar disponible en un formato de compresión con pérdidas debido a la velocidad de bits de gran tamaño y es necesario representar las imágenes en movimiento, ya sea en un formato no comprimido o sin pérdidas. La compresión con pérdida deja huellas características, que pueden ser detectados por el analista forense. Las arquitecturas de codificación del vídeo son más complejas que las adoptadas para imágenes. La mayoría de los estándares de codificación ampliamente utilizados como MPEG o las familias H.26x, heredan el uso de codificación en bloque por la transformada de

la norma JPEG. Sin embargo, la arquitectura de codificación es complicada por varios procesos adicionales, por ejemplo, la predicción espacial y temporal. Cuando cada trama se considera como una sola imagen, es posible aplicar técnicas de análisis forense basadas en las imágenes. Sin embargo, para permitir un análisis más completo, es necesario tener en cuenta las operaciones de codificación a lo largo de la dimensión temporal. A continuación, se detallan técnicas forenses destinadas a la codificación del contenido del vídeo.

La elección de los parámetros de codificación depende de la implementación específica del códec y de las características de la señal codificada. En la compresión de vídeo, el número de parámetros de codificación que se pueden ajustar es significativamente amplio. Como consecuencia de ello, el analista forense debe tener en cuenta un mayor grado de libertad cuando se detecta la identidad del códec. Esta información podría permitir identificar las implementaciones de otros proveedores que dependen de los códecs de vídeo. Los métodos con el objetivo de estimar diferentes parámetros de codificación y elementos de sintaxis que caracterizan a cada códec se pueden agrupar en tres categorías principales: detección de bloques, detección de parámetros de cuantificación e identificación de vectores de movimiento.

En [HS09] se calcula el tamaño de bloque en una secuencia de vídeo comprimida mediante el análisis de la imagen reconstruida en el dominio de la frecuencia y la detección de los picos que están relacionados con las discontinuidades en los límites de bloque, en lugar de las características intrínsecas de la imagen subyacente.

Cuando un vídeo ha sido comprimido y posteriormente se edita la escala, brillo, etc., el vídeo en mención es re-comprimido. Las técnicas para detectar las huellas dejadas por la doble compresión de vídeo se han enfocado en el estándar MPEG y explotan las mismas ideas usadas para la doble compresión JPEG. En [LWH08] se propone un método para la detección de doble

compresión de MPEG basada en el bloqueo de los artefactos. Se define un sistema de medición para calcular el *Block Artifact Strength* (BAS) de cada trama. La media BAS se calcula para la eliminación de secuencias obtenidas a partir de 1 a 11 tramas, obteniendo un vector de características de los valores BAS. Si la secuencia se ha manipulado previamente, sea eliminación de fotograma o re-compresión, el vector de características presenta un comportamiento característico.

Aunque la detección de doble compresión de las imágenes es un tema ampliamente investigado, la compresión doble de vídeo aún es un problema de investigación actual, debido a la complejidad y diversidad de arquitecturas de codificación de vídeo.

Cuando se transmite un vídeo a través de un canal ruidoso, éste deja huellas en el contenido del vídeo. Un claro ejemplo son las pérdidas de paquetes y errores que pueden afectar el flujo de bits que se recibe. Por tanto, algunos datos codificados harán falta o en su defecto estarán dañados.

En [RP07], los autores presentan un algoritmo basado en varias métricas de evaluación de calidad para estimar la pérdida de paquetes en el vídeo reconstruido. Sin embargo, la solución propuesta adopta métricas de calidad de referencia completa que requieren la disponibilidad de la secuencia de vídeo original sin comprimir. La solución propuesta en [GMT10] se basa en la estimación de calidad no-referenciada pero se lleva a cabo sin tener en cuenta la disponibilidad de la corriente de bits. Por lo tanto, la solución propuesta procesa sólo los valores de los píxeles, identificando las rebanadas que se perdieron del vídeo y produciendo un valor que presenta una correlación con el valor *Mean Square Error* (MSE). El método supone que las rebanadas se corresponden con las filas de macrobloques.

4.2.3. Herramientas Forenses para el Análisis de Manipulaciones

A pesar de ser más complicado que para las imágenes, la creación de un vídeo falsificado o manipulado actualmente es más fácil que antes, debido a la disponibilidad de los medios. Existen muchas maneras diferentes de manipulación de un vídeo como la sustitución o eliminación de algunos fotogramas, la replicación de un conjunto de fotogramas y la inserción o eliminación de objetos de la escena. A continuación se presentan algunas técnicas para detectar la falsificación o manipulación de los vídeos.

Los dispositivos móviles suelen dejar una huella característica en los vídeos grabados. Aunque estas huellas normalmente son explotadas sólo para la identificación de dispositivos, algunas investigaciones como [MCP+07] [HHLH08] [KTY10] han sido direccionadas para detectar las manipulaciones.

En [MCP+07] se propone una técnica basada a PRNU, en concreto a las secuencias de vídeo. El patrón PRNU característico de los dispositivos se estima en los primeros fotogramas de vídeo, y se utiliza para detectar varios tipos de ataques. Los autores evalúan el coeficiente de tres correlaciones y cada uno de estos coeficientes de correlación es el umbral para obtener un evento binario, y diferentes combinaciones de eventos permiten detectar diferentes tipos de manipulación, entre los cuales están la inserción de fotogramas, inserción de objetos dentro de un fotograma (cortar y pegar), la replicación de fotogramas, etc. Los experimentos se realizaron en vídeos MPEG sin comprimir; los resultados muestran que el método es fiable aunque solo informa de algunos estudios de casos, no los valores promediados.

La codificación del vídeo introduce artefactos que pueden ser aprovechados para investigar la integridad del contenido. Estos artefactos son notables en el contenido. En los últimos años, algunos analistas forenses investigan la presencia o inconsistencia de estos artefactos para evaluar la integridad de un vídeo y para localizar las regiones que no son originales. En [WF06], se exploran

dos fenómenos en vídeos MPEG comprimidos, uno estático es decir entre fotogramas y una temporal es decir intra-fotogramas. El fenómeno estático se basa en el hecho de que un vídeo MPEG manipulado casi seguro es comprimido dos veces, el primero que se realiza cuando se crea el vídeo, y la segunda cuando el vídeo se vuelve a guarda después de ser manipulada. Los fenómenos temporales se basan en la estructura GOP de archivos MPEG. Cuando un vídeo se vuelve a comprimirse después de la eliminación o la adición de un grupo de fotogramas, se producirá una desincronización en el patrón de GOP. La literatura indica que aún quedan muchos aspectos por descubrir en la detección de manipulación basada a la codificación de vídeos. Esto se debe a que los algoritmos de codificación de vídeo son mucho más complejos que la compresión JPEG. Esto hace que la detección de artefactos introducidos sea más difícil, ya que los modelos matemáticos no son fáciles de obtener.

Como se ha podido evidenciar es relativamente difícil entender la geometría y las propiedades físicas de una escena cuando es consistente o no. Para un analista forense sería más asequible comprobar las consistencias en una imagen. Lo contrario ocurre al comprobar las consistencias geométricas de un vídeo que contiene una gran cantidad de fotogramas. En [COF12] se propone un algoritmo para detectar trayectorias físicamente inverosímiles de objetos en secuencias de vídeo. La idea es modelar la trayectoria parabólica tridimensional de objetos en vuelo libre, es decir, como cuando una pelota vuela o desplaza hacia la canasta y la correspondiente proyección bidimensional en el plano de la imagen. El objeto volador se extrae del vídeo, compensando el movimiento de la cámara si es necesario. Después el movimiento en el espacio 3D se estima a partir marcos de 2D y se compara con una trayectoria satisfactoria. Si la desviación entre las trayectorias observada y esperada es grande, el objeto se clasifica como manipulado. Aunque el análisis es un escenario muy específico, el método hereda todas las ventajas que caracterizan técnicas forenses basadas

en aspectos físicos y geométricos; por ejemplo, el rendimiento no depende de la compresión y la calidad de vídeo.

Los ataques de copiar - mover en un vídeo afectan a los niveles intra e inter-fotograma. Un ataque de copiar - mover intra-fotograma es conceptualmente idéntico a las de las imágenes, y consiste en la replicación de una parte del fotograma en el propio fotograma, cuyo objetivo normalmente es ocultar o reproducir un objeto. Un ataque de copiar - mover inter-fotograma, consiste en la sustitución de algunos fotogramas con una copia de los anteriores, por lo general para ocultar algo entró en la escena en el vídeo original. En [WF07] se realiza la detección el procedimiento de copiar-mover en un vídeo. Para ello el vídeo se divide en sub partes y se calculan diferentes tipos de coeficientes de correlación con el fin de resaltar las similitudes entre las diferentes partes de la secuencia. También se efectúa un método para detectar la duplicación de regiones, tanto para la inter-fotograma e intra-fotograma. Los resultados logran una precisión superior al 90% para una cámara fija, y 80% para una cámara en movimiento.

4.3. Técnicas Anti-Forenses

En la actualidad un vídeo puede ser fácilmente manipulado por un atacante gracias a la infinidad de técnicas y software de edición existentes. Para hacer frente a este problema surgen las estrategias forenses enfocadas a los vídeos para detectar el origen, la manipulación y verificación de la autenticidad. Pero cuando el atacante tiene mayor conocimiento y un objetivo concreto, puede usar y crear técnicas anti-forenses sofisticadas para manipular vídeos sin dejar rastro alguno, es decir ocultar la huella del procedimiento realizado sobre el vídeo con la finalidad de engañar al analista forense y conllevar a tomar decisiones no acertadas. En este sentido, la ciencia forense debe hacer frente a las técnicas anti-forenses. Como se comentó en la sección 1.1, la forma de combatir a las técnicas anti-forenses es conocerlas en profundidad. Sin

embargo, esto se logra implementando dichas técnicas, y así aplicar medidas de prevención a la hora de desarrollar técnicas capaces de detectar este tipo de operaciones anti-forenses.

Las investigaciones realizadas sobre el análisis anti-forense se ha realizado en gran parte sobre imágenes fijas, pero alguna de las técnicas se pueden aprovechar para el análisis de los vídeos a partir de la extracción de fotogramas. Las técnicas anti-forenses dirigidas a vídeos son relativamente escasas por diversos factores como la complejidad y diversidad de arquitecturas de codificación del vídeo.

En cuanto a técnicas anti-forenses dirigidas a un vídeo, en [PPML15] se estudian técnicas forenses y anti-forenses aplicables a los procesos de edición de vídeos con codificación H.264/AVC. El trabajo se basa en el hecho de que la excesiva predicción residual en la codificación puede aparecer en los fotogramas que se han codificado con predicción intra e inter. En el primer método implementado los autores evalúan el resultado residual después del filtrado de desbloqueo para revelar las operaciones de edición realizadas. En el segundo, se utiliza un mecanismo de control de frecuencia para comprobar los parámetros de cuantificación. En [SRL11] se propone una técnica anti-forense para suprimir la huella digital temporal que surge en secuencias de vídeo MPEG cuando se agregan o eliminan fotogramas después de la re-compresión. Para ejecutar esta técnica inicialmente los autores identificaron las propiedades de la huella digital temporal. Estas propiedades fueron utilizadas para modelar el efecto de supresión y adición de fotogramas. La técnica funciona aumentando del error de predicción en ciertos *P-frames* del vídeo, mediante el establecimiento de los vectores de movimiento de algunos macrobloques dentro de ese fotograma a cero; para posteriormente calcular el error de predicción del fotograma. Después de realizar sus experimentos los autores consiguen eficientemente eliminar la huella digital temporal de los vídeos MPEG que se han sometido a una eliminación o adición de fotogramas.

En cuanto a imágenes fijas, en [RC13] [GKWB07] realizan una clasificación de los ataques a las técnicas de análisis forense de imágenes según el objetivo a cumplir: (1) camuflaje de post-procesamientos maliciosos sobre la imagen, (2) destrucción de la identificación correcta del origen de la imagen y (3) falsificación del origen de imagen.

4.3.1. El Camuflaje de Post-Procesamientos

El objetivo principal de estas técnicas es camuflar algún proceso que al que haya sido sometido una imagen mediante el análisis de rasgos que dejan dichos procesos para ser contrarrestados. Existe investigaciones enfocados a estos rasgos de los algoritmos en imágenes como [PF05] [LF03] [LF03] [CSS07] [GKWB07].

En [GKWB07] se presenta una técnica para ocultar el proceso de re-muestreo *resampling*. El re-muestreo es el redimensionamiento con interpolación de las imágenes. Para detectar el re-muestreo los algoritmos realizan una búsqueda de dependencias sistemáticas y periódicas entre píxeles vecinos. Para ocultar el re-muestreo se elimina las equidistancias periódicas insertando distorsiones geométricas. El ataque consiste básicamente en generar una imagen (y) a partir de una imagen (x) aplicando el re-muestreo, para esto se calcula el componente de baja frecuencia de (x), se calcula el componente de alta frecuencia, y por último se obtiene un la imagen (y) que es la suma de los dos pasos anteriores. Los autores obtuvieron una tasa de falsos positivos *False Acceptance Rate* (FAR) inferior al 1%.

4.3.2. Manipulación de la Identificación de la Fuente

Imaginemos que un analista forense utiliza una técnica que extrae el ruido del sensor de la imagen para identificar la fuente, un contraataque a esta técnica sería la eliminación del ruido del sensor de la imagen, que se podría complementar con la sustitución del ruido del sensor de otra cámara. La

manipulación de la identificación de la fuente, según [RC13] se puede dar a través de dos procedimientos: (1) la destrucción de la identidad de una imagen, y (2) la falsificación de identidad de la imagen.

Para la destrucción de la identidad de una imagen existe métodos como la resta de las características del dominio *wavelet* que no es suficiente para eliminar el ruido de una imagen y deja rasgos visibles en la imagen [GKWB07]. Otro método de eliminación del ruido de una imagen es la corrección de sensibilidad o *flatfielding*, usado en la astronomía y escaneado de planos para mejorar la calidad de las imágenes. Esta corrección de sensibilidad se ejecuta utilizando el ruido de patrón fijo *Fixed Pattern Noise* (FPN) y el ruido de respuesta no uniforme PRNU. En [LFG06] [GKWB07], los autores manifiestan que los atacantes pueden evitar la identificación correcta de la fuente por el hecho de existir la posibilidad de eliminar y extraer la huella de una imagen.

Para la falsificación de la identidad de una imagen también se puede eliminar el ruido utilizando la técnica de corrección de sensibilidad, se puede incrustar el ruido de la imagen de otra cámara diferente mediante la corrección de sensibilidad inversa con la ecuación 4.1 [GKWB07].

$$\bar{y} = \bar{x} \cdot f_{falsa} + d_{falsa} \tag{4.1}$$

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

En [SLFK10] se propone el Algoritmo 1 para falsificar la identidad de una cámara.

Algoritmo 1: Falsificación de la identidad de una cámara

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

Si las dimensiones de ambas imágenes no coinciden se recorta o reconstruye. Adicionalmente en el Algoritmo 2 se propone una mejora al algoritmo de falsificación anterior para enmascarar los rasgos de la cámara *C2*.

Algoritmo 2: Falsificación de la identidad de una cámara para imágenes con dimensiones diferentes

1. Calcular el promedio de las huellas *F(C1)* de la cámara *C1* con la que se atacará;
2. Calcular el promedio de las huellas *F(C2)* de la cámara *C2*;
3. Sumar *F(C1)* a la fotografía *P*;
4. Tomar una fotografía P con la cámara *C2*;
5. Restar F*(C2)* a P;

Cuando se resta *F(C2)* se intenta eliminar la correlación entre la fotografía *P* y la cámara *C2*.

4.3.3. Detección de Falsificación de la Identidad de una Imagen

Según [SLFK10] [GFC11] es posible detectar un ataque de falsificación de la huella de una imagen y la inyección de ésta en otra imagen, mediante el análisis de las diferencias entre las propiedades de un patrón de ruido copiado. En la Figura 4.1 se presenta diagrama de procesos de un escenario de ataque basado a las ideas de [RC13], con la finalidad de explicar el proceso de detección de la falsificación de identidad de una imagen.

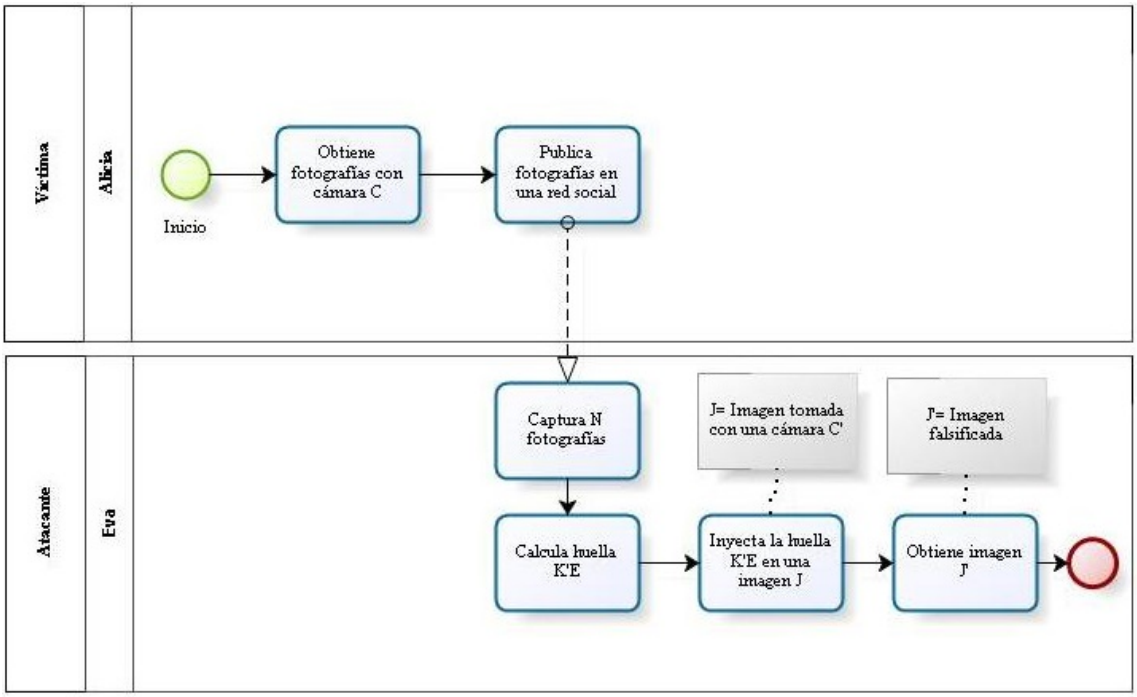

Figura 4.1: Escenario de ataque de falsificación de identidad

Alicia, la víctima publicó sus fotografías tomadas con su cámara ***C*** en una red social como muchas personas lo hacen en la actualidad, pero no se imaginó que **Eva** la atacante, descargaría sus ***N*** fotografías para calcular la huella ***K'E*.** **Eva** inyecta la huella obtenida ***K'E*** en una imagen ***J***, que fue tomada con una cámara ***C'***. Producto de los procesos anteriores se genera una imagen falsificada ***J'***. La finalidad de **Eva** al realizar este trabajo es hacer creer que **Alicia** fue la que generó la imagen falsificada.

En un plano legal, en caso que la imagen falsificada este inmiscuido en un delito, **Alicia** podría ser acusada y condenada. Los recursos para su defensa claramente serían un conjunto de fotografías ***C***, que son obtenidas de la cámara de **Alicia**, otro conjunto ***F*** de fotografías que **Eva** robó, y algunas de su propiedad.

En [GFC11] [RC13] se propone un escenario de defensa de Alicia, que comúnmente lo realiza una analista forense. Para mejor detalle en la Figura 4.2 se presenta el mencionado escenario.

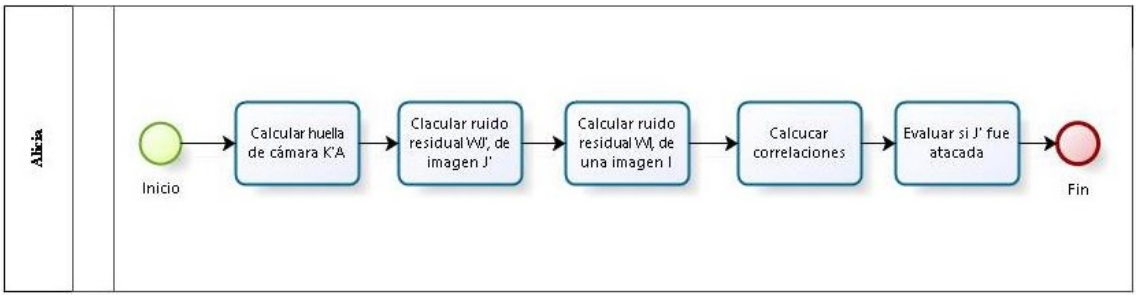

Figura 4.2: Escenario de defensa de falsificación de identidad

Alicia deberá calcular la huella de su cámara ***K'A*** utilizando imágenes planas inocentes para obtener una mejor estimación de la huella. Después calculara el ruido residual **PRNU** ***WJ'*** de la imagen ***J'***. Seguidamente calculara el ruido residual **PRNU** ***WI*** de una de las imágenes ***I*** utilizadas por **Eva** para realizar el ataque. Posteriormente **Alicia** calcula las correlaciones según la Tabla 4.1. Por último se evaluara si ***J'*** fue atacada realizando la prueba del triángulo como se muestra en la Figura 4.3, que se basa en el hecho de que el valor de correlación ***CI,J'*** de las imágenes ***I*** que no fueron utilizadas para falsificar ***J'*** puede ser estimado de las correlaciones ***CI,K'A*** y ***CJ',K'A***. En caso de que la imagen ***I*** haya sido utilizada para la falsificación la correlación ***CI,J'*** tendría un valor mayor que ***CI,K'A*** y ***CJ',K'A***.

Tabla 4.1: Calculo de correlaciones

Formula
$CI,J'=\text{corr}(WI, WJ')$
$CI,K'A=\text{corr}(WI, K'A)$
$CJ',K'A=\text{corr}(WJ',K'A)$

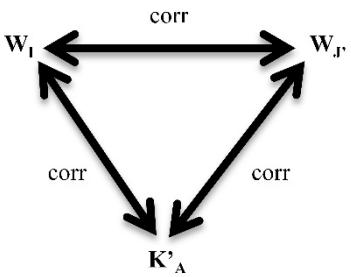

Figura 4.3: Correlaciones de la prueba del triángulo

En resumen esta propuesta refuerza los elementos necesarios para no culpar a alguien inocente, aunque el contraataque no es al 100% efectivo, porque los autores malintencionados pueden generar imágenes ilegales con características similares a las fotografías de la víctima.