Análisis de la Estructura del Contenedor de Vídeos Digitales de Dispositivos Móviles para Identificación de la Fuente de Vídeos en Escenarios Abiertos

Raquel Ramos López, Elena Almaraz Luengo, Ana Lucila Sandoval Orozco, Luis Javier García Villalba*, *Member, IEEE*

***Resumen*—La ciencia forense se ha servido de la tecnología multimedia para analizar, evidenciar e incluso dilucidar responsabilidades en los procesos judiciales. El análisis de vídeos digitales adquiere especial relevancia al permitir determinar tanto el origen como la autenticidad de un material y de relacionar a un individuo con un dispositivo, lugar o evento. El constante desarrollo de la tecnología hace que, a pesar de que los principios básicos sean inalterables, el análisis de vídeos digitales requiera, en el ámbito forense, de nuevos procedimientos y herramientas de enfoque. Por consiguiente, es necesario proporcionar al analista forense técnicas para identificar el contenido multimedia. En este trabajo se estudia el problema de la identificación de la fuente de vídeos en escenarios abiertos, esto es, aquéllos en los que no se conozca a priori el conjunto de cámaras a las que pertenezca el vídeo a fin de identificar su fuente, hecho éste que se produce en casos reales. En particular, se propone un algoritmo de identificación de la fuente de adquisición de vídeos digitales generados por los dispositivos móviles usando algoritmos no supervisados basados en el análisis de la estructura del contenedor de vídeo.**

***Palabras claves*—Análisis de Conglomerados, Análisis Forense, Átomos, Contenedor de vídeo, Estructura del Contenedor, Identificación de fuente**

I. Introducción

El análisis forense de dispositivos móviles se ha convertido en una de las áreas de investigación más importantes. En primer lugar, las capacidades de los dispositivos inteligentes han mejorado sustancialmente, siendo más utilizados que los portátiles ya que los usuarios los tienen a su alcance en cualquier momento, registrando constantemente sus actividades y movimientos proporcionando una visión del comportamiento del usuario [1].

La combinación de teléfonos móviles inteligentes con plataformas de social media y almacenamiento en la nube ha permitido que el vídeo se convierta en una importante fuente de información. Estos vídeos digitales se pueden realizar en cualquier momento y lugar para diferentes propósitos y distribuir en Internet en un corto período de tiempo. En ocasiones su contenido puede estar relacionado con actos ilegales como terrorismo, pornografía infantil, espionaje industrial, etc.

R. Ramos López Securitas Direct (rlopez@securitasdirect.es). R. Ramos López, E. Almaraz Luengo, A. L. Sandoval Orozco and L. J. 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, España. {raqram01, ealmaraz}@ucm.es, {asandoval, javiergv}@fdi.ucm.es.

Para abordar estos problemas, los investigadores han desarrollado algoritmos forenses que verifican la autenticidad y la fuente del contenido digital [2]. Las técnicas forenses que identifican información sobre la fuente cuando se genera contenido multimedia (imágenes o vídeos), se dividen en dos grupos: por un lado, aquéllas orientadas a la verificación del origen de un contenido multimedia y aquéllas destinadas a la detección de inconsistencias en la fuente dentro del contenido multimedia [2], [3].

Existen diversas investigaciones que desarrollan algoritmos forenses para determinar la identificación de la fuente de una imagen aunque los estudios son escasos en el caso de los vídeos digitales [3]. En [4] se sugiere que estos algoritmos utilizan trazas dejadas por una amplia variedad de componentes físicos y algorítmicos en la tubería de procesamiento de una cámara. Se han diseñado algoritmos de modelos de cámaras forenses que aprovechan las huellas dejadas por la demosaicing (método de los artefactos CFA y Demosaicing), [5], [6], [7], [8], [9] e información del encabezado JPEG [10]. La mayor parte del trabajo existente se ha centrado en el uso de huellas dactilares de sensores para identificar el dispositivo fuente específico de un vídeo, por ejemplo, ver [11], [12], [13], [14], [15] [16].

El análisis de la fuente de adquisición de vídeo es uno de los primeros problemas que han surgido en las técnicas de análisis forense. Dentro de la identificación de la fuente de adquisición existen dos enfoques principales: escenarios cerrados o escenarios abiertos. Un escenario cerrado es aquél en el que la identificación de la fuente del vídeo se realiza en un conjunto de cámaras específicas y conocidas. Para este enfoque, normalmente se utiliza un conjunto de vídeos de cada dispositivo para formar a un clasificador y, posteriormente, se predice la fuente de adquisición de los vídeos que se están investigando.

En [17] se presenta un esquema de identificación de fuentes de vídeo digital basado en máquinas de soporte vectorial (SVM) y ruido PRNU. Con un vídeo de entrada, los fotogramas con cambios de escena más significativos se extraen utilizando el histograma de color. Un total de 81 funciones, que son los componentes Wavelet del sensor, se utilizan para entrenar al clasificador SVM con vídeos de entrenamiento. Un total de 5 dispositivos diferentes de 5 marcas diferentes fueron utilizados para entrenar al clasificador SVM. Los resultados obtenidos muestran una tasa de éxito del 87 % o del 90 %, dependiendo de la resolución del vídeo.

https://doi.org/10.12804/si9789587844337.08

En un escenario abierto, no se conoce inicialmente el conjunto de dispositivos a los que pertenecen los vídeos para identificar su fuente de adquisición. El objetivo no es identificar la marca y el modelo de los vídeos, sino poder agruparlos en grupos en los que todos sus vídeos pertenezcan al mismo dispositivo. Este último enfoque es más realista, ya que en muchos casos el analista ignora completamente el conjunto de dispositivos a los que puede pertenecer un conjunto de vídeos. Identificar el dispositivo que genera el contenido digital es muy importante en el contexto de un proceso judicial porque puede incriminar o delimitar responsabilidades a un sospechoso antes de un acto delictivo. Para realizar cualquier tipo de clasificación de vídeo en escenarios abiertos o cerrados, es necesario obtener ciertas características que permitan a las técnicas de clasificación realizar su tarea. Dentro del análisis forense de vídeo digital, las principales técnicas de análisis se dividen en 5 grupos: 1) Metadatos, 2) Características de la imagen, 3) Defectos de la matriz CFA, 4) Imperfecciones del sensor y 5) Características de los contenedores multimedia.

Este trabajo propone una técnica de identificación de la fuente de adquisición de vídeos digitales generados por dispositivos móviles mediante el uso de algoritmos no supervisados basados en el análisis de la estructura de los contenedores de vídeo multimedia. Se ha dividido en 5 secciones, la primera consistente en esta introducción. En la Sección II se presentan las principales técnicas actuales de análisis en este contexto. El algoritmo propuesto en este trabajo se desarrolla en la Sección III. En la Sección IV se muestran los resultados numéricos en los que se ha aplicado el algoritmo diseñado. Finalmente, en la Sección V se presentan las conclusiones extraídas de este trabajo.

II. Herramientas de Análisis Forense de Vídeo

Un vídeo está formado por una secuencia de imágenes llamadas fotogramas que varían con el tiempo dando una sensación de movimiento. Debido al gran volumen de información que tiene un vídeo, éste se codifica y decodifica mediante un algoritmo matemático conocido como códec. A su vez, estas tramas ya codificadas se encapsulan junto con las pistas de audio, metadatos y subtítulos en un único archivo conocido como contenedor multimedia. En la Tabla I se muestran los diferentes elementos que componen un contenedor multimedia.

Tabla I
Elementos Compositivos de un Contenedor Multimedia.

Formato del contenedor: .avi, .mp4, .mov, .ogg, .flv, .mkv, etc			
Video códec	Audio códec	Captioning descripción de video	Metadatos
H.264.	AAC	SAMI	MPEG-7
VC-1	WMA	SMIL	CableLabs
Theora	Vorbis	Hi-Caption	TV-ANytime
Dirac 2.1	PCM, etc.	CMML	EBU
H.263, etc.		DFXP	XPM, etc.
		3GPP TS	
		MPSub, etc.	

Los contenedores multimedia o formatos de vídeo se definen como aplicaciones informáticas capaces de almacenar audio y vídeo y, en algunos casos, también subtítulos y otra información adicional.

Los contenedores multimedia más utilizados en la actualidad son:

- AVI (Audio Video Interleave): contenedor multimedia estándar de Windows.
- MP4: contenedor estandarizado para MPEG4 [18].
- FLV (Flash Video): formato utilizado para entregar vídeo MPEG a través de Flash Player.
- MKV (Mastroska): contenedor de especificaciones abierto orientado a animación.
- MOV: formato de contenedor QuickTime de Apple.
- OGG, OGM, OGV: contenedores estándar abiertos.

En la literatura más reciente, se puede encontrar que la mayoría de las investigaciones analizan la estructura interna de los contenedores multimedia en el caso del formato AVI, siendo casi inexistente el estudio de los contenedores MP4, 3GP y MOV.

Uno de los primeros trabajos donde se realiza un análisis detallado de las estructuras de los vídeos es [19] donde se analizan en detalle las secuencias de vídeo AVI y MP4 (MOV, 3GP, MP4) de teléfonos móviles y cámaras digitales. Uno de los principales resultados que se obtienen es que los vídeos de cámaras digitales y teléfonos móviles suelen emplear diferentes formatos de contenedores y codecs de compresión. Los teléfonos móviles optan por sofisticados algoritmos de compresión (MP4V, H.26x). La mayoría de las cámaras digitales de nuestro equipo de prueba prefieren una combinación de contenedores AVI y compresión básica MPEG. La estructura de los contenedores tipo AVI y MP4 no está estrictamente definida. Observaron diferencias considerables tanto en el orden como en la presencia de segmentos de datos individuales. Los archivos AVI a menudo contienen listas de información específicas o trozos de JUNK. Los archivos de tipo MP4 pueden emplear varios átomos no estándar y diferentes parametrizaciones de entradas de átomos específicos. La edición de vídeo sin pérdida deja intactos los ajustes de compresión del flujo de vídeo original, pero introduce sus propios artefactos distintivos en la estructura de los archivos de contenedor. Mientras que las peculiaridades del formato de archivo del dispositivo fuente original se pierden normalmente después de la edición de vídeo, todas las herramientas de software probadas tienen firmas de formato de archivo únicas en todo su conjunto de pruebas.

En [20] se introduce un método para el análisis no supervisado de contenedores de archivos de vídeo y sus autores presentan dos aplicaciones forenses principales de dicho método: la verificación de la integridad del vídeo (basado en la diferencia entre un contenedor de archivos de referencia y uno de consulta) y la identificación y clasificación de la marca del dispositivo de origen (basado en el análisis de la estructura y contenido de los contenedores). Se comprobó la eficacia de ambas aplicaciones en un conjunto de datos compuesto por 578 vídeos tomados con smartphones modernos de las principales marcas y modelos y se llegó a la conclusión de que la solución propuesta ofrece un coste computacional extremadamente bajo en comparación con todas las técnicas disponibles basadas en el análisis del flujo de vídeo o la inspección manual de los contenedores de archivos.

En [21] se investiga el contenido de vídeo almacenado en Video Event Data Recorders (VEDRs). En concreto, se estudia la estructura de los archivos de vídeo para cada tipo de software de edición de vídeo que dejaría huellas del procesamiento del software de edición de vídeo. Debido a que tales trazas son una característica inherente a cada paquete de software de edición de vídeo, pueden detectar el software de edición de vídeo específico utilizado para manipular el vídeo, además de si el vídeo fue, de hecho, manipulado. Para evaluar la precisión de su técnica, examinaron 296 archivos de vídeo no modificados de Audio Video Interleave (AVI). Se realizó este examen utilizando versiones populares de software de edición de vídeo. Como resultado, se encontró que las estructuras de datos AVI en los archivos de vídeo modificados aparecen consistentemente de acuerdo con cada paquete de software de edición de vídeo. Cada estructura de datos resultante no se ve afectada por la estructura del archivo de vídeo original.

III. Descripción del Algoritmo Propuesto

Este artículo presenta una técnica para identificar la fuente de adquisición de vídeo digital generada por los dispositivos móviles. La técnica se divide en dos tareas: 1) Extracción de la información de los átomos contenidos en cada uno de los vídeos y 2) Agrupación mediante técnicas de análisis de conglomerados de cara a la identificación del vídeo por modelo.

Para la realización de las tareas se han tenido en cuenta los formatos de contenedores MOV basados en los estándares QuickTime de Apple QuickTime [22], MP4 y 3GP compatibles con el estándar ISO / IEC 14496 Parte 12 [18]. Se han analizado los contenedores que existen tanto en la plataforma principal de compartición de vídeo, YouTube, como en una de las principales plataformas de mensajería de WhatsApps porque se encuentran dentro de uno de los conjuntos de datos utilizados para este trabajo [23].

III-A. Análisis de la Estructura del Contenedor

La estructura elemental de un vídeo es el átomo. Los metadatos, el vídeo y el sonido de un vídeo están dentro de ellos. Los átomos son de naturaleza jerárquica. Es decir, un átomo puede contener otros átomos, que pueden contener otros y así sucesivamente. El tipo de átomo se especifica mediante un número entero sin signo de 32 bits, normalmente interpretado como un código ASCII (American Standard Code for Information Interchange) de cuatro caracteres, generalmente en minúsculas. Debe tenerse en cuenta que no hay ninguna regla respecto a los átomos que deben aparecer y su orden, sin embargo, la mayoría sigue una estructura similar [24]. Este algoritmo se ha utilizado para extraer información de los átomos. Esta solución es capaz de analizar información múltiple de cualquier formato de vídeo como: Vídeo MP4 / H.264, MOV y 3GP, para extraer información de los átomos.

La extracción de átomos consiste en almacenar las etiquetas, valores, orden de aparición de los átomos y todo tipo de información relevante de un vídeo digital generado por un dispositivo móvil. El proceso comienza con la obtención del byte inicial del átomo, tamaño y tipo de átomo con una longitud máxima de 4 bytes formados como una cadena de caracteres (ejemplo: fytp, moov, mdat, etc.). A continuación se verifica la duplicidad de los átomos y la existencia de átomos hijos. Finalmente, se obtiene un diccionario de un conjunto de átomos y etiquetas (Path-tag) con sus respectivos valores y órdenes de apariencia. Para un estudio más profundo de los átomos, ver [19] y [24].

La Tabla II muestra el output que se obtiene tras la extracción de los átomos. El primer átomo es *"fytp"* tal y como se indica en las especificaciones de [19]. Como los átomos se organizan jerárquicamente *(ie./moov/)*, a su vez tienen átomos hijos *(ie./moov/trak)* y etiquetas (ie*./moov/mvhd/tkhd/flags)* que a su vez contienen valores (ie*./moov/mvhd/tkhd/version*,*value:0)*.

Tabla II
Muestra de la Información Extraída del Contenedor

Path	PathOrder	Field	Value
ftyp	ftyp-1	majorBrand	mp42
ftyp	ftyp-1	minorVersion	1
ftyp	ftyp-1	compatibleBrands	mp41mp42isom
beam	beam-2	byteInitial	28
beam	beam-2	size	42
moov/trak/tkhd	moov-4/trak-2/tkhd-1	version	0
moov/trak/tkhd	moov-4/trak-2/tkhd-1	flags	1
moov/trak/tkhd	moov-4/trak-2/tkhd-1	trackId	1
moov/trak/tkhd	moov-4/trak-2/tkhd-1	trackWidth	48
moov/trak/tkhd	moov-4/trak-2/tkhd-1	trackHeight	848

En esta propuesta, para realizar la agrupación de vídeos se ha tenido en cuenta que los vídeos son conjuntos de elementos que contienen las siguientes características:

- **PathField**: Se define como la unión de las etiquetas Path y Field separadas por el carácter ('/'). En la Tabla II, el campo PathField the la primera fila corresponde al valor *(ftyp/majorBrand)*.
- **PathFieldValue**: Se define como la unión de las etiquetas Path y Field separadas por '/' y se añade la etiqueta Value separada por '='. En la Tabla II, el campo PathFieldValue de la primera fila corresponde a: *ftyp/majorBrand = mp42*.
- **PathOrderField**: Se define como la unión de las etiquetas PathOrder y Field separadas por '/'. En la Tabla II, el campo PathOrderField de la primera fila corresponde con: *ftyp-1/majorBrand*.
- **PathOrderFieldValue**: Es la unión de las etiquetas PathOrder y Field separadas por '/' al que se le añade la etiqueta Value separada por '='. En la Tabla II, el campo PathOrderFieldValue de la primera fila se corresponde con: *ftyp-1/majorBrand = mp42*.

III-B. Técnicas de Análisis de Conglomerados

El análisis de conglomerados es un conjunto de técnicas para agrupar las observaciones por afinidad. Esta es la razón por la que esta técnica ha sido tradicionalmente considerada como parte de la Estadística Multivariante, aunque actualmente tiende a ser catalogada como Minería de Datos. En estas técnicas no se hacen suposiciones sobre el número de grupos o la estructura del grupo. La agrupación se hace sobre la base de similitudes o distancias (diferencias).

Cuando los artículos (unidades, cajas) están agrupados, la proximidad es usualmente indicada por algún tipo de de distancia.

Definición 1: La distancia $d(P,Q)$ entre dos puntos P y Q satisface las siguientes propiedades, donde R es otro punto intermedio:

1. $d(P,Q) = d(Q,P)$.
2. $d(P,Q) > 0$; si $P \neq Q$.
3. $d(P,Q) = 0$; si $P = Q$.
4. $d(P,Q) \leq d(P,R) + d(R,Q)$.

La tercera condición de la definición 1 se conoce como desigualdad triangular. Aquellas medidas que verifican las condiciones 1 y 2 de la definición 1 pero no la tercera, se conocen como semimétricas. Existen muchos tipos de distancia que se utilizan en el análisis de conglomerados [25]. Sean $\boldsymbol{x'} = (x_1, ..., x_p)$ y $\boldsymbol{y'} = (y_1, ..., y_p)$.

- Distancia Euclídea: $d(\boldsymbol{x},\boldsymbol{y}) = \sqrt{(\boldsymbol{x}-\boldsymbol{y})'(\boldsymbol{x}-\boldsymbol{y})}$.
- Distancia Estadística: $d(\boldsymbol{x},\boldsymbol{y}) = \sqrt{(\boldsymbol{x}-\boldsymbol{y})'\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{y})}$. Generalmente, $\boldsymbol{A} = \boldsymbol{S}^{-1}$, donde $\boldsymbol{S}$ contiene las varianzas y covarianzas muestrales. Sin embargo, sin tener un conocimiento previo de los diferentes grupos, esas cantidades muestrales no se pueden computar, es por ello que en el análisis de conglomerados se prefiere usar la distancia Euclídea.
- Métrica de Minkowski: $d(\boldsymbol{x},\boldsymbol{y}) = \left(\sum_{i=1}^{p} |x_i - y_i|^m\right)^{1/m}$. Para $m = 1$, $d(\boldsymbol{x},\boldsymbol{y})$ mide la distancia en manzanas de ciudad (city-block) entre dos puntos en dimensión p. Para $m = 2$, $d(\boldsymbol{x},\boldsymbol{y})$ es la distancia Euclídea.
- Otras: Métrica de Canberra, coeficiente de Czekanowski (para el caso de variables no negativas).

Por otro lado, cuando los items no se pueden representar de manera adecuada por una medida p-dimensional, cada par de ítem se comparan en base a si tienen o no determinada característica. En este caso, se utilizan variables binarias en las que 1 indicará la presencia de la característica y 0 su ausencia. Esta situación se representa mediante una tabla de contingencia.

Las técnicas de análisis de conglomerados se pueden clasificar en dos clases: métodos jerárquicos y no jerárquicos. En este artículo se usarán técnicas de clasificación jerárquicas.

Los métodos jerárquicos proceden, bien de una serie de fusiones sucesivas, bien de una serie de escisiones sucesivas:

- Métodos jerárquicos aglomerativos: parten de los elementos individuales y los añaden en grupos.
- Métodos jerárquicos divisivos: parten de un conjunto formado por todos los elementos y lo van dividiendo sucesivamente hasta llegar a alcanzar los elementos individuales.

Los algoritmos aglomerativos tienen siempre la misma estructura y solamente difieren en la manera en que calculan la distancia entre grupos. Su estructura se muestra en el Algoritmo 1.

Algoritmo 1: Algoritmo aglomerativo

1. Comenzar con N conglomerados, cada uno conteniendo un ítem y una matriz de dimensión $N \times N$ de distancias (o similitudes) entre items $\boldsymbol{D} = (d_{ik})$;
2. Seleccionar los dos elementos más cercanos según las distancias actuales en $\boldsymbol{D}$ y formar con ellos una nueva clase;
3. Sustituir los dos elementos utilizados en (2) para definir la clase por un nuevo elemento que represente la clase construida. Las distancias entre este nuevo elemento y el elemento se calculan utilizando uno de los criterios que se exponen a continuación;
4. Volver al paso (2) y repetir (2) y (3) hasta conseguir una única clase;

Como se mencionó anteriormente, existen diferentes criterios para medir distancias. Los más comunes se pueden ver en la Tabla III en la que $d(u,v)$ representa la distancia entre los elementos u y v, d_{UW} y d_{VW} son las distancias entre los vecinos más próximos de los conglomerados U y W y V y W respectivamente y $|U|$ y $|V|$ representan el cardinal de U y V respectivamente.

El resultado de los métodos aglomerativos y divisivos se puede representar en un diagrama bidimensional conocido como dendograma que muestra las sucesivas uniones (o divisiones) que se han realizado en cada iteración del algoritmo. Una vez que se obtiene el dendograma, se pueden obtener los diferentes conglomerados.

Otra metodología de análisis de conglomerados en el espacio es la basada en la densidad de aplicaciones con ruido, conocida usualmente con sus siglas en inglés DBSCAN (Density-Based Spatial Clustering of Applications with Noise). Esta metodología fue propuesta por Martin Ester, Hans-Peter Kriegel, Jörg Sander y Xiaowei Xu en 1996 [26].

Considérese un conjunto de puntos que se desea agrupar en un determinado espacio. El algoritmo DBSCAN clasifica los puntos en tres tipos: puntos núcleo, puntos densamente-alcanzables o ruido.

Definición 2: (Puntos núcleo, directamente alcanzable y ruido)

- Un punto p es un punto núcleo si al menos $minPoints$ puntos están a una distancia ϵ de él (incluyendo p), y esos puntos son directamente alcanzables desde él. No es posible tener puntos directamente alcanzables desde un punto que no sea un núcleo.

Tabla III
CRITERIOS MÁS COMUNES DE VINCULACIÓN EN ANÁLISIS JERÁRQUICO DE CONGLOMERADOS

Criterio	**Descripción: la distancia entre dos grupos nuevos es la**	**Fórmula**
Encadenamiento simple	menor de las distancias entre grupos antes de la fusión	$d_{(UV),W} = \min\{d_{UW}, d_{VW}\}$
Encadenamiento completo	mayor de las distancias entre los grupos antes de la fusión	$d_{(UV),W} = \max\{d_{UW}, d_{VW}\}$
Media ponderada	semisuma de las distancias entre grupos antes de la fusión	$d((U \cup V, W)) = \frac{d(U,W)+d(V,W)}{2}$
Promedio no ponderado	dist. media entre todos los pares (i,j) con i en un grupo y j en otro	$\frac{1}{\lvert U\rvert\lvert V\rvert}\sum_{u\in U}\sum_{v\in V} d(u,v)$

- Un punto q es directamente alcanzable desde p si existe una secuencia de puntos $p_1, \ldots, p_n$ donde $p_1 = p$ y $p_n = q$, tales que, cada punto p_{i+1} es directamente alcanzable desde p_i.
- Un punto que no sea alcanzable desde cualquier otro punto es considerado ruido.

Si p es un punto núcleo, éste forma un conglomerado con otros puntos (núcleo o no) que sean alcanzables desde él.

Cada conglomerado contiene al menos un punto núcleo. Los puntos no núcleos alcanzables pueden pertenecer a un conglomerado pero actúan como una barrera puesto que no es posible alcanzar más puntos desde estos. Obsérvese que la propiedad de ser alcanzable no es simétrica.

Por definición, ningún punto puede ser alcanzable desde un punto que no sea núcleo, independientemente de la distancia a la que se encuentre, por tanto es necesario definir el concepto de conectividad para especificar la noción de conglomerado usada en el algoritmo DBSCAN.

Definición 3: Dos puntos p y q están conectados densamente si existe otro punto o tal que p y q son directamente alcanzables desde o. Esta propiedad es simétrica.

El algoritmo DBSCAN utiliza dos parámetros: ϵ (que especifica cuán cerca deben estar los puntos entre sí para ser considerados parte de un mismo conglomerado) y $minPoints$ (que representa el número mínimo de puntos para formar una región densa). Los pasos fundamentales de esta metodología se muestran en el Algoritmo 2.

Algoritmo 2: Algoritmo DBSCAN

1. Para cada punto p_i calcular la distancia $d(p_i, p_j)$, $\forall j \neq i$. Encontrar todos los puntos vecinos en un radio de ϵ del punto de partida p_i. Cada punto, con un vecino cuya distancia sea mayor o igual que $minPoints$, está marcado como punto núcleo o punto visitado;
2. Para cada punto núcleo que aún no haya sido asignado a un conglomerado, crear un nuevo conglomerado. Encontrar de manera iterativa todos los puntos conectados densamente y asignarlos al mismo conglomerado que el punto núcleo;
3. Iterar a través de los puntos no visitados restantes en el conjunto de datos;

El algoritmo OPTICS (Ordering Points to Identify the Clustering Structure) se puede considerar como una generalización del algoritmo DBSCAN en el caso de múltiples rangos, reemplazando el parámetro ϵ por el radio máximo de búsqueda. Fue presentado por Michael Ankerst, Markus M. Breunig, Hans-Peter Kriegel y Jörg Sander en 1999 [27]. La idea principal es similar a la de DBSCAN pero aborda el problema de detectar conglomerados significativos en datos de densidad variable, en concreto, OPTICS calcula un orden de los puntos aumentados por información adicional, es decir, la distancia de alcanzabilidad, que representa la estructura jerárquica intrínseca del conglomerado. A continuación mostraremos la propuesta específica de este trabajo.

- Extracción y almacenamiento de datos: para extraer la información de los átomos se utilizará una solución escrita en Python y desarrollada por el Grupo de Análisis, Seguridad y Sistemas (GASS) del departamento de Ingeniería del Software e Inteligencia Artificial de la Universidad Complutense de Madrid. Esta solución es capaz de analizar información múltiple de cualquier vídeo MP4/H.264, MOV y 3gp, para extraer la información de los átomos. Los valores del interior han sido convertidos a utf-8 cuando ha sido posible y de lo contrario permanecen como hexadecimales.
- Recopilar los datos: con la información obtenida, consultar la base de datos para descargar los vídeos necesarios. Posterior agrupación de los mismos para obtener un único cuadro de datos.
- Vectorización: agrupación de los datos con el objetivo de que el uso posterior de los mismos sea más abarcable y facilite la manipulación de los mismos.
- Filtrado: mediante pandas, filtrar las columnas correspondientes a los átomos y campos no deseados. Este filtrado es una extensión de la realizada en [20].
- Procedimiento de Análisis de Conglomerados: recuperar la matriz de valores binarios del marco de datos que representa el conjunto de datos. Cada línea contiene una observación, un archivo de vídeo y dimensiones de columna. Para agrupar los datos, se proponen dos algoritmos de clasificación jerárquica y OPTICS. Los detalles se pueden consultar en el Algoritmo 3.

Algoritmo 3: Algoritmo propuesto

1. Extracción y almacenamiento de los datos;
2. Recopilación los datos;
3. Vectorización;
4. Filtrado;
5. Agrupación: algoritmos de conglomerados jerárquicos y OPTICS;

IV. Experimentos y Resultados

IV-A. Conjuntos de Datos

Para llevar a cabo los experimentos se han utilizado los dos conjuntos de datos más recientes en la literatura, concretamente: VISION dataset [23] y ACID dataset [4]. En concreto se han utilizado muestras de ambos conjuntos de datos. La información concreta de cada muestra se puede ver en las Tablas IV y V.

IV-B. Condiciones Experimentales

Como se comentó anteriormente para el procedimiento de obtención de conglomerados, es necesario definir una medida. Sin embargo, en un procedimiento de análisis de conglomerados, es posible incluso trabajar con semimétricas. En particular, las medidas utilizadas en nuestro trabajo han sido: Euclídea, correlación, Rogers-Tanimoto y métrica de Sokal-Sneath.

Para la realización de los experimentos se han tenido en cuenta ciertas consideraciones. En primer lugar, la etiqueta Field no siempre es válida para identificar el origen, ya que tiene valores específicos que dependen del propio vídeo, en el caso de las etiquetas relacionadas con la fecha de creación,

Tabla IV
COMPOSICIÓN DE LA MUESTRA DE ACID

Marca	Modelo	Dispositivo	# Vídeos
Apple	iPhone 8 plus	M00	223
Asus	Zenfoe 4 Laser	M01	239
Canon	VIXIA HF R800	M06	25
Google	Pixel 1	M10	25
	Pixel 2	M11	25
Huawei	Honor 6X Pixel 2	M12	25
	Honor Mate SE 2	M13	25
Kodak	Ektra	M15	25
LG	Q6	M16	25
	X Charge	M17	25
Moto	E4	M18	25
	G5 plus	M19	25
Nikon	Coolpix S33	M20	25
	Coolpix S3700	M21	25
	Coolpix S7000	M22	25
Olympus	Stylus Tough TG-860	M24	25
Samsung	Galaxy J7 Pro	M27	25
	Galaxy S5	M29	25
	Galaxy S7	M30	25
	Galaxy Tab A	M31	25

Tabla V
COMPOSICIÓN DE LA MUESTRA DE VISION

Marca	Modelo	# Vídeos
Apple	iPad2	16
	Ipad mini	16
	iPhone 4	19
	iPhone 4S	28
Asus	Zenphone 2 Laser	19
Huawei	Ascend G6-U10	19
	Honor 5C NEM-L51	19
	P8 GRA-L09	19
	P9 EVA-L09	19
	P9 Lite VNS-L31	19
Lenovo	Lenovo P70-A	19
LG	D290	19
Microsoft	Lumia 640 LTE	10
OnePlus	A3000	19
	A3003	19
Samsung	Galaxy S III Mini GT-I8190	16
	Galaxy S III Mini GT-I8190N	22
	Galaxy S3 GT-I9300	19
	Galaxy S4 mini GT-I9195	19
	Galaxy S5 SM-G900F	19
	Galaxy Tab 3 GT-P5210	37
	Galaxy Tab A SM-T555	16
	Galaxy Trend Plus GT-S7580	16
Sony	Xperia Z1 Compact D5503	19
Wiko	Ridge 4G	11
WhatsApp	WhatsApp	644
Xiaomi	Redmi Note 3	19
Youtube	Youtube	622

la duración, etc. Se han eliminado algunos átomos: *modificationTime*, *creationTime*, *entryCount*, *sampleCount*, *freeSpace* y *duration*. En segundo lugar, como universo se han definido todas las representaciones posibles de las etiquetas de los átomos, específicamente las siguientes: PathField, PathFieldValue, PathOrderField y PathOrderFieldValue.

Para elegir la mejor representación del conjunto de datos a agrupar y la mejor métrica, las alternativas se han evaluado siguiendo el criterio del Coeficiente de Silueta. La representación y medida con el mayor coeficiente de silueta será la más probable de ser correctamente separada.

Definición 4: El Coeficiente de Silueta es una medida de la consistencia de los conglomerados. Mide tanto la cohesión como la separación de los mismos. Sean C_i, $i = 1, ..., k$ los conglomerados. Dado $i \in C_i$, sean

$$a(i) = \frac{1}{|C_i| - 1} \sum_{j \in C_i, j \neq i} d(i, j) \quad (1)$$

la distancia media entre i y todos los demás puntos de datos en el mismo conglomerado, donde $d(i, j)$ es la distancia entre i y j en el conglomerado C_i, y

$$b(i) = \min_{k \neq i} \frac{1}{|C_k|} \sum_{j \in C_k} d(i, j) \quad (2)$$

la distancia media más pequeña de i a todos los puntos de cualquier otro grupo, del cual i no es miembro. El coeficiente de silueta se define como:

$$s(i) = \begin{cases} 1 - a(i)/b(i), & \text{si } a(i) < b(i) \\ 0, & \text{si } a(i) = b(i) \\ a(i)/b(i) - 1, & \text{si } a(i) > b(i) \end{cases} \quad (3)$$

El Coeficiente de Silueta se ha utilizado ampliamente en otros trabajos de análisis forense multimedia como por ejemplo [28], [29] o [30].

En la Tabla VI se muestra el Coeficiente de Silueta máximo (para cualquier métrica) en cada uno de los conjuntos de datos utilizados en la experimentación. Además, las Tablas VII y VIII muestran el resultado de las 4 mejores métricas que han dado buenos resultados en ambos conjuntos de datos que han sido calculados usando el Coeficiente de Silueta.

Tabla VI
COEFICIENTE DE SILUETA MEDIO MÁXIMO PARA CADA MÉTRICA EN LOS CONJUNTOS DE DATOS DE LA EXPERIMENTACIÓN

Universo	Clase	Coeficiente de Silueta	
		VISION	ACID
PathField	Marca	0.042925	0.532204
PathField	Dispositivo	-0.044301	0.405100
PathField	Modelo	-0.104576	0.405133
PathFieldValue	Marca	0.490063	0.586225
PathFieldValue	Dispositivo	0.492146	0.653571
PathFieldValue	Modelo	0.465963	0.653581
PathOrderField	Marca	0.747372	0.907504
PathOrderField	Dispositivo	0.666355	0.805020
PathOrderField	Modelo	0.609535	0.805028
PathOrderFieldValue	Marca	0.585747	0.814979
PathOrderFieldValue	Dispositivo	0.538426	0.795484
PathOrderFieldValue	Modelo	0.485761	0.795490

IV-C. Evaluación del Desempeño del procedimiento de Análisis de Conglomerados

Existen numerosas medidas de comparación de los resultados de un procedimiento de análisis de conglomerados [31]. En este trabajo se hará uso del Índice Rand (RI) o medida de Rand. Este índice es un valor perteneciente al intervalo $[0, 1]$ que calcula una medida de similitud entre dos conglomerados considerando todos los pares de muestras y contando los pares que se asignan en los mismos o diferentes conglomerados en los conglomerados predichos y verdaderos.

Tabla VII
TOP MÉTRICAS MUESTRA DE VISION

Universo	Clase	Métrica	Resultado
PathField	Marca	Euclídea	0.0429
PathField	Marca	Correlación	-0.0029
PathField	Marca	Rogers-Tanimoto	0.0047
PathField	Marca	Sokal-Sneath	0.0103
PathField	Dispositivo	Euclídea	-0.0443
PathField	Dispositivo	Correlación	-0.0504
PathField	Dispositivo	Rogers-Tanimoto	-0.0503
PathField	Dispositivo	Sokal-Sneath	-0.0488
PathField	Modelo	Euclídea	-0.1045
PathField	Modelo	Correlación	-0.1123
PathField	Modelo	Rogers-Tanimoto	-0.1127
PathField	Modelo	Sokal-Sneath	-0.1112
PathFieldValue	Marca	Euclídea	0.3073
PathFieldValue	Marca	Correlación	0.4900
PathFieldValue	Marca	Rogers-Tanimoto	0.4782
PathFieldValue	Marca	Sokal-Sneath	0.4021
PathFieldValue	Dispositivo	Euclídea	0.3042
PathFieldValue	Dispositivo	correlation	0.4921
PathFieldValue	Dispositivo	Rogers-Tanimoto	0.4860
PathFieldValue	Dispositivo	Sokal-Sneath	0.4137
PathFieldValue	Modelo	Euclídea	0.2919
PathFieldValue	Modelo	Correlación	0.4659
PathFieldValue	Modelo	Rogers-Tanimoto	0.4586
PathFieldValue	Modelo	Sokal-Sneath	0.3940
PathOrderField	Marca	Euclídea	0.7473
PathOrderField	Marca	Correlación	0.7253
PathOrderField	Marca	Rogers-Tanimoto	0.7211
PathOrderField	Marca	Sokal-Sneath	0.7329
PathOrderField	Dispositivo	Euclídea	0.6663
PathOrderField	Dispositivo	Correlación	0.6267
PathOrderField	Dispositivo	Rogers-Tanimoto	0.6337
PathOrderField	Dispositivo	Sokal-Sneath	0.6480
PathOrderField	Modelo	Euclídea	0.6095
PathOrderField	Modelo	Correlación	0.5647
PathOrderField	Modelo	Rogers-Tanimoto	0.5705
PathOrderField	Modelo	Sokal-Sneath	0.5849
PathOrderFieldValue	Marca	Euclídea	0.3838
PathOrderFieldValue	Marca	Correlación	0.5857
PathOrderFieldValue	Marca	Rogers-Tanimoto	0.5711
PathOrderFieldValue	Marca	Sokal-Sneath	0.4820
PathOrderFieldValue	Dispositivo	Euclídea	0.3545
PathOrderFieldValue	Dispositivo	Correlación	0.5379
PathOrderFieldValue	Dispositivo	Rogers-Tanimoto	0.5384
PathOrderFieldValue	Dispositivo	Sokal-Sneath	0.4673
PathOrderFieldValue	Modelo	Euclídea	0.3230
PathOrderFieldValue	Modelo	Correlación	0.4857
PathOrderFieldValue	Modelo	Rogers-Tanimoto	0.4823
PathOrderFieldValue	Modelo	Sokal-Sneath	0.4239

Tabla VIII
TOP MÉTRICAS MUESTRA DE ACID

Universo	Clase	Métrica	Resultado
PathField	Marca	Euclídea	0.5322
PathField	Marca	Correlación	0.5238
PathField	Marca	Rogers-Tanimoto	0.5238
PathField	Marca	Sokal-Sneath	0.5243
PathField	Dispositivo	Euclídea	0.4046
PathField	Dispositivo	Correlación	0.4048
PathField	Dispositivo	Rogers-Tanimoto	0.4050
PathField	Dispositivo	Sokal-Sneath	0.4050
PathField	Modelo	Euclídea	0.4051
PathField	Modelo	Correlación	0.4048
PathField	Modelo	Rogers-Tanimoto	0.4050
PathField	Modelo	Sokal-Sneath	0.4050
PathFieldValue	Marca	Euclídea	0.4127
PathFieldValue	Marca	Correlación	0.5862
PathFieldValue	Marca	Rogers-Tanimoto	0.5781
PathFieldValue	Marca	Sokal-Sneath	0.5349
PathFieldValue	Dispositivo	Euclídea	0.4521
PathFieldValue	Dispositivo	Correlación	0.6535
PathFieldValue	Dispositivo	Rogers-Tanimoto	0.6448
PathFieldValue	Dispositivo	Sokal-Sneath	0.5999
PathFieldValue	Modelo	Euclídea	0.4521
PathFieldValue	Modelo	Correlación	0.6535
PathFieldValue	Modelo	Rogers-Tanimoto	0.6448
PathFieldValue	Modelo	Sokal-Sneath	0.5999
PathOrderField	Marca	Euclídea	0.8945
PathOrderField	Marca	Correlación	0.8921
PathOrderField	Marca	Rogers-Tanimoto	0.8934
PathOrderField	Marca	Sokal-Sneath	0.9075
PathOrderField	Dispositivo	Euclídea	0.8045
PathOrderField	Dispositivo	Correlación	0.8050
PathOrderField	Dispositivo	Rogers-Tanimoto	0.8049
PathOrderField	Dispositivo	Sokal-Sneath	0.8047
PathOrderField	Modelo	Euclídea	0.8045
PathOrderField	Modelo	Correlación	0.8050
PathOrderField	Modelo	Rogers-Tanimoto	0.8049
PathOrderField	Modelo	Sokal-Sneath	0.8047
PathOrderFieldValue	Marca	Euclídea	0.6659
PathOrderFieldValue	Marca	Correlación	0.8205
PathOrderFieldValue	Marca	Rogers-Tanimoto	0.8149
PathOrderFieldValue	Marca	Sokal-Sneath	0.7622
PathOrderFieldValue	Dispositivo	Euclídea	0.6196
PathOrderFieldValue	Dispositivo	Correlación	0.7954
PathOrderFieldValue	Dispositivo	Rogers-Tanimoto	0.7896
PathOrderFieldValue	Dispositivo	Sokal-Sneath	0.7374
PathOrderFieldValue	Modelo	Euclídea	0.6196
PathOrderFieldValue	Modelo	Correlación	0.7954
PathOrderFieldValue	Modelo	Rogers-Tanimoto	0.7896
PathOrderFieldValue	Modelo	Sokal-Sneath	0.7374

IV-D. Resultados del Algoritmo de Conglomerado Jerárquico

Todas las ejecuciones se han completado con las diferentes configuraciones, mostradas en las Tablas VII y VIII, para cada uno de los conjuntos de datos utilizados en este trabajo. La Tabla IX muestra el resumen de las condiciones experimentales que se han utilizado en el algoritmo de conglomerados jerárquico.

Tabla IX
CONFIGURACIÓN DEL CONGLOMERADO JERÁRQUICO.

Parámetros	Valores
Universo	PathField
Métrica	Euclídea
Umbral	1.132
Criterio	Encadenamiento
Encadenamiento	Simple

IV-D1. Resultados obtenidos para la muestra de VISION: Los resultados del experimento se muestran en la Figura 1 donde se puede ver que el número de conglomerados que han sido identificados, son 17 de los 13 modelos disponibles en el conjunto de datos de VISION. Los vídeos de las plataformas YouTube y WhatsApp se han identificado casi por completo en un conglomerado cada uno. La marca Apple identifica varios conglomerados con un comportamiento diferente al de las otras marcas. Otros dispositivos de diferentes marcas,

por el contrario, no muestran ninguna diferencia con esta representación, como el Asus Zenfone, que produce vídeos exactamente como los de Honor 5c y P8 de Huawei. Siempre es imposible distinguir un vídeo de Asus de uno de Huawei con esta representación de los datos. Por último, la marca OnePlus también se distingue, como las marcas Sony o Wiko. Las Figuras 1 y 2 muestran el resultado de la agrupación de la muestra según marca y modelo respectivamente.

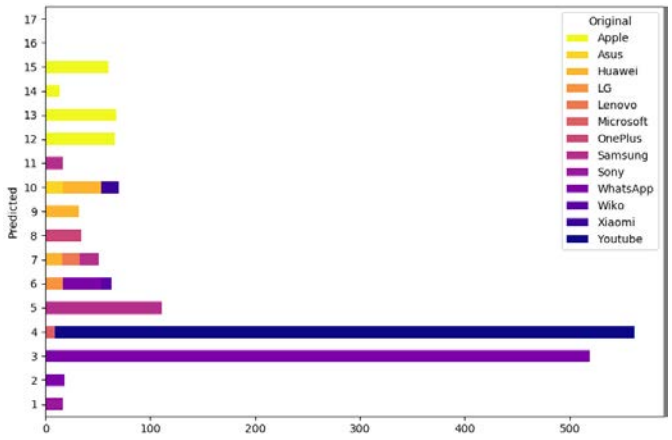

Figura 1. Conglomerados Jerárquico por Marca para la Muestra de VISION.

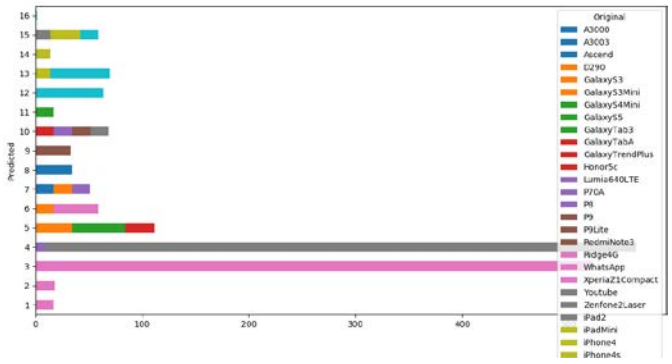

Figura 2. Conglomerados Jerárquico por Modelo para la Muestra de VISION.

IV-D2. Resultados Obtenidos para la Muestra de ACID: Como se puede ver en la Figura 3 hay 11 conglomerados de las 11 marcas que pertenecen al conjunto de datos. En este caso, a diferencia de la muestra de VISION, la marca Apple está correctamente clasificada en un solo grupo. Las marcas LG y Moto no pueden distinguirse con esta representación. En cuanto a las cámaras digitales, se observa que las marcas Canon y Olympus están correctamente clasificadas, pero no ocurre lo mismo con la marca Kodak, que no puede distinguirse de la marca Samsung.

El resumen de estos resultados por marca para las muestras de VISION y ACID se muestran en la Tabla X.

La Figura 4 muestra gráficamente los conglomerados que se han generado según Modelo para la muestra de ACID y en la Tabla XI muestra los resultados numéricos asociados.

IV-E. Resultados del Algoritmo OPTICS

Todas las configuraciones mostradas en las Tablas VII y VIII se han utilizado para ejecutar OPTICS. En las ejecuciones se han fijado *minPoints*$= 5$ y $\epsilon = 0{,}01$ que son

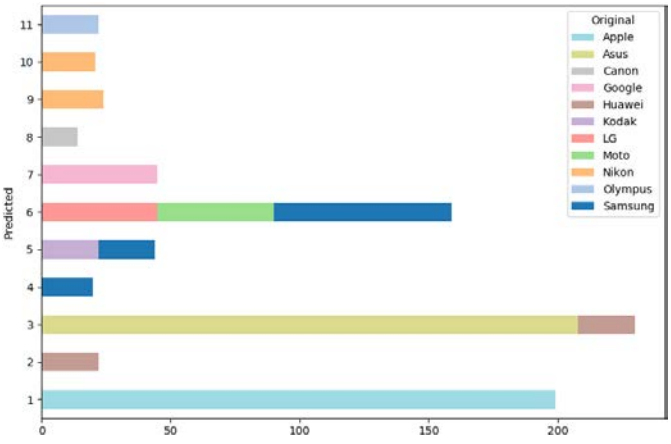

Figura 3. Resultado del Procedimiento de Conglomerados Jerárquico por Marca para la Muestra de ACID.

Tabla X
CONGLOMERADO JERÁRQUICO AGRUPADOS POR MARCA.

Parámetro	VISION	ACID
#Marcas	13	11
#Conglomerados	17	11
RI	0.8839517587	0.8128426754
Homogeneidad	0.9195359995	0.8324380092
Completitud	0.7970734665	0.8923993328

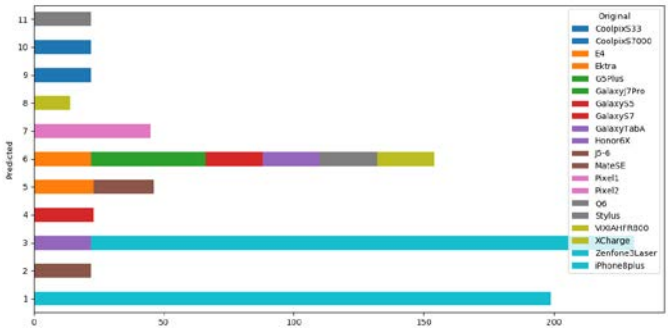

Figura 4. Resultado del Procedimiento de Conglomerados Jerárquico por Modelo para la Muestra de ACID

Tabla XI
PROCEDIMIENTO DE CONGLOMERADOS JERÁRQUICO POR MODELO PARA LA MUESTRA DE ACID

#Modelos	20
#Conglomerados	11
RI	0.8233019504
Homogeneidad	0.778805082778
Completitud	1.0

los valores que han proporcionado mejores resultados en las experimentaciones. La Tabla XII muestra el resumen de las condiciones experimentales del algoritmo OPTICS.

IV-E1. Resultados para la Muestra de VISION: Como se puede ver en la Figura 5 el algoritmo ha generado 23 conglomerados de las 13 marcas que pertenecen al conjunto de datos de VISION.

Al igual que en el caso del procedimiento de conglomerados jerárquicos, la marca Apple necesita varios conglomerados para identificarse, aunque un aspecto positivo a señalar es que en esos conglomerados no hay mezcla de

Tabla XII
CONFIGURACIÓN DEL EXPERIMENTO CON OPTICS QUE PROPORCIONA MEJORES RESULTADOS.

Parámetros	Valores
Universe	PathOrderField
Metric	Roger-Stanimoto
Epsilon	0.01
MinPoints	5

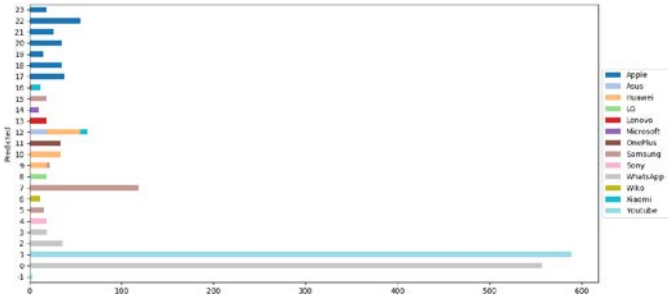

Figura 5. Resultado del Algoritmo OPTICS por Marca para la Muestra de VISION.

otra marca en esos conglomerados. Los vídeos de YouTube o WhatsApp se clasifican principalmente en un conglomerados por modelo. Por lo tanto, el algoritmo es capaz de agrupar vídeos nativos de dispositivos móviles y también vídeos que se han descargado de plataformas en línea como YouTube o WhatsApp. Los resultados de la experimentación por Modelo se pueden ver gráficamente en la Figura 6.

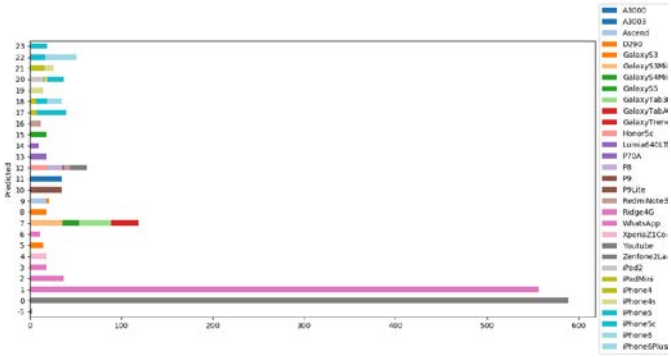

Figura 6. Resultado del Algoritmo OPTICS por Modelo para la Muestra de VISION.

IV-E2. Resultados para la Muestra de ACID: En la Figura 7 se puede ver que el algoritmo ha originado 16 conglomerados de las 11 marcas disponibles en el conjunto de datos ACID. La clasificación es correcta tanto en los vídeos originados por dispositivos móviles como en los generados por cámaras digitales. Este algoritmo tiene mejores resultados que el algoritmo jerárquico.

Los resultados detallados por Modelo se muestran en la Figura 8

El resultado de la ejecución del Algoritmo OPTICS por marcas para las muestras de VISION y ACID se pueden ver en detalle en la Tabla XIII.

Los resultados del Algoritmo OPTICS por modelos se pueden ver en la Tabla XIV.

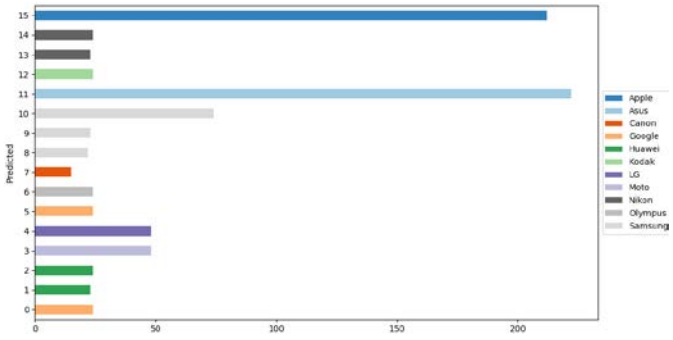

Figura 7. Resultado del Algoritmo OPTICS por Marca para la Muestra de ACID.

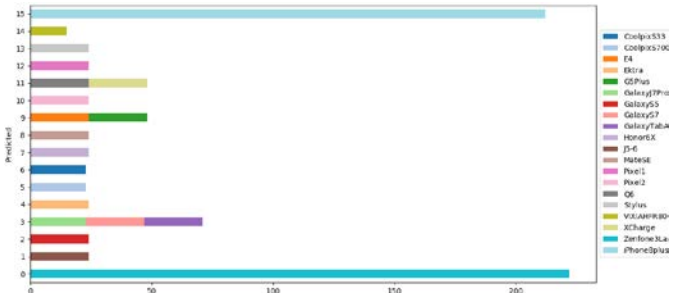

Figura 8. Resultado del Algoritmo OPTICS por Modelo para la Muestra de ACID.

Tabla XIII
ALGORITMO OPTICS POR MARCA.

Parámetro	VISION	ACID
#Marcas	13	11
#Conglomerados	25	16
RI	0.8930095982	0.8839517587
Homogeneidad	0.9759981843	0.9195359995
Completitud	0.7737539175	0.7970734665

Tabla XIV
ALGORITMO OPTICS POR MODELO.

Parámetro	VISION	ACID
#Modelos	13	16
#Conglomerados	25	20
RI	0.9210856392	0.9687571912
Homogeneidad	0.8758900022	0.9313286618
Completitud	0.8950395170	1.0

V. CONCLUSIONES

En este trabajo se ha mostrado cómo la información de los archivos de vídeo se puede explotar para agrupar vídeos por fuente de datos, sin formación previa de un clasificador. En la literatura actualmente disponible hay una gran escasez en la investigación de la fuente de adquisición de vídeo que utiliza la estructura del contenedor de vídeo para obtener las características.

Un punto esencial de la metodología propuesta ha sido la correcta adquisición de los datos para su posterior procesado y tratamiento. Con una buena adquisición preliminar, el tratamiento posterior a través del uso de algoritmos de clasificación se ha visto eficaz a la hora de determinar mediante el uso de técnicas de Minería de Datos la agrupación final de lo mismos.

La metodología propuesta ha sido validada a través de dos conjuntos de datos a los que ha sido aplicada con la misma selección de parámetros a fin de obtener resultados comparables. Los conjuntos de datos utilizados se han obtenido mediante el muestreo sobre las dos bases de datos más actuales de la literatura. Las bases de datos contienen vídeos de diversas tecnologías: vídeos nativos de dispositivos móviles, vídeos nativos de cámaras digitales y vídeos que se han descargado de las plataformas WhatsApp y YouTube de tal manera que se ha intentado obtener muestras suficientemente significativas para poder llevar a acabo el estudio.

La metodología propuesta es lo suficientemente general como para poder aplicarla y adaptarla a otro tipos de datos (modificación de la obtención preliminar de los mismos, modificación de la representación primitiva mediante la combinación de otros campos, etc.) , así como aplicar otras técnicas de clasificación presentes en Análisis Multivariante (técnicas de clasificación no jerárquica, utilización de métodos basados en modelos estadísticos [25], entre otros). Como se ha visto en la los resultados numéricos obtenidos de las muestras, los algoritmos de agrupación propuestos han proporcionado buenos resultados desde el punto de vista de la clasificación, obteniéndose precisiones superiores al 89 %.

Agradecimientos

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 700326. This paper has also received funding from THEIA (Techniques for Integrity and authentication of multimedia files of mobile devices) UCM project (FEI-EU-19-04).

Referencias

[1] D. R. Hayes, *A Practical Guide to Computer Forensics Investigations*. Pearson Education, August 2015.

[2] M. C. Stamm, M. Wu, and K. Liu, "Information Forensics: An Overview of the First Decade," *IEEE Access*, vol. 1, pp. 167–200, March 2013.

[3] 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*, Bucharest, Romania, August 2012, pp. 1229–1233.

[4] B. Hosler, O. Mayer, B. Bayar, X. Zhao, C. Chen, J. A. Shackleford, and M. C. Stamm, "A Video Camera Model Identification System Using Deep Learning and Fusion," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Brighton, UK, May 2019, pp. 8271–8275.

[5] B. Sevinc, H. Sencar, and N. Memon, "Classification of Digital Camera-Models Based on Demosaicing Artifacts," *Digital Investigation*, vol. 5, no. 1-2, pp. 49–59, September 2008.

[6] H. Cao and A. C. Kot, "Accurate Detection of Demosaicing Regularity for Digital Image Forensics," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 4, pp. 899–910, December 2009.

[7] X. Zhao and M. C. Stamm, "Computationally Efficient Demosaicing Filter Estimation for Forensic Camera Model Identification," in *Proceedings of the IEEE International Conference on Image Processing*, Phoenix, Arizona, USA, September 2016, pp. 151–155.

[8] M. A. Qureshi, C.-H. Deriche, M. Choi, H.-Y. Lee, and H.-K. Lee, "Estimation of Color Modification in Digital Images by CFA Pattern Change," *Forensic Science International*, vol. 226, no. 1-3, pp. 94–105, March 2013.

[9] F. M.Peng and D.-L. Zhou, "Discriminating Natural Images and Computer Generated Graphics Based on the Impact of CFA Interpolation on the Correlation of PRNU," *Digital Investigation*, vol. 11, no. 2, pp. 111–119, June 2014.

[10] P. Mullan, C. Riess, and F. Freiling, "Forensic Source Identification Using JPEG Image Headers: The Case of Smartphones," *Digital Investigation*, vol. 28, no. Supplement, pp. S68–S76, April 2019.

[11] J. Lukáš, J. Fridrich, and M. Goljan, "Determining Digital Image Origin Using Sensor Imperfections," in *Proceedings of the Image and Video Communications and Processing 2005*, San Jose, California, USA, March 2005, pp. 249–260.

[12] ——, "Detecting Digital Image Forgeries Using Sensor Pattern Noise," in *Proceedings of the SPIE-The International Society for Optical Engineering*, San Jose, California, USA, February 2006, pp. 362–372.

[13] M. Chen, J. Fridrich, and M. Goljan, "Digital Imaging Sensor Identification (Further Study)," in *Proceedings of the SPIE Electronic Imaging, Security, Steganography, Watermarking of Multimedia Contents IX*, February 2007, pp. 1–14.

[14] T. Filler, J. Fridrich, and M. Goljan, "Using Sensor Pattern Noise for Camera Model Identification," in *Proceedings of the 15th IEEE International Conference on Image Processing*, October 2008, pp. 1296–1299.

[15] Y. Hu, B. Yu, and C. Jian, "Source Camera Identification Using Large Components of Sensor Pattern Noise," in *Proceedings of the Second International Conference on Computer Science and its Applications*, Jeju Island, Korea, January 2010, pp. 1–5.

[16] C.-T. 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.

[17] L. J. García Villalba, A. L. Sandoval Orozco, R. Ramos López, and J. Hernández Castro, "Identification of Smartphone Brand and Model via Forensic Video Analysis," *Expert Systems with Applications*, vol. 55, pp. 59–69, August 2016.

[18] "ISO/IEC 14496-12:2015: Coding of Audio-Visual Objects-Part 12: ISO base media file format," https://www.iso.org/standard/68960.html, 2017.

[19] T. Gloe, A. Fischer, and M. Kirchner, "Forensic Analysis of Video File Formats," *Digital Investigation*, vol. 11, no. Supplement 1, pp. 68–76, May 2014.

[20] M. Iuliani, D. Shullani, M. Fontani, S. Meucci, and A. Piva, "A Video Forensic Framework for the Unsupervised Analysis of MP4-Like File Container," *Institute of Electrical and Electronics Engineers Transactions on Information Forensics and Security*, vol. 14, no. 3, pp. 635–645, March 2019.

[21] J. Song, K. Lee, Y. Lee, Wan, and H. Lee, "Integrity Verification of the Ordered data Structures in Manipulated Video Content," *Digital Investigation*, vol. 18, pp. 1–7, September 2016.

[22] "QuickTime File Format Specification: Overview," https://developer.apple.com/library/archive/documentation/QuickTime/QTFF/QTFFChap1/qtff1.html, September 2016.

[23] D. Shullani, M. Fontani, M. Iuliani, O. Alshaya, and A. Piva, "VISION: a Video and Image dataset for Source Identification," *EURASIP Journal on Information Security*, vol. 1, no. 15, pp. 1–16, October 2017.

[24] M. Esteban Cobo, "Herramienta para la Extracción Automática de Metadatos en Vídeos de Dispositivos Móviles," Facultad de Informática, Universidad Complutense de Madrid, Spain, Trabajo de Fin de Grado, August 2016.

[25] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed. Pearson Education Inc.(US), 2007.

[26] M. Ester, H. Kriegel, J. Sander, and X. Xiaowei, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, Portland, Oregon, September 1996, pp. 226–231.

[27] M. Ankerst, M. Breunig, H. P. Kriegel, and J. Sander, "OPTICS: Ordering Points To Identify the Clustering Structure," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, New York, USA, June 1999, pp. 49–60.

[28] S. Khan and T. Bianchi, "Fast Image Clustering Based on Camera Fingerprint Ordering," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, Shanghai, China, July 2019, pp. 766–771.

[29] L. J. García Villalba, A. L. Sandoval Orozco, and J. Rosales Corripio, "Smartphone Image Clustering," *Expert Systems with Applications*, vol. 42, no. 4, pp. 1927–1940, March 2015.

[30] F. Marra, G. Poggi, C. Sansone, and L. Verdoliva, "Blind PRNU-Based Image Clustering for Source Identification," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 9, pp. 2197–2211, September 2017.

[31] N. X. Vinh, J. Epps, and J. Bailey, "Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance," *Journal of Machine Learning Research*, vol. 11, pp. 2837–2854, October 2010.

Raquel Ramos López received his Computer Science Engineering degree and a M.S. degree in Computer Science, both from the Universidad Complutense of Madrid. She is currently a Ph.D. student of Computer Engineering in the GASS Group (Group of Analysis, Security and Systems) in Universidad Complutense de Madrid. Her main research interests are information security and its applications.

Elena Almaraz Luengo received a Mathematics degree from the University Complutense of Madrid in 2005, a Statistical Sciences and Techniques degree from the University Complutense of Madrid in 2007 and a Bussines and Administration degree from the National Distance Education University in 2015. She is Doctor in Mathematics from the University Complutense of Madrid since 2007 and hold a Master's Degree in Advanced Mathematics with specialization in Statistics and Operations Research, from the National Distance Education University in 2010. She is currently an Assistant Professor in the Department of Statistic and Operational Research in the Faculty of Mathematics Sciences of the University Complutense of Madrid. Her main interest are statistic techniques, probability, information security and applications.

Ana Lucila Sandoval Orozco received a Computer Science Engineering degree from the Universidad Autónoma del Caribe (Colombia) in 2001. She holds a Specialization Course in Computer Networks (2006) from the Universidad del Norte (Colombia), and holds a M.Sc. in Research in Computer Science (2009) and a Ph.D. in Computer Science (2014), both from the Universidad Complutense de Madrid (Spain). She is currently a postdoctoral researcher at Universidad Complutense de Madrid (Spain). Her main research interests are coding theory, information security and its applications.

Luis Javier García Villalba received a Telecommunication Engineering degree from the Universidad de Málaga (Spain) in 1993 and holds a Ph.D. in Computer Science (1999) from the Universidad Politécnica de Madrid (Spain). Visiting Scholar at COSIC (Computer Security and Industrial Cryptography, Department of Electrical Engineering, Faculty of Engineering, Katholieke Universiteit Leuven, Belgium) in 2000 and Visiting Scientist at IBM Research Division (IBM Almaden Research Center, San Jose, CA, USA) in 2001 and 2002, he is currently Associate Professor of the Department of Software Engineering and Artificial Intelligence at the Universidad Complutense de Madrid (UCM) and Head of Complutense Research Group GASS (Group of Analysis, Security and Systems) which is located in the Faculty of Computer Science and Engineering at the UCM Campus. His professional experience includes the management of both national and international research projects and both public (Spanish Ministry of R&D, Spanish Ministry of Defence, Horizon 2020 - European Commission, ...) and private financing (Hitachi, IBM, Nokia, Safelayer Secure Communications, TB Solutions Security, ...). Author or co-author of numerous international publications is editor or guest editor of numerous journals such as Entropy MPDI, Future Generation Computer Systems, Future Internet MDPI, IEEE Latin America Transactions, IET Communications, IET Networks, IET Wireless Sensor Systems, International Journal of Ad Hoc and Ubiquitous Computing, International Journal of Multimedia and Ubiquitous Engineering (IJMUE), Journal of Supercomputing, Sensors MDPI, etc.