Clasificación sin Supervisión de Imágenes de Dispositivos Móviles

David Manuel Arenas González, Jocelin Rosales Corripio, Ana Lucila Sandoval Orozco,
Jorge Alberto Zapata Guridi, 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, jorge.zapata@jazg.net

***Resumen*—Cada día el uso de imágenes de dispositivos móviles como evidencias en procesos judiciales es más habitual y común. Por ello, el análisis forense de imágenes de dispositivos móviles cobra especial importancia. En este trabajo se estudia la rama del análisis forense que se basa en la identificación de la fuente, concretamente en la agrupación o *clustering* de imágenes según la fuente de adquisición. Como diferencia con otras técnicas del estado del arte para la identificación de la fuente, en el *clustering* no se tiene un conocimiento a priori del número de imágenes ni dispositivos a identificar, ni se tienen datos de entrenamiento para una futura fase de clasificación. Es decir, se realiza un agrupamiento por clases con todas las imágenes de entrada. La propuesta se basa en la combinación de *clustering* jerárquico y plano y en el uso del patrón de ruido del sensor. Se han realizado un conjunto de experimentos que emulan situaciones similares a las que se pueden dar en la realidad para mostrar la robustez y fiabilidad de los resultados de la técnica. Los resultados obtenidos son satisfactorios en todos los experimentos realizados superando en tasa de acierto a otras propuestas descritas en el estado del arte.**

***Palabras clave*—Análisis forense de imágenes, clustering de imágenes, patrón de ruido del sensor, PRNU. *(Image forensics analysis, image clustering, sensor pattern noise, PRNU)*.**

I. Introducción

En la actualidad, el número de cámaras integradas a dispositivos móviles ha proliferado permitiendo a millones de consumidores tomar fotografías e incluso compartir de manera sencilla el contenido capturado. La industria de los dispositivos móviles ha desarrollado la tecnología necesaria para abaratar los costos y de esta manera hacerlos muy accesibles al público.

El gran número de cámaras en dispositivos móviles constituye un mayor número de evidencias presentadas ante la ley en delitos como robo de información de tarjetas de crédito, pornografía infantil, espionaje industrial, etc. Por tanto, el análisis forense de este tipo de imágenes cobra especial importancia en las investigaciones judiciales. Dentro de análisis forense de imágenes digitales existen dos grandes ramas: la identificación de la fuente de adquisición y la detección de manipulaciones malintencionadas. Este trabajo se centra en la primera rama, es decir, dada una imagen o conjunto de imágenes identificar la marca y modelo de la cámara que realizó la foto mediante la clasificación por agrupamiento o *clustering*. Asimismo, dado que las cámaras de dispositivos móviles tienen unas características propias que las hacen diferentes a la restantes, este trabajo se enfoca en las fotos de este tipo de dispositivos.

Dentro de la identificación de la fuente existen dos grandes enfoques: escenarios cerrados o escenarios abiertos. Un escenario cerrado es aquel en el cual la identificación de la fuente de la imagen se realiza sobre un conjunto de cámaras concreto y conocidas a priori. Para este enfoque normalmente se utiliza un conjunto de imágenes de cada cámara para entrenar un clasificador y posteriormente se predice la fuente de adquisición de las imágenes objeto de investigación. La técnica más utilizada para la tarea de clasificación de imágenes digitales es *Support Vector Machine* (SVM). Este trabajo se centra en la identificación de la fuente en escenarios abiertos, es decir, el analista forense no conoce a priori el conjunto de cámaras a las que pertenece la imagen a identificar su fuente. Obviamente en este tipo de clasificación, en la que no se tienen datos de cámaras a priori, el objetivo no es identificar la marca y modelo de la cámara, sino poder agrupar distintas imágenes en grupos disjuntos en los que todas sus imágenes pertenecen al mismo dispositivo. Este planteamiento es muy cercano a situaciones de la vida real, ya que en muchos casos el analista desconoce por completo el conjunto de cámaras a las que pueden pertenecer un conjunto de imágenes. Además, es prácticamente imposible tener un conjunto de imágenes para entrenar un clasificador con todas las cámaras de dispositivos móviles existentes en el mundo.

Este trabajo está estructurado en 5 secciones, siendo la primera de ellas la presente introducción. En la sección 2 se presentan brevemente los trabajos previos relacionados con las técnicas de análisis forense para la identificación de la fuente de imágenes de dispositivos móviles. En la sección 3 se presenta la técnica propuesta. Los experimentos realizados y sus resultados se presentan en la sección 4. Por último en la sección 5 se presentan las conclusiones obtenidas de este trabajo.

II. Trabajos Relacionados

La mayoría de las investigaciones realizadas sobre la identificación de la fuente de adquisición de imágenes se centran en cámaras digitales tradicionales o DSC (*Digital Still Camera*), no siendo en su mayoría estas técnicas válidas para imágenes

ISBN: 978-84-9717-323-0

de dispositivos móviles. La principal razón por la que se necesitan técnicas específicas para imágenes de dispositivos móviles es que muchas de ellas se basan en la extracción de características de aspectos relacionados con el sensor. En general, los sensores de las DSC utilizan la tecnología CCD (*Charge Coupled Device*), siendo estos sensores de mayor calidad que los que utilizados en cámaras de los dispositivos móviles, los cuales se utilizan la tecnología CMOS (*Complementary Metal Oxide Semiconductor*). Dada la alta calidad de muchos de los sensores de las DSC, las técnicas forenses que utilizan las características del sensor tienen un enfoque diferente al que se utiliza para las cámaras de dispositivos móviles. Asimismo, existen otros aspectos diferenciadores entre DSC y cámaras de dispositivos móviles que se deben tener en cuenta en las distintas técnicas. Algunos de estos aspectos son el sistema de lentes y filtros o el algoritmo de interpolación utilizado. En [1] puede verse una panorámica de las distintas investigaciones realizadas.

Para cualquier tipo de clasificación de imágenes, ya sea en escenarios abiertos o cerrados, se necesita obtener ciertas características que permitan a las técnicas de clasificación realizar su tarea. Según [2] se pueden establecer cuatro grupos de técnicas para este fin: basadas en la aberración de las lentes, basadas en la interpolación de la matriz CFA, basadas en las imperfecciones del sensor y basadas en el uso de las características de la imagen. Dentro de este último grupo puede hacerse una subdivisión en las basadas en características del color (*Color Features*), características de la calidad (*Quality Features*) y estadísticas del dominio *wavelet*. Este trabajo utiliza las técnicas basadas en las imperfecciones del sensor, concretamente aquellas basadas en el patrón de ruido del sensor *Sensor Pattern Noise* (SPN) el cual es originado por las imperfecciones en el proceso de fabricación de los semiconductores o las producidas por la utilización de la cámara en el día a día.

El objetivo del análisis de *clusters* o *clustering* es agrupar una colección de objetos en clases representativas llamadas *clusters*, sin información a priori, de forma que los objetos pertenecientes a cada *cluster* guarden una mayor similitud con respecto de objetos en otros *clusters*. La agrupación de imágenes puede llevarse a cabo mediante técnicas de aprendizaje supervisadas o sin supervisión. En el primer caso es indispensable conocer información del dispositivo a priori, es decir se identifica claramente con la clasificación en escenarios cerrados en donde se requiere una fase de entrenamiento con las características extraídas de las imágenes y una segunda fase de clasificación conforme al resultado anterior. Sin embargo, en un caso real puede ser difícil contar con la cámara en cuestión o con un subconjunto de fotografías tomadas por la misma para llevar a cabo un entrenamiento, de ahí la necesidad de técnicas de aprendizaje sin supervisión, que se corresponden directamente con los escenarios abiertos. El *clustering* tradicional se caracteriza por ser una técnica de aprendizaje sin supervisión.

Para poder determinar la similitud entre objetos pertenecientes a un mismo *cluster* existen medidas de distancia como pueden ser: distancia euclideana, distancia Manhattan y distancia Chebychev, entre otras. Alternativamente, es posible usar funciones de similitud $S(X_i, X_j)$ las cuales comparan dos vectores X_i y X_j en forma simétrica, es decir, $S(X_i, X_j) = S(X_j, X_i)$. Estas funciones alcanzan sus valores más altos cuando X_i y X_j son más similares. La medida más usada en la identificación de fuente de imágenes es la correlación normalizada [3], [4], [5] definida como:

$$corr\,(X_i, X_j) = \frac{\left(X_i - \overline{X_i}\right) \odot \left(X_j - \overline{X_j}\right)}{\left\|X_i - \overline{X_i}\right\| \cdot \left\|X_j - \overline{X_j}\right\|} \qquad (1)$$

Donde $\overline{X_i}$ y $\overline{X_j}$ representan la media del vector, $X_i \odot X_j$ es el producto punto de dos vectores y $\|X_i\|$ es la norma L_2 de X_i. Dado que el patrón de ruido del sensor es una matriz bidimensional, previamente a la aplicación de las funciones del cálculo de la correlación, se realiza una transformación a vector unidimensional.

De acuerdo a la clasificación de algoritmos de *clustering* propuesta en [6] encontramos los métodos jerárquicos cuyo propósito es lograr una estructura denominada dendograma que representa la agrupación de los objetos de acuerdo a sus niveles de similitud. Esta agrupación puede realizarse de distintas formas: aglomerativa o decisiva. La agrupación aglomerativa considera inicialmente a cada objeto como una clase independiente hasta, de forma iterativa, lograr agrupar todos los objetos en una clase única. La agrupación de forma divisiva se basa en la idea de partir de una sola clase hasta lograr separar todos los objetos en clases individuales. También existen los algoritmos de particionamiento en donde iniciando de una partición, el algoritmo se encarga de mover objetos de un *cluster* a otro hasta minimizar cierto criterio de error. Dentro de esta categoría el método más famoso es el k-means, sin embargo la mayoría de estos métodos requieren conocer de antemano el número de *clusters*, por lo cual no son muy utilizados en temas de análisis forense de imágenes. Por último, existen otros algoritmos de *clustering* como: [7] que produce *clusters* por medio de grafos, [8] basado en la densidad donde los puntos dentro de un *cluster* vienen dados por cierta función de probabilidad, *clusters* basados en modelos como árboles de decisión [9] o redes neuronales [10] y *clustering* con métodos de *soft-computing* como *fuzzy clustering* [11], métodos evolucionarios de *clustering* y recocido simulado en *clustering* [12].

Existen trabajos previos sobre agrupación de imágenes por métodos sin supervisión, todos ellos consideran al SPN como el criterio más fiable para representar la huella digital de un dispositivo, es de ahí que utilizan concretamente el PRNU (*Photo Response Non-Uniformity*) como huella y la correlación normalizada como medida de similitud para lograr el agrupamiento de imágenes por dispositivo.

En [13] se utiliza una técnica de clasificación con aprendizaje no supervisado donde mediante la maximización de grafos se logra una agrupación. El *clustering* se realiza a partir de grafos no dirigidos con pesos, comenzando con una matriz de afinidad donde los pesos de conexión entre vértices es el

valor de correlación entre cada SPN, iniciando con un nodo aleatorio. En cada iteración conectan los nodos restantes y eligen los nodos más cercanos al central obteniendo una nueva matriz de afinidad en cada paso, el algoritmo se detiene cuando el número de nodos más cercanos es menor a un parámetro k. Posteriormente el grafo es particionado hasta el punto en donde la similitud en un conjunto sea máxima y mínima con respecto a otros conjuntos.

En [4] se realizan agrupamientos mediante campos markovianos aleatorios. Se propone un algoritmo de *clustering* partiendo de una matriz que contiene todas las correlaciones entre SPN de diversas cámaras. En cada iteración el algoritmo agrupa dentro de clases los SPN más similares haciendo uso de las características locales de los campos markovianos aleatorios y asigna una nueva etiqueta de clase a cada SPN maximizando una función de probabilidad. El criterio para detener el algoritmo se cumple cuando no hay cambios en las etiquetas después de cierto número de iteraciones.

El algoritmo propuesto en [5], en el cual se basa esta investigación, utiliza *clustering* jerárquico para agrupar las imágenes. Previo al algoritmo de *clustering*, los autores aplican una función de mejora del ruido del sensor, que fortalece los componentes bajos y atenúa los componentes altos en el dominio wavelet, con la finalidad de eliminar los detalles de la escena en el mismo. Con una matriz de similitud que contiene todas las correlaciones entre los diferentes SPN y tomando como punto de partida a cada imagen como un *cluster* único, el algoritmo de *clustering* agrupa los dos *clusters* con un valor de correlación más alta formando un solo *cluster* y actualiza la matriz con una nueva fila y columna que vienen a sustituir las filas y columnas de los *clusters* agrupados. El criterio de enlace elegido para mezclar dos *clusters* fue el de enlace promedio. En cada iteración del algoritmo se almacena en una partición el estado de los *clusters* en ese momento y se calcula el coeficiente silueta global. Al final del algoritmo se elige la partición cuyo valor del coeficiente silueta sea el mínimo. En esa partición el número de *clusters* debería corresponderse con número de dispositivos que existen inicialmente, así como el contenido de cada *cluster* con los SPN de cada dispositivo. Los autores realizan una etapa de entrenamiento con el algoritmo descrito y una etapa de clasificación para las imágenes restantes. Para realizar esto basta obtener el promedio de los SPN por cada *cluster* y compararlos contra las imágenes restantes, la imagen se clasificará dentro del *cluster* cuya correlación sea más alta.

III. Descripción de la Técnica

El algoritmo de agrupación sin supervisión propuesto está basado en el presentado en [5]. Se trata de una combinación entre un *clustering* jerárquico y un *clustering* plano. Es decir, a pesar de formar una estructura de dendrograma con cada iteración del algoritmo, al final los *clusters* son tomados como entidades sin relación alguna ya que cada uno de ellos debe corresponder a un dispositivo específico.

Previo a realizar el *clustering*, es necesario obtener los patrones de ruido del sensor del conjunto de imágenes $I^{(i)},\ i = 1, ..., N$ utilizando el algoritmo de extracción y el parámetro de supresión de ruido $s_0 = 5$ propuestos en [14]:

$$n^{(i)} = I^{(i)} - F\left(I^{(i)}\right) \tag{2}$$

Donde n es el patrón de ruido de cada imagen i, I es el conjunto de imágenes con ruido del sensor y F es el filtro de extracción del ruido basado en la transformada wavelet. Para esto se utilizó el algoritmo desarrollado por Goljan et-al en [15]. En nuestra propuesta no se ha utilizado ningún algoritmo de mejoramiento de ruido, como los propuestos por [5] y [4]. El filtro de Wiener en el dominio de la frecuencia es suficiente para eliminar la mayoría de los detalles de la escena presentes al extraer el SPN.

Para cada uno de los N ruidos $(n_1, ..., n_N)$ se obtiene el valor de correlación usando la ecuación 1 y esto genera una matriz de similitud H de $N \times N$. Dicha matriz es simétrica y está compuesta de unos en su diagonal principal (ya que la correlación de un ruido consigo mismo es 1) . Una vez generada la matriz no será necesario volver a calcular las correlaciones entre ruidos a lo largo del algoritmo de *clustering* ahorrando tiempo y capacidad de procesamiento.

El algoritmo de *clustering* jerárquico seleccionado consiste en encontrar dentro de la matriz H el par de ruidos k y l con un valor de correlación más alto. Cabe mencionar que los valores de correlación en la diagonal principal no se toman en cuenta. A continuación las filas y columnas correspondiente a k y l son eliminadas y tanto una nueva fila como una nueva columna son agregadas a la matriz. Los valores de esta nueva fila y columna son el resultado de una función de criterio de enlace. La función elegida para este trabajo fue el criterio de enlace promedio puesto que sus resultados son más satisfactorios que con otros criterios de enlace como criterio simple o criterio completo, tal como se sugiere en [5]. La ecuación 3 muestra la función del criterio de enlace promedio entre dos *clusters* A y B.

$$H\left(A, B\right) = \frac{1}{\|A\| \, \|B\|} \sum_{ni \in A, nj \in B} corr\left(n_i, n_j\right) \tag{3}$$

Donde el valor $corr(n_i, n_j)$ se calcula con la ecuación 1 y puede ser tomado de la matriz H para simplificar el procesamiento computacional. $\|A\|$ y $\|B\|$ son la cardinalidad de los *clusters* A y B respectivamente.

Cada iteración del algoritmo toma los dos *clusters* con el valor de correlación más alto en la matriz y mezcla los objetos contenidos en éstos para crear un nuevo *cluster*, al mismo tiempo que almacena el estado de los distintos *clusters* en particiones $P_0, ..., P_{N-1}$ con el objetivo de conocer el contenido de los *clusters* en cada momento. En el *clustering* jerárquico, el resultado final del algoritmo es un *cluster* que contiene a todos los objetos. Sin embargo, en este trabajo para el agrupamiento de fotografías, cada *cluster* debería representar un dispositivo al final de la ejecución. Por este motivo se usó el coeficiente silueta como medida de validación de *clusters*. El coeficiente silueta mide el índice de similitud entre los elementos de un mismo *cluster* (cohesión) y la

similitud entre los elementos de un *cluster* con respecto a los demás (separación). A diferencia de Caldelli et al. [5] el cálculo del coeficiente silueta se realiza por cada *cluster* contenido en la partición P_i y no por cada patrón de ruido, como observamos en la ecuación 4.

$$s_j = \max(b_j) - a_j \tag{4}$$

donde, a_j (cohesión) es la correlación promedio entre todos los patrones de ruido dentro del *cluster* c_j. b_j (separación) es la correlación promedio de los patrones de ruido contenidos en el *cluster* c_j con respecto a los patrones de ruidos en los *clusters* restantes. Se toma el *cluster* vecino más cercano, es decir, aquel con la correlación más alta.

Para cada iteración q del algoritmo se obtiene una medida global de todos los coeficientes siluetas calculados a partir de los K *clusters*. Esto equivale a promediar los valores s_j en q. La ecuación 5 muestra dicho cálculo.

$$SC_q = \frac{1}{K}\sum_{j=1}^{K} s_j \tag{5}$$

Una vez concluido el *clustering* jerárquico se procede a buscar el SC_q con el valor mínimo, lo cual indica que los *clusters* de la partición P_q^* están en un nivel de correlación mayor. El número de *clusters* en ese instante debería corresponder al número real de dispositivos. El objetivo de almacenar la partición en cada momento del algoritmo es evitar volver a ejecutar el *clustering* ya que se tiene información de todos los *clusters* en cada iteración q.

En el Algoritmo 1 se muestra el pseudocódigo de la propuesta.

Algorithm 1: Algoritmo de clustering

1. Calcular el patrón de ruido $n^{(i)}$ de cada imagen donde $i \in 1, ..., N$;
2. Generar matriz de similitud $H \in R^{NxN}$;
3. **foreach** $q \in 1, ..., N-1$ **do**
4. Encontrar el par de *clusters* $H(k,l)$ con la mayor similitud;
5. Eliminar el par de filas y columnas correspondientes a los *clusters* k y l;
6. Calcular los valores del nuevo *cluster* usando el criterio de enlace promedio y agregar tanto la fila como columna correspondientes;
7. Determinar el coeficiente silueta global SC_q;
8. Almacenar la partición P_q;
9. Encontrar la partición donde el coeficiente silueta mínimo $min_q(SC_q)$;

IV. Experimentos y Resultados

Los experimentos fueron realizados con un conjunto total de 1050 fotografías de 7 modelos diferentes de cámaras de dispositivos móviles (Apple iPhone 5, Huawei U8815, Nokia 800 Lumia, Samsung GT-S5830M, LG E400, Sony ST25a y Zopo ZP980). Del conjunto total hay 150 fotografías de cada modelo.

Todas las imágenes fueron recortadas a 1024x1024 píxeles, poseen una orientación horizontal y son tanto de interiores como de exteriores con el objetivo de simular un escenario más realista. En la extracción del patrón de ruido de todas las imágenes se utilizó el promedio a cero (zero-mean) de filas y columnas, los 3 canales de color RGB fueron convertidos a una sola matriz de intensidades de grises, eliminando la información correspondiente al tono y la saturación, pero conservando la luminancia.

Para medir el grado de certeza en los resultados se utilizó la tasa de verdaderos positivos TPR (*True Positive Rate*). El TPR promedio para cada uno de los siguientes experimentos se calcula, computando para cada *cluster* el número de fotos que han sido bien clasificadas (TPR de cada *cluster*) y promediando los TPR de todos los *clusters* resultantes (si hay menos *clusters* que dispositivos se promedia teniendo en cuenta el número de dispositivos). Para calcular el TPR de cada *cluster*, hay que detectar en el *cluster* cual es el dispositivo que tiene el mayor número de imágenes con respecto al total de imágenes por dispositivo, siendo ese el *cluster* predominante del dispositivo, posteriormente hay que calcular el porcentaje de fotos que ha sido bien clasificadas para ese dispositivo en ese *cluster*. Realmente en la inmensa mayoría de los casos puede verse fácilmente que un *cluster* se asocia a uno o varios dispositivos como puede apreciarse en matrices de confusión de las Tablas I, II y III. Si hay varios *clusters* con el mismo número de fotos de un dispositivo o un *cluster* con igual número de fotos de varios dispositivos y a su vez éstos son los máximos, se toma como *cluster* predominante para el dispositivo el que se desee de entre las distintas opciones. Puede darse el caso que si hay un *cluster* de más, un *cluster* no sea predominante de ningún dispositivo (ver Tabla II) y su TPR para ese *cluster* sea 0. También puede que se forme un *cluster* menos (ver Tabla III), en este caso este se tendrá en cuenta la asociación del *cluster* al dispositivo y utilizar para el promedio el número de dispositivos como se indicó anteriormente.

Tabla I
TPR CON IGUAL NÚMERO DE DISPOSITIVOS QUE CLUSTERS

Marca - Modelo	Clusters					TPR
	1	2	3	4	5	promedio
Apple Iphone 5	49	0	0	1	0	
Huawei U8815	0	50	0	0	0	
LG E400	0	1	49	0	0	
Nokia 800 Lumia	0	0	0	50	0	
Samsung GT5830m	0	0	0	0	50	
TPR por *cluster*	98 %	100 %	98 %	100 %	100 %	99.2 %

En los resultados de los experimentos se consideran 3 posibles casos: a) Número de *clusters* identificados igual al número de dispositivos, b) número de *clusters* identificados mayor al número de dispositivos, y c) número de *clusters* identificados menor al número de dispositivos. Aunque el

Tabla II
TPR CON MENOR NÚMERO DE DISPOSITIVOS QUE CLUSTERS

Marca - Modelo	Clusters				TPR
	1	2	3	4	promedio
Apple I- phone 5	100	0	0	0	
Huawei - U8815	0	100	0	0	
LG - E400	0	0	97	3	
TPR por *cluster*	100 %	100 %	97 %	0 %	99 %

Tabla III
TPR CON MAYOR NÚMERO DE DISPOSITIVOS QUE CLUSTERS

Marca - Modelo	Clusters				TPR
	1	2	3	4	promedio
Apple Iphone 5	100	0	0	0	
Huawei U8815	0	100	0	0	
LG E400	0	0	100	0	
Nokia 800 Lumia	100	0	0	0	
Samsung GT 5830M	0	0	0	100	
TPR por *cluster*	100 %	100 %	100 %	100 %	80 %

primer caso es el ideal, el segundo caso también tiene un valor alto de TPR puesto el algoritmo deja ciertas imágenes desasociadas, es decir, fuera del *cluster* que les corresponde y estas no son consideradas en el TPR por no ser acierto. Por otro lado, en el último caso es donde se tienen porcentajes de acierto más bajo porque el algoritmo une dos o más dispositivos dentro de un mismo *cluster*.

Se realizaron varios experimentos para comparar los resultados entre recortar la imagen desde el centro o desde la esquina superior izquierda, teniendo este último criterio un TPR más alto. Una de las posibles razones por la que las dos zonas de recortes obtienen distintos resultados, es porque generalmente, las fotografías se toman enfocando en el centro el objeto de interés, el cual normalmente tiene un mayor grado de detalle. Este alto grado de detalle en el recorte de la imagen, en ciertos casos, puede dificultar la clasificación de la misma. La Tabla IV muestra el TPR en función del número distinto de dispositivos utilizados y el número de fotos utilizadas por dispositivo. Todos los dispositivos tienen el mismo número de fotos. En la Tabla IV se puede observar como el TPR aumenta en el caso del recorte en el centro a medida que se agrupan más dispositivos mientras que en el recorte por la esquina se mantienen buenos resultados.

Tabla IV
TPR EN FUNCIÓN DEL NÚMERO DISTINTO DE DISPOSITIVOS Y EL NÚMERO DE FOTOS POR DISPOSITIVO

Número de Fotos	Crop Corner			Crop Center		
	Número de Dispositivos			Número de Dispositivos		
	3	5	7	3	5	7
50	99.33 %	99.20 %	99.71 %	66.67 %	80 %	99.71 %
100	99 %	100 %	99.57 %	66.67 %	80 %	99.71 %

En un escenario cerrado no es muy probable contar con el mismo número de imágenes de cada dispositivo a identificar, por esa razón se realizaron experimentos en donde los conjuntos de imágenes por cada dispositivo no poseen una distribución simétrica para comprobar la adaptabilidad del algoritmo propuesto en un escenario real. En las Tablas V y VI se presentan los resultados obtenidos de agrupar las imágenes de 5 y 7 dispositivos respectivamente. El número de imágenes por dispositivo es variado y aún así podemos observar un muy alto grado de acierto (97.76 % TPR promedio de los experimentos de las Tablas V y VI).

Como se puede observar en los casos de número de imágenes asimétrico se ha experimentado con grupos de bastante disparidad numérica y en algunos casos con grupos pequeños (5 imágenes de un tipo de dispositivo), aun así se han logrado resultados de agrupación satisfactorios. Cabe destacar que existe una diferencia significativa en el resultado del experimento del grupo C de la Tabla VI, ya que se obtiene un TPR sensiblemente más bajo que el obtenido en el resto de experimentos. La causa a esta situación es que como se puede observar este experimento hay una cámara (Zopo Zp980) con una sola imagen. El hecho de que haya una sola imagen de un dispositivo hace que exista una alta probabilidad de que ese *cluster* no se genere correctamente, ya que sólo existen dos casos, la generación correcta al 100 % o la fusión de este *cluster* con otro. Concretamente en este experimento la imagen del Zopo Zp980 no se ha clasificado correctamente como *cluster* independiente y se ha fusionado con el *cluster* del Huawei U8815, bajando considerablemente el TPR. Simplemente para este experimento si se hubiera clasificado en un *cluster* independiente esa única imagen, el TPR habría sido del 99,71 %. Como puede observarse para *clusters* con una única imagen si se da una clasificación incorrecta el TPR baja sensiblemente, ya que esa única imagen hace que el TPR parcial del *cluster* sea del 0 %, aunque la práctica totalidad de las imágenes se hayan clasificado correctamente.

Tabla V
TPR PARA CLUSTERING ASIMÉTRICO DE 5 DISPOSITIVOS

Grupo	Apple Iphone 5	Huawei U8815	LG E400	Nokia 800 Lumia	Samsung GT 5830m	TPR
A	100	95	90	85	80	99.78 %
B	50	45	40	35	30	99.1 %
C	100	75	50	25	10	99.6 %
D	100	30	20	10	5	99 %

Tabla VI
TPR PARA CLUSTERING ASIMÉTRICO DE 7 DISPOSITIVOS

Grupo	Apple Iphone 5	Huawei U8815	LG E400	Nokia 800 Lumia	Samsung GT 5830m	Sony ST25a	Zopo Zp980	TPR
A	100	95	90	85	80	75	70	99.84 %
B	50	45	40	35	30	25	20	99.36 %
C	100	75	50	25	10	5	1	85.43 %
D	100	50	40	30	20	10	5	99.21 %

V. CONCLUSIONES

En este trabajo se ha realizado un análisis de las principales técnicas de agrupación de imágenes sin supervisión, siendo estas de suma importancia en el análisis forense de imágenes

digitales. A pesar del auge que han tenido las cámaras de dispositivos móviles en estos tiempos, aún no existen en el estado del arte muchas referencias para la agrupación no supervisada de imágenes de dispositivos móviles. La mayor parte de los trabajos se refieren a la clasificación supervisada y en muchos casos no se centran en imágenes de dispositivos móviles, las cuales tienen características peculiares. La comparación de los resultados de este trabajo con los de otros trabajos del estado del arte no puede realizarse de forma precisa, ya que en los mismos no se hace referencia al número final de *clusters* generados, lo cual es un tema fundamental. Además en estos trabajos no se detalla como se han calculado las tasas de acierto, ni se hace referencia a las mismas cuando los *clusters* generados por la clasificación son diferentes en número a la cantidad dispositivos utilizados, haciendo esto que la comparativa de sus tasas con respecto a nuestra interpretación del TPR carezca de sentido. El ruido agregado en cada fotografía por el sensor de la cámara, debido a las fallos en el proceso de fabricación de este o defectos por el uso diario, ha demostrado ser una fuente fiable de identificación de un dispositivo. Asimismo, el cálculo de correlación normalizada entre ruidos de sensor extraídos de dos o más fotografías es una medida de similitud bastante utilizada en las técnicas de aprendizaje sin supervisión de imágenes, siendo las técnicas de *clustering* aquellas que tienen mejores resultados.

El algoritmo de esta propuesta está basado en la combinación de un *clustering* jerárquico y un *clustering* plano para la separación entre *clusters*. El uso del coeficiente silueta para la validación de los *clusters* demostró dar buenos resultados al obtener elevados TPR, también el número de *clusters* correspondió al número de dispositivos reales en la mayoría de los casos.

El porcentaje de aciertos al utilizar el recorte de la imagen desde la esquina izquierda era más estable que aquellos recortados por el centro, pese a encontrar diferentes observaciones en la literatura argumentando la saturación y ausencia de iluminación encontrada en esas regiones.

Los experimentos realizados en este trabajo han permitido comprobar gran diversidad de situaciones con respecto a la simetría o no de los conjuntos del fotos, el tamaño de los mismos, el número de dispositivos utilizados y el uso de dispositivos de la misma marca. Tras todos los experimentos realizados se concluye que los resultados de la aplicación de la técnica son buenos (98.01 % TPR promedio de todos los experimentos realizados).

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] A. L. Sandoval Orozco, D. M. Arenas González, J. Rosales Corripio, L. J. García Villalba, and J. C. Hernandez-Castro, "Techniques for Source Camera Identification," in *Proceedings of the 6th International Conference on Information Technology*, May 2013, pp. 1–9.

[2] T. Van Lanh, K. S. Chong, S. Emmanuel, and M. S. Kankanhalli, "A Survey on Digital Camera Image Forensic Methods," in *Proceedings of the IEEE International Conference on Multimedia and Expo*. IEEE, July 2007, pp. 16–19.

[3] J. Fridrich, "Digital Image Forensics," *IEEE Signal Processing Magazine*, vol. 26, no. 2, pp. 26–37, March 2009.

[4] C.-T. Li, "Unsupervised Classification of Digital Images Using Enhanced Sensor Pattern Noise," in *Proceedings of the IEEE International Symposium on Circuits and Systems*. IEEE, May 2010, pp. 3429–3432.

[5] R. Caldelli, I. Amerini, F. Picchioni, and M. Innocenti, "Fast Image Clustering of Unknown Source Images," in *Proceedings of the IEEE International Workshop on Information Forensics and Security*. IEEE, December 2010, pp. 1–5.

[6] L. Rokach, "A Survey of Clustering Algorithms," in *Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach, Eds. Springer US, 2010, pp. 269–298.

[7] C. Zahn, "Graph-Theoretical Methods for Detecting and Describing Gestalt Clusters," *IEEE Transactions on Computers*, vol. C-20, no. 1, pp. 68–86, January 1971.

[8] J. D. Banfield and A. E. Raftery, "Model-Based Gaussian and Non-Gaussian Clustering," *Biometrics*, vol. 49, no. 3, pp. pp. 803–821, September 1993.

[9] D. Fisher, "Knowledge Acquisition Via Incremental Conceptual Clustering," *Machine Learning*, vol. 2, no. 2, pp. 139–172, 1987.

[10] J. Vesanto and E. Alhoniemi, "Clustering of the Self-Organizing Map," *IEEE Transactions on Neural Networks*, vol. 11, no. 3, pp. 586–600, May 2000.

[11] F. Hoppner, *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*, ser. Jossey-Bass higher and adult education series. Wiley, 1999.

[12] S. Z. Selim and K. Alsultan, "A Simulated Annealing Algorithm for the Clustering Problem," *Pattern Recogn.*, vol. 24, no. 10, pp. 1003–1008, Oct. 1991.

[13] B.-b. Liu, H.-K. Lee, Y. Hu, and C.-H. Choi, "On Classification of Source Cameras: A Graph Based Approach," in *Proceedings of the IEEE International Workshop on Information Forensics and Security (WIFS),*. IEEE, December 2010, pp. 1–5.

[14] 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.

[15] M. Goljan, J. Fridrich, and T. Filler, "Large Scale Test of Sensor Fingerprint Camera Identification," in *Proceedings of the SPIE on Media Forensics and Security*, vol. 7254. International Society for Optics and Photonics, February 2009, pp. 72 540I–72 540I.