

Identificación de la Fuente en Vídeos de Dispositivos Móviles

David Manuel Arenas González, Jocelin Rosales Corripio, Ana Lucila Sandoval Orozco, Hiram Jafet Romo Torres, Luis Javier García Villalba

Grupo de Análisis, Seguridad y Sistemas (GASS), Departamento de Ingeniería del Software e Inteligencia Artificial
Facultad de Informática, Despacho 431, Universidad Complutense de Madrid (UCM)
Calle Profesor José García Santesmases, 9, Ciudad Universitaria, 28040 Madrid
Email: {darenas, jocerosa, asandoval, javiergv}@fdi.ucm.es, hromot@hotmail.com

Resumen—La realización de vídeos con dispositivos móviles se ha convertido en una actividad común dado su alto grado de utilización y el gran número de usuarios. Además, la portabilidad de este tipo de dispositivos hace que estén a mano de los usuarios gran cantidad de tiempo facilitando que se utilicen para generar vídeos en una gran diversidad de situaciones. Por tanto, estos vídeos pueden ser utilizados como evidencias en procesos judiciales. Todo lo anterior hace necesario contar con técnicas de análisis forense enfocadas en vídeos de dispositivos móviles dada las características peculiares de sus cámaras. En este trabajo se estudia la identificación de la fuente de adquisición de los vídeos de dispositivos móviles y se presenta una técnica basada en la extracción del ruido del sensor y la transformada wavelet de los fotogramas extraídos del vídeo. Estos fotogramas son extraídos mediante un algoritmo que tiene en cuenta la naturaleza de los mismos, mejorando la selección de los fotogramas a analizar. Finalmente se presentan experimentos con vídeos de dispositivos móviles para evaluar la validez de las técnicas utilizadas.

Palabras clave—Análisis forense de vídeos, fuente de adquisición de vídeos, patrón de ruido del sensor, PRNU. (*Video forensics analysis, video source acquisition, sensor pattern noise, PRNU*).

I. INTRODUCCIÓN

Si las imágenes capturadas por dispositivos electrónicos son consideradas parte de la verdad como hechos reales, en pocos minutos, un vídeo puede comunicar una enorme cantidad de información. Según el medidor de tráfico “*Alexa, The Web Information Company*” [1], Youtube es actualmente el tercer sitio con más visitas del mundo, lo cual nos deja un claro indicio de la popularidad de la que gozan los vídeos entre los diferentes medios en los que puede desplegarse. Existe una amplia gama de dispositivos móviles que pueden reproducirlo y/o grabarlo, como por ejemplo: teléfonos móviles, tablets, vídeoconsolas portátiles y cámaras digitales o de vídeo. En cuanto a los dispositivos móviles, *Gartner Inc.* [2], afirma que las ventas de teléfonos inteligentes creció un 36 % en el cuarto trimestre del 2013. Asimismo, este tipo de dispositivos representó el 57.6 % de las ventas globales de teléfonos móviles en el cuarto trimestre de 2013, frente al 44 % del año anterior. Al igual que las cámaras digitales han desplazado en términos de uso a las cámaras tradicionales de película, actualmente, los dispositivos móviles equipados con cámaras, tienen un papel importante poniendo fin al rápido crecimiento

de las cámaras digitales. En los dispositivos móviles, se ha visto una gran competencia entre fabricantes que se esfuerzan en integrar una videocámara de alta definición al alcance del usuario. Como consecuencia de este fenómeno y de la gran cantidad de tiempo que una persona pasa junto a un teléfono inteligente, este se ha convertido en el primer dispositivo de grabación de vídeos para muchos usuarios en la sociedad actual.

Debido al frecuente uso de los dispositivos móviles, en ciertos casos existen restricciones legales sobre el uso de este dispositivo, así como también de su uso en distintos lugares, tales como: colegios, universidades, oficinas de gobierno, empresas, etc. Actualmente los vídeos se exhiben con mayor frecuencia, ya sea directa o indirectamente en procesos judiciales como pruebas o evidencias para la aplicación de la ley [3]. Por tanto, dada la importancia de los vídeos en estas situaciones, el análisis forense cobra especial relevancia. Dentro de las distintas ramas del análisis forense, destaca la que nos permite identificar la fuente de adquisición, en este caso de la videocámara que generó el vídeo. En este trabajo se presentan técnicas de análisis forense para la identificación de la fuente de adquisición de vídeos, centrándonos especialmente en los vídeos generados por dispositivos móviles.

Este trabajo está estructurado en 6 secciones, siendo la primera de ellas la presente introducción. En la sección 2 se presentan brevemente las diferencias entre el pipeline en la creación de una imagen y un vídeo. La sección 3 realiza un estado del arte del análisis forense para imágenes y vídeos generados por dispositivos móviles. En la sección 4 se presenta la técnica propuesta. Los experimentos realizados y sus resultados son presentados en la sección 5. Por último en la sección 6 se presentan las conclusiones obtenidas de este trabajo.

II. PIPELINE DE UNA VIDEOCÁMARA

Antes de mencionar alguna de las técnicas existentes para la identificación de la fuente, es importante comprender cuál es el procedimiento realizado para generar un vídeo. Este proceso es similar en la generación de una imagen y de un vídeo, salvo que en un vídeo finalmente existe un último paso que consiste

en codificar los fotogramas resultantes para la creación de un archivo único final de vídeo. Esta codificación tiene como objetivo transformar todos los fotogramas capturados en una secuencia de ellos a lo largo de un tiempo. También se busca conseguir un tamaño lo más óptimo posible del archivo final, ya que en un vídeo existen fotogramas capturados que son redundantes entre sí. Es decir, en ocasiones entre un fotograma y otro, se puede compartir características de la escena que facilitan el poder optimizar el tamaño del vídeo final sin perder contenido visual. Por ejemplo, para la codificación MPEG, existe una estructura llamada GOP (*Group of Pictures*) que especifica el orden en el que las imágenes son ordenadas y soluciona el problema de redundancia en la codificación. Para la codificación habitualmente se utilizan los códecs: MPEG-x o H.26x para cámaras digitales y 3GP para teléfonos móviles, este último tiene la capacidad de ser compatible con los códecs: MPEG-4, H.263 o H.264 [4].

III. TÉCNICAS DE IDENTIFICACIÓN DE LA FUENTE

La mayor parte de las investigaciones realizadas sobre la identificación de la fuente se han realizado para imágenes fotográficas estáticas. La mayoría de las técnicas que se pueden aplicar a una imagen se pueden emplear con los diferentes fotogramas de un vídeo [5].

En [6] 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.

El área que está 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 [7], [8], [9] y [10] se utilizan los metadatos con fines de clasificación de imágenes digitales.

En [11] se trata el tema de la identificación de la fuente utilizando las características de la misma. Se contemplan tres tipos de características: características de color, características de calidad y características de la imagen en el dominio de la frecuencia. La clasificación de las imágenes es realizada por una Máquina de Soporte Vectorial (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 %. Un último experimento se realizó con un conjunto de 8 cámaras que alcanzo una precisión del 95,46 %.

En [12] 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 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 [13], [14] y [15].

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 [16] se demostró que los sensores de las cámaras generan un patrón de ruido (*Sensor Pattern Noise*) que podría ser utilizado como método único de identificación.

En [17] se demostró que el ruido del sensor extraído de las imágenes podían 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 [18], [19] y [20] se presentan otros métodos de identificación de fuente basados en las imperfecciones del sensor.

Por último, en el área de las transformadas wavelets existen diversos enfoques. Por ejemplo en [21] se propone una nueva técnica de identificación basada en las características de probabilidad condicional. Este tipo de características fueron propuestas inicialmente para propósitos de estegoanálisis en [22]. 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 [23] 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 frames aplicando algún método de clasificación para imágenes fijas.

En [24], 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 (20 de ellas procedentes de VQEG (*Video Quality Experts Group*)[25] y 80 de DVDs). 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 [26] 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.

IV. DESCRIPCIÓN DE LA TÉCNICA

Al completar la generación de un vídeo es posible que se introduzcan, en cada uno de sus fotogramas, algunos defectos que se vean reflejados como ruido, llamados comúnmente “huellas”. Estas “huellas” se pueden utilizar para detectar la fuente de adquisición del vídeo. A continuación, se describe el método de extracción de fotogramas del vídeo y se referencia el algoritmo que permite extraer el patrón de ruido del sensor.

Los fotogramas seleccionados para la clasificación son obtenidos mediante el algoritmo 1.

Algoritmo 1: Algoritmo de extracción de fotogramas

Input: vídeo: Vídeo a procesar
 nFotogramas: Número de fotogramas deseados
 $umbral_I$: Umbral inicial
 inc_U : Incremento del umbral

Result: fotogramas: Vector de fotogramas

- ① $histogramas \leftarrow extraerHistogramas(vídeo)$;
 - ② $umbral \leftarrow estimarUmbral(histogramas, nFotogramas, umbral_I, inc_U)$;
 - ③ $fotogramas \leftarrow extraerFotogramas(vídeo, umbral, histogramas)$;
-

El algoritmo calcula y compara los fotogramas contenidos en un vídeo. Los fotogramas que presenten un cambio de escena significativo serán utilizados para la clasificación e identificación. Esto se debe a que el ruido del sensor extraído de una imagen puede estar severamente contaminada por los detalles de la escena [17], además de que los datos de un vídeo contienen redundancia temporal, espacial y espectral. En el algoritmo son necesarios 4 parámetros para su funcionamiento:

- Video del cual serán extraídos los fotogramas.
- Numero de fotogramas deseados a extraer.
- Umbral inicial que será la referencia para determinar cuando existe un cambio de escena.
- Valor del incremento para el umbral que se realizará en cada iteración.

El primer paso consiste en extraer el histograma (frecuencia de los valores de color) de los fotogramas, y calcular mediante la correlación de cada par de fotogramas contiguos la similitud existente. La correlación se calcula con la ecuación (1).

$$correlacion(H_1, H_2) = \frac{\sum_i H'_1(i)H'_2(i)}{\sqrt{\sum_i H'_1(i)^2 H'_2(i)^2}} \quad (1)$$

donde,

$$H'_k(i) = H_k(i) - \frac{1}{N} \left(\sum_j H_k(j) \right)$$

siendo N el número de niveles de intensidad para cada canal de color RGB.

Existen diversos métodos para calcular la diferencia de histogramas de color de dos dimensiones. En este trabajo se optó por el cálculo de la correlación, ya que es un vector aleatorio (variable aleatoria multidimensional) y además, los resultados obtenidos con el coeficiente de correlación son mejores que otras medidas [27]. Se puede mencionar por ejemplo la distribución de probabilidad continua (chi-cuadrado) o la intersección o distancia de Bhattacharyya.

El primer fotograma del vídeo se toma como parte del conjunto de fotogramas elegidos. Se realiza la comparación tomando el primer y el segundo fotograma, si no hay una diferencia significativa entre ellos en base al umbral, se toma el siguiente fotograma y se realiza una nueva comparación con el primero, esto se realiza hasta que el resultado de la correlación sea menor al umbral, para tomar en cuenta al fotograma para la clasificación e identificación. Si al final la cantidad de cambios de escena en base al umbral, es menor a la cantidad necesaria, se repite el proceso de comparación incrementando el umbral, hasta que la cantidad de cambios de escena sea mayor o igual a los deseados.

Para poder determinar el umbral inicial, se realizaron varios experimentos sobre los vídeos, y se halló que mediante la comparación de los histogramas de un vídeo, la correlación promedio más baja fue de -0.27, presentando al menos 1 o 2 cambios de escena, definiendo así el umbral inicial. Para el valor del incremento se experimentó con diferentes valores, tales como: 0.1, 0.01, 0.001, 0.0001 y 0.0001. El valor de 0.001 fue el elegido, ya que demostró ser un valor ideal para llegar a el número de fotogramas deseados en un menor tiempo y con más exactitud. Estos incrementos se realizan porque, si el umbral se encuentra más cercano al valor máximo de correlación directa, es decir al valor de 1, se pueden encontrar más cambios de escena, y así extraer la cantidad de fotogramas definidos por el usuario para la clasificación e identificación.

Mediante el análisis de los trabajos de la literatura, se llegó a la conclusión de que el patrón de ruido del sensor y la transformada wavelet, ayudan a definir una huella, siendo métodos efectivos para la identificación de fuente. Este artículo extiende el uso del patrón de ruido del sensor y la transformada wavelet de [23]. La técnica está enfocada en representar las huellas en vectores de características.

El esquema presentado en la Figura 1 muestra el diagrama funcional de la técnica.

La obtención del patrón de ruido del sensor de las imágenes, se basa en el método descrito en [23].

El siguiente paso es obtener las características que caracterizan el ruido del sensor para fines de la clasificación. Un total de 81 características son obtenidas utilizando el algoritmo de extracción de características descrito en [23].

V. EXPERIMENTOS Y RESULTADOS

Para probar la efectividad del técnica propuesta, se capturaron vídeos sin ninguna consideración en las características temporales o espaciales, debido a que deben representar casos reales. Como actualmente los dispositivos móviles presentan grandes mejoras en la calidad del vídeo, se consideró usar

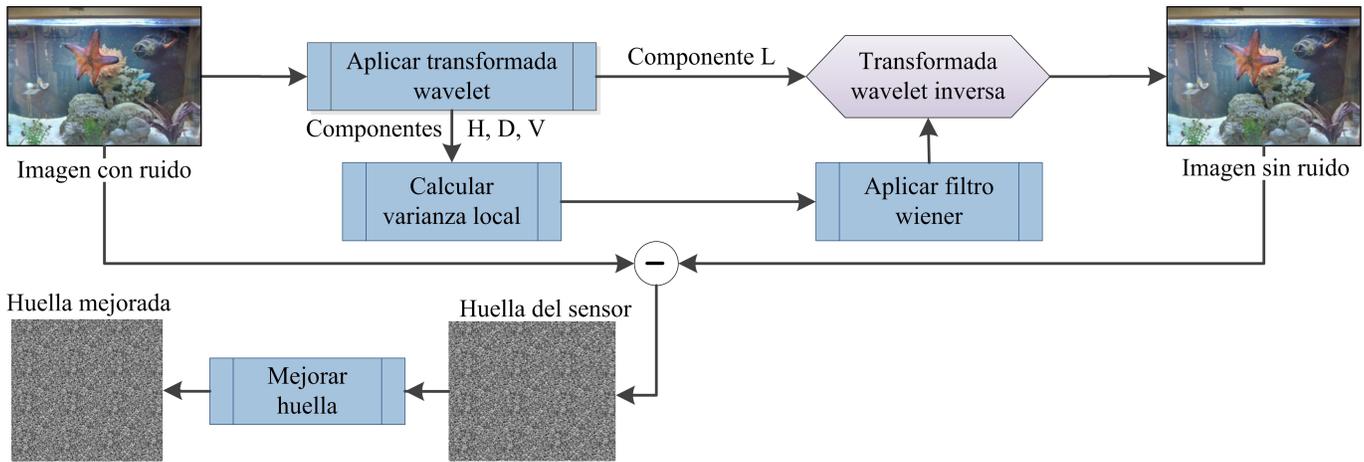


Figura 1. Esquema funcional de extracción de patrón de ruido del sensor.

vídeos con calidad de 1080p (vídeos de alta definición), es decir con una resolución de 1920x1080 píxeles.

La Tabla I muestra las especificaciones básicas y los modelos de dispositivos móviles considerados para los experimentos.

Tabla I
CONFIGURACIÓN DE LAS CÁMARAS DE DISPOSITIVOS MÓVILES

| Marca (Modelo) | FPS | Formato | Códec | Condiciones de Captura |
|-------------------------|-----|---------|--------|--|
| Apple iPhone 5 (M1) | 24 | mov | H.264 | Resolución: 1080p Tipo de escena: Cualquiera Orientación: Vertical Flash: Deshabilitado Luz: Natural Balance de blancos: Autom Zoom digital: 0 Tiempo de vídeo: 2 min |
| Nokia 808 Pureview (M2) | 30 | mp4 | MPEG-4 | |
| Samsung Galaxy S4 (M3) | 30 | mp4 | MPEG-4 | |
| Wiko Cink Slim (M4) | 12 | 3gp | MPEG-4 | |
| Zopo ZP-980 (M5) | 15 | 3gp | MPEG-4 | |

En la Tabla II se resumen las condiciones experimentales usadas en la evaluación de la técnica propuesta.

La clasificación se ha realizado utilizando el método de aprendizaje supervisado SVM con kernel RBF, ya que es uno de los más utilizados para este tipo de clasificaciones [6]. Los parámetros utilizados en el clasificador SVM son los mismos que los empleados en [23]. Se utilizó el paquete LibSVM [28] que permite la clasificación de múltiples clases.

Con la finalidad de mostrar a grandes rasgos tiempos de ejecución para la obtención del conjunto de características, para una imagen con un recorte de 1024x768, en un procesador Intel Core i7 de 1.6GHz con 8Gb de RAM, se consumen aproximadamente 2 segundos. Para 500 imágenes con el recorte anterior y utilizando la misma máquina, en las fases de entrenamiento y clasificación de la SVM se emplearon aproximadamente unos 650 segundos y 1 segundo respectivamente.

Tabla II
PARÁMETROS DE LOS EXPERIMENTOS

| Parámetro | Valor |
|--|------------|
| Numero de vídeos para entrenamiento por cámara | 5 |
| Numero de vídeos para pruebas por cámara | 5 |
| Método de extracción | Histograma |
| Umbral inicial | -0.27 |
| Incremento umbral | 0.001 |
| Numero de fotogramas deseados por vídeo | 100 |

V-A. Influencia de la Resolución en la Tasa de Acierto

Para analizar la influencia que tiene tamaños de recortes de los fotogramas en la tasa de acierto, se realizó la identificación de la fuente de los 5 dispositivos móviles de la Tabla I con cada una de las siguientes resoluciones de los fotogramas:

- Resoluciones estándar: 128x128, 320x240, 640x480, 800x600, 1024x768.
- Resolución recomendada: 1024x1024
- Resolución real del fotograma: 1920x1080.

Los parámetros utilizados para la extracción de las características definido en [23] son: Daubechies 8 wavelet, recorte del fotograma centrado y estimación de varianza adaptativa.

En la Tabla III se muestra el porcentaje de acierto medio en la identificación de la fuente de cada dispositivo con respecto a los distintos tamaños de recortes de los fotogramas. Por porcentaje de acierto se entiende al porcentaje de fotogramas de los 5 vídeos de cada dispositivo móvil clasificados correctamente. Cada vídeo obtuvo un porcentaje de acierto en la identificación de la fuente que se muestra en la tabla III.

En la mayoría de los casos, los porcentajes de acierto por dispositivo, aumentan cuanto más grande sea el recorte de los fotogramas (esto se da para todos los casos si se tiene en cuenta la tasa de acierto promedio). Obteniendo para la mayor resolución la mayor tasa de acierto promedio, un 85.56%. En todos los experimentos realizados se supera la tasa por vídeo individual del 50%. Esto indica que en todos los casos, para

Tabla III
TASA DE ACIERTO \times RESOLUCIÓN DE FOTOGRAMAS.

| Resolución | Dispositivos | | | | | % Acierto |
|------------------|----------------|----------------|----------------|----------------|----------------|----------------|
| | M1 | M2 | M3 | M4 | M5 | Medio |
| 128x128 | 65.80 % | 79.20 % | 66.60 % | 75.00 % | 64.00 % | 70.12 % |
| 320x240 | 73.80 % | 88.20 % | 78.00 % | 78.80 % | 63.40 % | 76.44 % |
| 640x480 | 79.60 % | 95.60 % | 85.00 % | 85.60 % | 66.20 % | 82.40 % |
| 800x600 | 81.80 % | 96.80 % | 84.80 % | 86.00 % | 68.80 % | 83.64 % |
| 1024x768 | 80.80 % | 97.40 % | 88.80 % | 85.40 % | 75.40 % | 85.56 % |
| 1024x1024 | 81.20 % | 97.20 % | 92.20 % | 88.00 % | 78.40 % | 87.40 % |
| 1920x1080 | 87.40 % | 98.80 % | 93.00 % | 89.80 % | 82.60 % | 90.32 % |

todos los fotogramas de un vídeo concreto de un dispositivo concreto, al menos el 50 % de los fotogramas son identificados correctamente. Finalmente la identificación de la fuente de un vídeo debe responder a la pregunta concreta de a qué fuente de adquisición pertenece ese vídeo. Como criterio lógico, puede estimarse que el vídeo pertenece a la fuente con el mayor número de fotogramas clasificados con respecto a las otras fuentes (mayor porcentaje de acierto con respecto a las otras fuentes). Se podría dar el caso en el que varias fuentes tengan exactamente el mismo número de fotogramas clasificados y a su vez sean el mayor número con respecto a las otras fuentes. En este caso, poco habitual, se diría que la fuente del vídeo no puede ser identificada con determinación y estaría entre la duda de esas distintas fuentes.

Los resultados obtenidos no dejan lugar a dudas sobre la identificación de la fuente de adquisición del vídeo teniendo en cuenta el criterio definido anteriormente, ya que en todos los casos el acierto superó el 50 %. Asimismo puede verse que las tasas de acierto en muchos casos son mucho mayores (llegando en ocasiones hasta el 100 %). Por tanto, según este experimento, tomando el criterio antes definido y teniendo en cuenta el vídeo como entidad unitaria (es decir un vídeo se clasifica bien o no), se puede concluir que esta técnica identifica la fuente de un vídeo con un 100 % de acierto.

Como se puede observar en la Tabla III, utilizando la imagen completa existe una mayor tasa de acierto promedio en la identificación de fuente, aunque el incremento es pequeño. Sin embargo, la influencia que tiene la resolución en la tasa de aciertos de la identificación de la fuente que adquirió un vídeo se refleja en la Figura 2, donde se muestra que la mejora en la tasa de acierto para la tamaño de recorte de 1920x1080 es del 2.92 % con respecto a un tamaño de recorte de 1024x1024. Por tanto, a partir de un cierto tamaño de recorte el incremento de la tasa de acierto es pequeña, e incluso en algunos casos este puede disminuir un poco. También hay que tener en cuenta que a mayor tamaño de recorte mayor tiempo de ejecución del algoritmo de extracción de características.

V-B. Influencia de los Parámetros de Ejecución en la Tasa de Acierto

Para analizar cómo afecta el uso de los distintos parámetros del algoritmo propuesto, en [23] en la identificación de la fuente de vídeos, se realizó un conjunto de experimentos utilizando el mismo tamaño de recorte centrado del fotograma

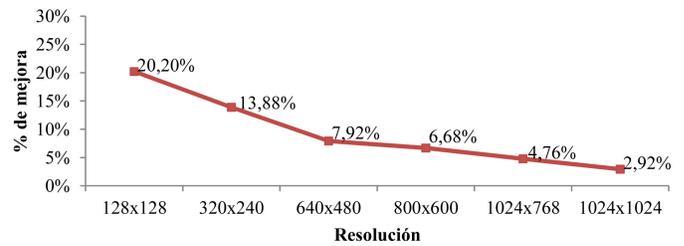


Figura 2. Porcentaje de mejora de la tasa de acierto \times por resolución.

(640x480) para identificar la fuente de los 5 dispositivos móviles de la Tabla I. En la Tabla IV se muestra un resumen de los experimentos realizados y los parámetros de configuración del algoritmo de extracción de las características utilizados en cada uno de ellos.

Tabla IV
TASA DE ACIERTO UTILIZANDO DIFERENTES CONFIGURACIONES

| Configuración | Varianza | Aplicar Zero Meaning | % de Acierto |
|---------------|---------------|----------------------|--------------|
| 1 | Adaptativa | No | 82.4 % |
| 2 | Adaptativa | Sí | 82.32 % |
| 3 | No adaptativa | Sí | 81.56 % |
| 4 | No adaptativa | No | 82.96 % |

La diferencia entre la mejor y la peor tasa de acierto mostrada en la Tabla IV es del 1.4 %. Este resultado indica que los parámetros de configuración del algoritmo de extracción de características no influye significativamente en el porcentaje de acierto de identificación de la fuente de vídeos. Puede verse que la tasa óptima de acierto se consigue con los parámetros de estimación de la varianza no adaptativa y sin utilizar el filtro zero-meaning. Asimismo los peores resultados se obtuvieron con los parámetros de estimación de varianza no adaptativa y el uso del filtro zero-meaning. Dado el estrecho margen entre las tasas de acierto que hay entre las diferentes configuraciones y los resultados de la configuración óptima y las restantes, las conclusiones no pueden extrapolarse de forma categórica para cualquier experimento, aunque éstas deben de tenerse en cuenta para futuros experimentos y aplicación de la técnica.

VI. CONCLUSIONES

Una vez presentada la técnica y realizados los experimentos variando los distintos parámetros, se llega a la conclusión general de que esta técnica obtiene buenos resultados y es válida para la identificación de la fuente en vídeos de dispositivos móviles. La aplicación a escenarios reales de esta técnica la consideramos realista y viable, siempre que los vídeos a clasificar pertenezcan a un conjunto cerrado y conocido de dispositivos móviles.

El algoritmo de extracción de fotogramas presentado tiene en cuenta la naturaleza de un vídeo y sus fotogramas, optimizando la extracción de los fotogramas claves. Es decir, extrae los fotogramas teniendo en cuenta que si los fotogramas obtenidos tienen mayor variación de escena entre ellos (buscando los cambios de escena), el proceso de clasificación

obtendrá mejores resultados. Sin embargo, para la clasificación utilizando SVM se necesita un número determinado de fotografías para el entrenamiento, y esto el algoritmo también lo tiene en cuenta ya que se puede dar el caso en el que el vídeo cambie poco de escena y tenga que obtener los fotogramas más distantes en escenas entre los que hay. Una vez obtenidos los fotogramas nos basamos en la extracción de características que se obtienen del patrón de ruido del sensor y la transformada wavelet especificado en [23].

Los resultados promedios de clasificación varían dependiendo de los parámetros utilizados. Teniendo en cuenta un tamaño de recorte centrado, se concluye que a mayor tamaño de recorte, mejores son los resultados. Asimismo, se ha evaluado como afecta el uso de los distintos parámetros de configuración definidos en [23] en la identificación de la fuente de vídeos de dispositivos móviles usando un recorte centrado de fotograma de 640x480. En este sentido, no se han podido obtener conclusiones categóricas y extrapolables sobre el uso de los parámetros de configuración, ya que en todos los experimentos realizados la tasa de acierto están comprendida en un margen muy pequeño.

Una vez clasificados los fotogramas seleccionados, se debe responder a la pregunta de cuál es la fuente de adquisición del vídeo como entidad unitaria. Nuestro criterio ha sido que el vídeo pertenece a la fuente cuyo mayor número de fotogramas se han clasificado de ese tipo. Este criterio debe tenerse en cuenta en futuras comparaciones con otras técnicas.

AGRADECIMIENTOS

El Grupo de Investigación GASS agradece la infraestructura proporcionada por el Campus de Excelencia Internacional (CEI) Campus Moncloa - Clúster de Cambio Global y Nuevas Energías (y, más concretamente, el sistema EOLO como recurso de computación de alto rendimiento HPC - High Performance Computing), infraestructura financiada por el Ministerio de Educación, Cultura y Deporte (MECD) y por el Ministerio de Economía y Competitividad (MINECO).

REFERENCIAS

- [1] "Alexa Top 500 Global Sites," <http://www.alexa.com/topsites>, 2014.
- [2] "Gartner Says Annual Smartphone Sales Surpassed Sales of Feature Phones for the First Time in 2013," <http://www.gartner.com/newsroom/id/2665715>, 2014.
- [3] C. Wen and K. Yang, "Image authentication for digital image evidence," *Forensic Science Journal*, vol. 5, no. 1, pp. 1–11, September 2006.
- [4] "Formatos, códecs y extensiones de archivos de audio y vídeo compatibles - Cómo - BlackBerry Q10 Smartphone - 10.1," http://docs.blackberry.com/es-es/smartphone_users/deliverables/50637/mba1344953159594.jsp, 2014.
- [5] P. Bestagini, M. Fontani, S. Milani, M. Barni, A. Piva, M. Tagliasacchi, and S. Tubaro, "An Overview on Video Forensics," in *Proceedings of the 20th European Signal Processing Conference*, August 2012, pp. 1229–1233.
- [6] A. Sandoval Orozco, D. Arenas González, J. Rosales Corripio, L. García Villalba, and J. Hernández-Castro, "Techniques for Source Camera Identification," in *Proceedings of the 6th International Conference on Information Technology*, May 2013, pp. 1–9.
- [7] N. L. Romero, V. G. Chornet, J. S. Cobos, A. S. Carot, F. C. Centellas, and M. C. Mendez, "Recovery of Descriptive Information in Images From Digital Libraries by Means of EXIF Metadata," *Library Hi Tech*, vol. 26, no. 2, pp. 302–315, 2008.
- [8] M. Boutell and J. Luo, "Beyond Pixels: Exploiting Camera Metadata for Photo Classification," *Pattern Recognition*, vol. 38, no. 6, pp. 935–946, June 2005.
- [9] J. Tesic, "Metadata Practices for Consumer Photos," *IEEE Multimedia*, vol. 12, no. 3, pp. 86–92, September 2005.
- [10] M. Boutell and J. Luo, "Photo Classification by Integrating Image Content and Camera Metadata," in *Proceedings of the 17th International Conference on Pattern Recognition*, vol. 4. IEEE Computer Society, August 2004, pp. 901–904.
- [11] M. J. Tsai, C. L. Lai, and J. Liu, "Camera/Mobile Phone Source Identification for Digital Forensics," in *Proceedings of the International Conference on Acoustics Speech and Signal Processing*. IEEE, April 2007, pp. II–221–II–224.
- [12] O. Celiktutan, I. Avciabas, B. Sankur, N. P. Ayerden, and C. Capar, "Source Cell-Phone Identification," in *Proceedings of the IEEE 14th Signal Processing and Communications Applications*. IEEE, April 2006, pp. 1–3.
- [13] Y. Long and Y. Huang, "Image Based Source Camera Identification using Demosaicking," in *Proceedings of the IEEE 8th Workshop on Multimedia Signal Processing*. IEEE, October 2006, pp. 419–424.
- [14] S. Bayram, H. Sencar, N. Memon, and I. Avciabas, "Source Camera Identification Based on CFA Interpolation," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 3, September 2005, pp. III–69–72.
- [15] S. Bayram, H. T. Sencar, and N. Memon, "Classification of Digital Camera-Models Based on Demosaicing Artifacts," *Digital Investigation*, vol. 5, no. 1, pp. 49–59, June 2008.
- [16] J. Lukas, J. Fridrich, and M. Goljan, "Digital Camera Identification from Sensor Pattern Noise," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, June 2006.
- [17] C. Li, "Source Camera Identification Using Enhanced Sensor Pattern Noise," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 280–287, June 2010.
- [18] Z. J. Gerads, J. Bijhold, M. Kieft, K. Kurosawa, K. Kuroki, and N. Saitoh, "Methods for Identification of Images Acquired with Digital Cameras," in *Proceedings on Enabling Technologies for Law Enforcement and Security*, vol. 4232. SPIE-International Society for Optical Engineering, February 2001, pp. 505–512.
- [19] T. Lanh, K. Chong, S. Emmanuel, and M. Kankanhalli, "A Survey on Digital Camera Image Forensic Methods," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, July 2007, pp. 16–19.
- [20] F. D. O. Costa, M. Eckmann, W. J. Scheirer, and A. Rocha, "Open Set Source Camera Attribution," in *Proceedings of the Patterns and Images Proceedings of the 25th Conference on Graphics*. IEEE, August 2012, pp. 71–78.
- [21] A. Wahab, A. Ho, and S. Li, "Inter-Camera Model Image Source Identification with Conditional Probability Features," in *Proceedings of the 3rd Image Electronics and Visual Computing Workshop*, November 2012.
- [22] A. Wahab, J. Briffa, H. Schaathun, and A. T. S. Ho, "Conditional Probability Based Steganalysis for JPEG Steganography," in *Proceedings of the International Conference on Signal Processing Systems*, May 2009, pp. 205–209.
- [23] J. Rosales Corripio, A. L. Sandoval Orozco, L. J. García Villalba, J. Hernández-Castro, and S. J. Gibson, "Source Smartphone Identification Using Sensor Pattern Noise and Wavelet Transform," in *Proceedings of the 5th International Conference on Imaging for Crime Detection and Prevention*, December 2013, pp. 1–6.
- [24] Y. Su, J. Xu, and B. Dong, "A Source Video Identification Algorithm Based on Motion Vectors," in *Proceedings of the Second International Workshop on Computer Science and Engineering*, vol. 2, October 2009, pp. 312–316.
- [25] "Video Quality Experts Group (VQEG)," <http://www.its.bldrdoc.gov/vqeg/vqeg-home.aspx>, 2014.
- [26] S. Yahaya, A. Ho, and A. Wahab, "Advanced Video Camera Identification Using Conditional Probability Features," in *Proceedings of the IET Conference on Image Processing*, July 2012, pp. 1–5.
- [27] E. Bashkov and N. Shozda, "Content-Based Image Retrieval Using Color Histogram Correlation," *Graphics Proceedings*, 2002.
- [28] C. C. Chang and C. J. Lin, "LIBSVM: A Library for Support Vector Machines. Version 3.17, April 26, 2013," <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.