Shullani *et al. EURASIP Journal on Information Security* (2017) 2017:15
DOI 10.1186/s13635-017-0067-2

EURASIP Journal on Information Security

RESEARCH Open Access

VISION: a video and image dataset for source identification

Dasara Shullani[1], Marco Fontani[1,2], Massimo Iuliani[1,2], Omar Al Shaya[1,3] and Alessandro Piva[1,2*]

Abstract

Forensic research community keeps proposing new techniques to analyze digital images and videos. However, the performance of proposed tools are usually tested on data that are far from reality in terms of resolution, source device, and processing history. Remarkably, in the latest years, portable devices became the preferred means to capture images and videos, and contents are commonly shared through social media platforms (SMPs, for example, Facebook, YouTube, etc.). These facts pose new challenges to the forensic community: for example, most modern cameras feature digital stabilization, that is proved to severely hinder the performance of video source identification technologies; moreover, the strong re-compression enforced by SMPs during upload threatens the reliability of multimedia forensic tools. On the other hand, portable devices capture both images and videos with the same sensor, opening new forensic opportunities. The goal of this paper is to propose the VISION dataset as a contribution to the development of multimedia forensics. The VISION dataset is currently composed by 34,427 images and 1914 videos, both in the native format and in their social version (Facebook, YouTube, and WhatsApp are considered), from 35 portable devices of 11 major brands. VISION can be exploited as benchmark for the exhaustive evaluation of several image and video forensic tools.

Keywords: Dataset multimedia forensics, Image forensics, Video forensics, Source identification

1 Introduction

In the last decades, visual data gained a key role in providing information. Images and videos are used to convey persuasive messages to be used under several different environments, from propaganda to child pornography. The wild world of web also allows users to easily share visual contents through social media platforms. Statistics [1] show that a relevant portion of the world's population owns a digital camera and can capture pictures. Furthermore, one third of the people can go online and upload their pictures on websites and social networks. Given their digital nature, these data also convey several information related to their life cycle (e.g., source device, processing they have been subjected to). Such information may become relevant when visual data are involved in a crime. In this scenario, multimedia forensics (MF) has been proposed as a solution for investigating images and videos to determine information about their life cycle [2]. During the years, the research community developed several tools to analyze a digital image, focusing on issues related to the identification of the source device and the assessment of content authenticity [3].

Generally, the effectiveness of a forensic technique should be verified on image and video datasets that are freely available and shared among the community. Unfortunately, these datasets, especially for the case of videos, are outdated and non-representative of real case scenarios. Indeed, most multimedia contents are currently acquired by portable devices that keep updating year by year. These devices are also capable to acquire both videos and images exploiting the same sensor, thus opening new investigation opportunities in linking different kind of contents [4]. This motivates the need for a new dataset containing a heterogeneous and sufficiently large set of visual data—both images and videos—as benchmark to test and compare forensic tools.

In this paper, we present a new dataset of native images and videos captured with 35 modern smartphones/tablets

*Correspondence: alessandro.piva@unifi.it
[1]Department of Information Engineering, University of Florence, Via di S. Marta, 3, 50139 Florence, Italy
[2]FORLAB, Multimedia Forensics laboratory, PIN Scrl, Piazza G. Ciardi, 25, 59100 Prato, Italy
Full list of author information is available at the end of the article

belonging to 11 different brands: Apple, Asus, Huawei, Lenovo, LG electronics, Microsoft, OnePlus, Samsung, Sony, Wiko, and Xiaomi.

Overall, we collected 11,732 native images; 7565 of them were shared through Facebook, in both high and low quality, and through WhatsApp, resulting in a total of 34,427 images. Furthermore, we acquired 648 native videos, 622 of which were shared through YouTube at the maximum available resolution, and 644 through WhatsApp, resulting in a total of 1914 videos[1].

To exemplify the usefulness of the VISION dataset, we test the performance of a well-known forensic tool, i.e., the detection of the sensor pattern noise (SPN) left by the acquisition device [5] for the source identification of native/social media contents; moreover, we describe some new opportunities deriving by the availability of images and videos captured with the same sensor to find a solution to current limits present in the literature. In particular, the proposed dataset contains several devices featuring in-camera digital stabilization, that is known to threaten source identification based on sensor pattern noise. Indeed, in most papers related to SPN [6–9], digitally stabilized videos are ignored, either by turning the stabilization off or considering non-stabilized devices only. This is unrealistic, considering that most common modern devices (e.g., Apple iPhones) are equipped with an in-camera digital stabilization system that cannot be turned off without resorting to third party applications.

The remaining part of the paper is organized as follows: in Section 2, we review the currently available datasets for image and video forensics and their current limitations; in Section 3, a complete description of the VISION dataset is provided for both native and social media contents; in Section 4, the dataset is exploited to evaluate some well-known forensic applications and to taste new research opportunities. Section 5 draws concluding remarks.

Eventually, we include in the Appendix additional information that can be useful to perform a deeper analysis on the available visual contents.

2 Motivation

In the field of digital image and video forensics, only few datasets have been made available to the research community, especially for the source device identification problem. This fact is a strong limitation for research advancement in our area.

One of the first datasets adopted in the multimedia forensic community is the UCID database [10], originally designed for the evaluation of image retrieval techniques. Such dataset includes 1338 uncompressed images stored in the TIFF format, but their size is very small, either 512×384 or 384×512 pixels.

The first sufficiently large and publicly available image database specifically designed for forensic applications is the Dresden Image Database [11, 12]. This dataset includes images of various indoor and outdoor scenes acquired from 73 devices, selected from 25 camera models spanning most important manufacturers and quality ranges. All cameras were configured to the highest available JPEG quality setting and maximum available resolution, and, when supported by the device, also lossless compressed images were stored. The image resolution ranges from 3072×2304 to 4352×3264 pixels, for a total of 16,961 JPEG images, 1491 RAW (unprocessed) images, 1491 RAW images processed in Lightroom 2.5, and 1491 RAW images processed in DCRaw 9.3. Since 2010, this dataset has been used by most of the works dealing with benchmarking of source identification methods.

More recently, RAISE (RAw ImageS datasEt) was presented [13]: it is a collection of 8156 raw images including a wide variety of both semantic contents and technical parameters. Three different devices (a Nikon D40, a Nikon D90, and a Nikon D7000) are employed, and the images are taken at very high resolution (3008×2000, 4288×2848, and 4928×3264 pixels) and saved in an uncompressed format (Compress Raw 12-bit and Lossless Compress Raw 14-bit) as natively provided by the employed cameras. Each image is also assigned one of seven possible categories, namely, "outdoor," "indoor," "landscape," "nature," "people," "objects," and "buildings." In the framework of the European project REWIND, a set of 200 uncompressed images acquired with a Nikon D60 camera were also made available [14] (among other sets for splicing detection, copy-move forgeries and recapture videos, all including a few number of samples). There are also other datasets, not cited here, that have been designed more for image tampering detection than for source identification, and thus no or little information is provided about the device generating the images.

As to digital videos, in the literature, there are very few datasets designed to be used in forensic scenarios; one of them is the SULFA [15], created by the University of Surrey. It collects 150 videos, each 10-s long, at 30 fps with a resolution of 320×240 pixels. The native videos are given compressed in H.264/AVC and MJPEG, for each camera: a Canon SX220, a Nikon S3000, and a Fujifilm S2800HD. Authors designed the dataset to be used for cloning detection, performed by means of Adobe Photoshop CS3 and Adobe After Effect CS5 [15]. The SULFA dataset was also extended by the REWIND dataset [16]; anyway, these datasets are less interesting for video source identification, since they contain few digital cameras only and no smartphone, while we know smartphones are the most representative kind of device today, especially for applications on social media platforms. Recently, the video tampering dataset (VTD) was provided by Al-Sanjary et al. [17]. The VTD, focused on video tampering detection on videos collected from the YouTube

platform, is composed by 33 downloaded videos, 16-s long, at 30 fps with a HD resolution. The original dataset is subdivided into four subsets: one containing unaltered videos; one with videos created by splicing; one with videos manipulated by copy-move; and one with videos tampered by swapping frames. Although they use a social media platform to acquire videos and provide interesting tampering techniques, there are not useful information related to the camera or device used.

The previous review shows that all currently available datasets consider mainly images, and the ones containing videos are not significant for video source identification; moreover, it is not possible to investigate relationships between images and videos acquired with the same sensor: this fact is a strong limitation, since 85% of shared media are captured using smartphones, which use the same sensor to capture both images and videos. Finally, another limit in the state-of-the-art is represented by the lack of a collection of controlled content coming from social media platforms, like Facebook, YouTube, and WhatsApp; indeed, recent multimedia forensic applications would take advantage in having a large dataset containing such kind of contents: for instance, in [6], the authors address the performance of identifying the source of YouTube videos but limiting to a scenario with videos belonging to 8 webcams of the same model (Logitech Quickcam STX). Similarly, Bertini et al. [18] propose to extract the Sensor Pattern Noise from images to identify fake social media accounts, but the technique was tested on 5 mobile devices only, with 200 images each.

3 The VISION Dataset

Images and videos have been acquired from each mobile device by following a specific procedure. First of all, the captured contents refer to the best-quality camera available in the device; in general, the one positioned on the upper rear of the device. Moreover, the devices were configured, when possible, with the highest quality and resolution available (usually the default one for Apple devices but not necessarily for Android ones).

VISION is mainly thought for video and image source identification applications; as a consequence, we organized the data collected from each device into two folders, (see Fig. 1 for an example), namely:

- *Images*: containing native and social exchanged images. We captured images, mainly in landscape mode, representing flat surfaces (e.g., skies or walls), here defined as *Flat*, and generic images, here defined as *Nat*, for which there are no limitations on orientation or scenario, as it can be seen in Fig. 2. In addition, the *Nat* images were exchanged via the Facebook and WhatsApp social media platforms.

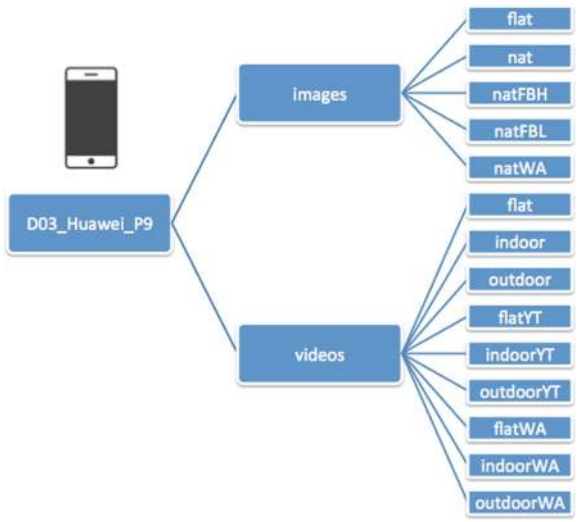

Fig. 1 VISION folder organization

- *Videos*: containing native and social exchanged videos, acquired mainly in landscape mode. The collected videos represent flat, indoor, and outdoor scenarios. The *flat scenario* includes videos belonging to flat surfaces such as walls and skies. The *indoor scenario* comprises videos representing offices or stores, and the *outdoor scenario* contains videos of open areas such as gardens. For each scenario, we used three different acquisition modes: *still mode*, where the user stands still while capturing the video; *move mode*, where the user walks while capturing the video; *panrot mode*, where the user performs a recoding combining a pan and a rotation. Furthermore, the videos belonging to each scenario were exchanged via YouTube and WhatsApp social media platforms.

The structure depicted in Fig. 1 is maintained also in the naming convention. The contents collected from each device are stored in its root folder named `ID_Brand_Model` as in `D01_Samsung_GalaxyS3 Mini`. Then, we distinguish between *images* and *videos*, within each of them, we have the native content folders and the social ones. A native flat image is called by convention as `ID_I_flat_XXXX.jpg` as in `D01_I_flat_0001.jpg`, where `ID` is the device identifier, `I` identifies it as an image content, `flat` identifies the subfolder and the type of image, while `XXXX.jpg` is an incremental number. Similarly, the video content naming is `ID_V_scenario_mode_XXXX.mp4` as in `D01_V_flat_panrot_0001.mp4`, where `V` identifies the video content, `scenario` and `mode` refer respectively to the area and the modality of the acquisition procedure. The so described naming

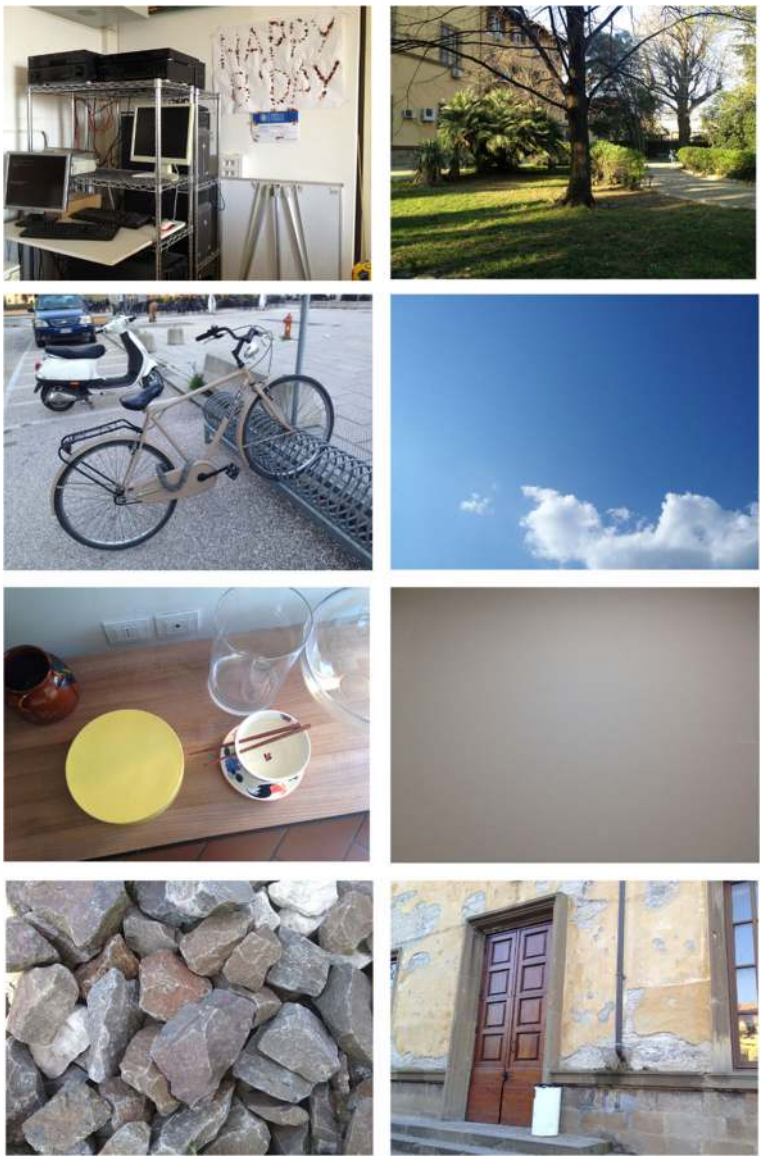

Fig. 2 Some examples of the images included in the proposed dataset

convention is also applied to the social folders represented in Fig. 1: an image uploaded to Facebook in low quality will be named `D01_I_natFBL_0001.jpg`, an image uploaded to Facebook in high quality will be named `D01_I_natFBH_0001.jpg`, while a video exchanged through WhatsApp will be named `D01_V_flatWA_panrot_0001.mp4`.

3.1 Main features

VISION is composed by 35 mobile devices from low-, middle-, and high-price range. There are 13 Apple devices, including iPhones and iPads. There are 8 Samsung devices including Galaxy phones and tablets. There are 5 Huawei and 2 OnePlus phones. Furthermore, we gathered one device for the following brands: Asus, Lenovo, LG electronics, Microsoft, Sony, Wiko, and Xiaomi. We collected a few devices of the same brand and model namely: two iPhone 4S, two iPhone 5, three iPhone 5c, two iPhone 6, and two GalaxyS3Mini. The employed devices had installed the following operating systems: iOS from 7.x to 10.x, Android from 6.x Marshmallow to 7.x Nougat, and the Windows Phone OS 8.1 Update 2.

In Table 1, we summarize the main features of the complete dataset. For each device, we report the *Brand*, *Model*, a unique identifier *ID*, and the number of collected videos and images with their corresponding resolutions.

In Table 1, we also clarify whether videos were captured using in-camera digital stabilization: the reader can see

Table 1 Devices main features

Brand	Model	ID	DStab	HDR	VR	#Videos	IR	#Images	#Flat	#Nat
Apple	iPad 2	D13	Off	F	1280 × 720	16	960 × 720	330	159	171
Apple	iPad mini	D20	On	F	1920 × 1080	16	2592 × 1936	278	119	159
Apple	iPhone 4	D09	Off	T	1280 × 720	19	2592 × 1936	326	109	217
Apple	iPhone 4S	D02	On	T	1920 × 1080	13	3264 × 2448	307	103	204
Apple	iPhone 4S	D10	On	T	1920 × 1080	15	3264 × 2448	311	133	178
Apple	iPhone 5	D29	On	T	1920 × 1080	19	3264 × 2448	324	100	224
Apple	iPhone 5	D34	On	T	1920 × 1080	32	3264 × 2448	310	106	204
Apple	iPhone 5c	D05	On	T	1920 × 1080	19	3264 × 2448	463	113	350
Apple	iPhone 5c	D14	On	T	1920 × 1080	19	3264 × 2448	339	130	209
Apple	iPhone 5c	D18	On	T	1920 × 1080	13	3264 × 2448	305	101	204
Apple	iPhone 6	D06	On	T	1920 × 1080	17	3264 × 2448	281	149	132
Apple	iPhone 6	D15	On	T	1920 × 1080	18	3264 × 2448	337	110	227
Apple	iPhone 6 Plus	D19	On	T	1920 × 1080	19	3264 × 2448	428	169	259
Asus	Zenfone 2 Laser	D23*	On	F	640 × 480	19	3264 × 1836	327	117	210
Huawei	Ascend G6-U10	D33	Off	T	1280 × 720	19	2448 × 3264	239	84	155
Huawei	Honor 5C NEM-L51	D30	Off	T	1920 × 1080	19	4160 × 3120	351	80	271
Huawei	P8 GRA-L09	D28	Off	T	1920 × 1080	19	4160 × 2336	392	126	266
Huawei	P9 EVA-L09	D03	Off	F	1920 × 1080	19	3968 × 2976	355	118	237
Huawei	P9 Lite VNS-L31	D16	Off	T	1920 × 1080	19	4160 × 3120	350	115	235
Lenovo	Lenovo P70-A	D07	Off	F	1280 × 720	19	4784 × 2704	375	158	217
LG electronics	D290	D04	On	F	800 × 480	19	3264 × 2448	368	141	227
Microsoft	Lumia 640 LTE	D17	Off	T	1920 × 1080	10	3264 × 1840	285	97	188
OnePlus	A3000	D25	On	T	1920 × 1080	19	4640 × 3480	463	176	287
OnePlus	A3003	D32	On	T	1920 × 1080	19	4640 × 3480	386	150	236
Samsung	Galaxy S III Mini GT-I8190	D26	Off	F	1280 × 720	16	2560 × 1920	210	60	150
Samsung	Galaxy S III Mini GT-I8190N	D01	Off	F	1280 × 720	22	2560 × 1920	283	78	205
Samsung	Galaxy S3 GT-I9300	D11	Off	T	1920 × 1080	19	3264 × 2448	309	102	207
Samsung	Galaxy S4 Mini GT-I9195	D31	Off	T	1920 × 1080	19	3264 × 1836	328	112	216
Samsung	Galaxy S5 SM-G900F	D27	Off	T	1920 × 1080	19	5312 × 2988	354	100	254
Samsung	Galaxy Tab 3 GT-P5210	D08	Off	F	1280 × 720	37	2048 × 1536	229	61	168
Samsung	Galaxy Tab A SM-T555	D35	Off	F	1280 × 720	16	2592 × 1944	280	126	154
Samsung	Galaxy Trend Plus GT-S7580	D22	Off	F	1280 × 720	16	2560 × 1920	314	151	163
Sony	Xperia Z1 Compact D5503	D12	On	T	1920 × 1080	19	5248 × 3936	316	100	216
Wiko	Ridge 4G	D21	Off	T	1920 × 1080	11	3264 × 2448	393	140	253
Xiaomi	Redmi Note 3	D24	Off	T	1920 × 1080	19	4608 × 2592	486	174	312

DStab shows the presence or absence of digital stabilization on the acquired content, HDR indicates whether the device supports it, VR stands for video resolution and IR for image resolution

that for most Apple devices if the stabilization is present it is also enabled (the only exceptions are D9 and D13), as it is also for the Sony Xperia, D12. On the contrary, this is not true for all other devices where the in-camera digital stabilization is set off by default. In addition, Table 1 clarifies whether the device can acquire images in HDR-*High Dynamic Range* mode: T (*True*) is used if HDR is available and F (*False*) if it is not. Several additional metadata and coding statistics are collected and reported in the Appendix.

We also make available a reduced version of VISION for researchers convenience. This baseline version is composed by 16,100 images and 315 videos, both native and social, *equally distributed* among all the devices. In Section 5, we provide instructions for accessing the dataset.

3.2 Social contents

The collected contents in VISION were also exchanged through social media platforms; in particular, for images in *Nat*, we provide their corresponding uploaded version on Facebook and WhatsApp. We chose to upload only natural images since, from a forensic point of view, having flat surfaces shared through social media is rather unrealistic. In addition, we shared all videos through YouTube and WhatsApp. In the rest of this Section, we explain the procedure used for uploading and downloading media contents through each social media platform.

Facebook web platform In order to exchange images via Facebook, we created two albums in which we uploaded all images belonging to *Nat* in high and low quality respectively (FBH and FBL from now on), as allowed by the social media platform. Indeed, as deeply explained in [19], these uploading options cause a significantly different compression strategy for the image.

For what concerns the download, we performed single-image downloads and album downloads, although there is no difference between the resulting contents. Album download functionality was recently added to the Facebook website[2] options. The one click album-download button allows downloading a zip version of each album; in each zip-file, the images are renamed by an incremental number as: 1.jpg, 2.jpg, ... *n*.jpg , where *n* is the number of images in the folder.

Since the collection of VISION lasted over a year, we exchanged data both before and after this update. We took care to provide a matching naming between the original content and the social media one: we used the SSIM index [20] as a metric to determine whether the two images depict the same content. Consequently, if the native image name is `D01_I_nat_0001.jpg`, its Facebook high quality counterpart will be named `D01_I_natFBH_0001.jpg`.

YouTube web platform All video contents were uploaded to YouTube with the *Public privacy* flag and collected into a playlist. During the collection of VISION, we exploited different solutions to speed-up the downloading process but maintaining the constraints of highest resolutions and no download compression. We encountered two software solutions to accomplish this goal, namely *ClipGrab*[3] and *Youtube-dl*[4]. Both software are freely available and can be used on several operating systems such as Unix and Windows. The main difference between the two is that the *ClipGrab* GUI can download one video at a time, while the *youtube-dl* command line can download also playlists.

As an example, we provide the following *youtube-dl* command line call to download a playlist:[5].

```
youtube-dl -f 137+140/bestvideo+best
audio -o "%(title)s.%(ext)s" -yes-play
list "device_url_playlist"
```

The options after the **-f** refers to the quality of video resolution and audio settings; here, the meaning is to choose the highest video resolution and audio quality, if not available choose the second-best pair and so on from left to right. Then with option **-o** we set the output video name and extension to be the YouTube video name and the default extension, i.e. mp4. For the complete documentation we advise the reader to refer to [21].

Similarly to the image naming convention, a video recorded in an outdoor scenario with a panrot movement has the following name: `D01_V_outdoor_panrot_0001.mp4`, while its YouTube counterpart will be named `D01_V_outdoorYT_panrot_0001.mp4`.

WhatsApp mobile application: All native video contents and images belonging to *Nat* were exchanged via WhatsApp v2.17.41 using an iPhone7 A1778 with iOS v10.3.1. We decided to use the mobile-application instead of the desktop one since the latter does not compute any compression to the shared file, while the mobile one does so. We used an iPhone since it produces a media file that is less compressed than the Android one, due to WhatsApp implementation choices. In this way, we provided an equilibrate spectrum of social image contents qualities: namely high and low provided by Facebook, and medium from WhatsApp. As to the naming convention, for these files we had the same issue as in Facebook: since downloaded images are renamed, we matched images using the SSIM index.

The videos downloaded from WhatsApp follow the same name structure, (e.g., `D01_V_outdoorWA_panrot_0001.mp4`.

4 Possible applications with experimental evaluations

This dataset was created to provide a benchmark for the forensic analysis of images and videos. In this Section, we exploit all the collected contents to test the source identification technique based on the sensor pattern noise. In this scenario, the aim is to identify the source of an image or video by evaluating the correlation between the SPN fingerprint estimated from the investigated content, and the device reference fingerprint, computed from a set of images or a video taken by this device.

We tested different application scenarios:

- Image source identification (ISI), where a query image is matched with a device reference computed from a set of images taken by the device;

- Video source identification (VSI), where a query video is matched with a device reference computed from the frames of a video taken by the device.

The identification is performed according to the classical workflow [22]: a camera fingerprint $\mathbf{K}$ is estimated from N still images or video frames $\mathbf{I}^{(1)}, \ldots, \mathbf{I}^{(N)}$ captured by the source device. A denoising filter [5] is applied to each image/frame, and the noise residuals $\mathbf{W}^{(1)}, \ldots, \mathbf{W}^{(N)}$ are obtained as the difference between each frame and its denoised version. Then, the camera fingerprint estimate $\widetilde{\mathbf{K}}$ is derived by the maximum likelihood estimator [22]:

$$\widetilde{\mathbf{K}} = \frac{\sum_{i=1}^{N} \mathbf{W}^{(i)} \mathbf{I}^{(i)}}{\sum_{i=1}^{N} \left(\mathbf{I}^{(i)}\right)^2}. \quad (1)$$

The fingerprint of the query is estimated in the same way by the available image or video frames. Then, the Peak to Correlation Energy (PCE) between the reference and the query pattern is computed and compared to a threshold [23]: if the PCE is higher than the threshold, then it is decided that the query content has been acquired by the reference device.

4.1 Image source identification

In this scenario, the reference SPN for each device is estimated using 100 still flat field images. Then, we run four experiments using natural, WhatsApp, Facebook high-quality, and Facebook low-quality images as queries. In all experiments, we consider for each device 100 matching cases (images from the same device) and the same number of mismatching cases (images randomly chosen from other devices). The achieved results are reported using ROC curves that plot true positive rate against false positive rate (see Fig. 3). The overall performance are summarized in Table 2 where, for each experiment, we also reported the dataset path of the query images and the Area Under Curve. `ID_Brand_Model` stands for any of the available device e.g., `D03_Huawei_P9`.

4.2 Video source identification

Here, the source of a test video is determined based on references estimated from a flat-field video. In particular, the reference SPN for each device is estimated from the first 100 frames of a flat video. Then, three experiments are performed using natural, YouTube and WhatsApp videos as queries, respectively. The fingerprint of each tested video is estimated from the first 100 frames. We consider for each device all available matching cases (videos from the same device) and the same number of mismatching cases (videos randomly chosen from other devices). The achieved results are reported in Fig. 4, where only non-stabilized cameras are analyzed, and in Fig. 5, where all

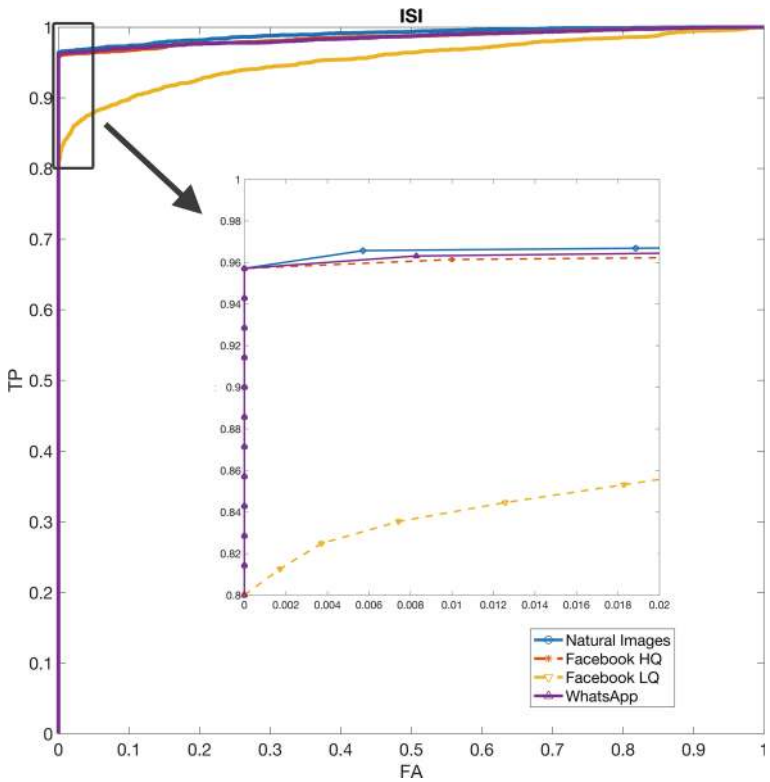

Fig. 3 (Best viewed in colors) ISI performance on Native, Facebook (HQ and LQ), and WhatsApp images using flat field references

Table 2 Performance of image source identification in growing difficulty scenarios

Experiment	Test Path	AUC
1	ID_Brand_Model/images/nat	0.9906
2	ID_Brand_Model/images/natWA	0.9860
3	ID_Brand_Model/images/natFBH	0.9859
4	ID_Brand_Model/images/natFBL	0.9544

devices in the dataset are considered. This experiment shows that performance of VSI strongly drop when digitally stabilized videos are involved. In Table 3, we briefly summarize for each test the paths in the dataset of tested videos and the Area Under Curve values obtained with and without stabilized videos.

For in-camera stabilized videos, possible solutions are still under development, as the one proposed in [24]. Anyway the solution in [24] is proved to be effective only on third party (out-camera) digital stabilization (ffmpeg), and when a non-stabilized video is available as reference. Unfortunately, most of the considered devices enforce in-camera digital stabilization, without an option to turn it off in the standard camera application.

4.3 Image vs video SPN fingerprint

In the research community, ISI and VSI applications are separately studied so that there is still no better way to perform image and video source identification for the same device than computing two different reference SPNs, one for still images and one for videos, respectively.

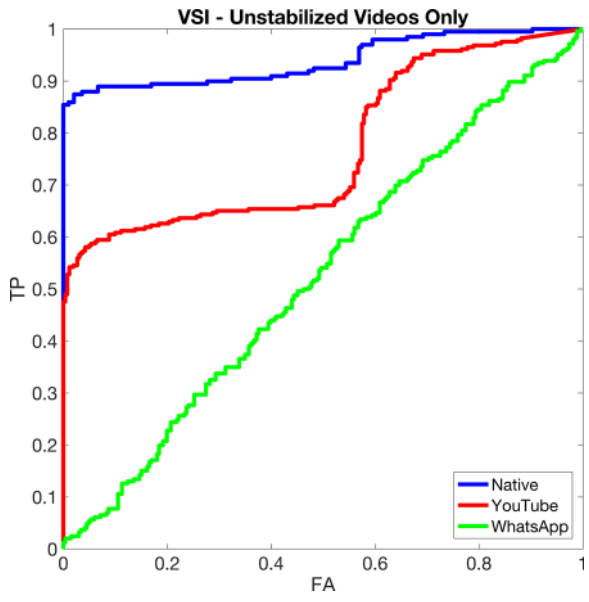

Fig. 4 (Best viewed in colors) The VSI performance on Native, YouTube, and WhatsApp videos (in blue, red and green respectively) considering only devices devices without in-camera digital stabilization

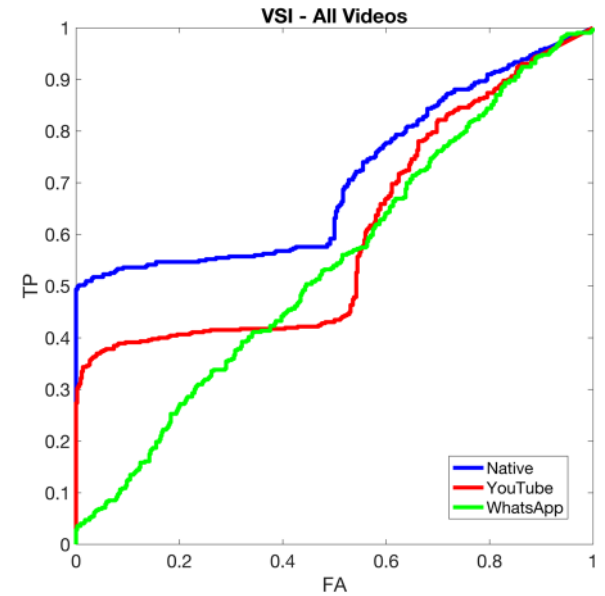

Fig. 5 (Best viewed in colors) The VSI performance on Native, YouTube, and WhatsApp videos (in blue, red and green respectively) considering all available devices

A first step towards an integration of these cases is a hybrid source identification (HSI) approach that exploits still images for estimating the fingerprint that will be used to verify the source of a video, as proposed in [4]. Authors of [4] investigate the geometrical relation between image and video acquisition processes. Indeed, even if the sensor is the same, videos are usually acquired at a much lower resolution than images: top-level smartphones reach 4K video resolution at most (8 megapixels per frame), but can easily capture 20-megapixel images. To achieve that, in video recording, a central crop is carried so to adapt the sensor size to the desired aspect ratio (commonly 16:9), then the selected pixels are scaled to match the desired video resolution. As a direct consequence, the fingerprints extracted from images and videos cannot be directly compared and most of the times, because of cropping, it is not sufficient to just scale them to the same resolution. Instead, image-based and video-based fingerprints are linked by the cropping and scaling factors between image and video sensor portion, that usually change across different device models.

With the aim of facilitating researchers exploring the HSI framework within the VISION dataset, we provide the cropping and scaling factor for several devices contained therein. For simplicity, we limit to non-stabilized devices; the hybrid analysis for stabilized devices is even more complex, and is one of the future research scopes this dataset has been built for. In order to estimate

Table 3 Dataset paths for VSI experiments

Experiment	Test path	AUC	
		All Videos	Unstab. Videos
1	ID_Brand_Model/videos/flat ID_Brand_Model/videos/indoor ID_Brand_Model/videos/outdoor	0.7069	0.9394
2	ID_Brand_Model/videos/flatYT ID_Brand_Model/videos/indoorYT ID_Brand_Model/videos/outdoorYT	0.6032	0.7700
3	ID_Brand_Model/videos/flatWA ID_Brand_Model/videos/flatWA ID_Brand_Model/videos/flatWA	0.5262	0.5437

cropping and scaling factors, for each device we estimated the video and image references from the videos contained in `ID_Brand_Model/videos/flat` and from the images in `ID_Brand_Model/images/flat`, respectively. Specifically, we estimated each image reference fingerprint from 100 flat field images and each video reference fingerprint from 100 frames of a flat field video. The cropping and scaling factors are estimated by a brute force search, as suggested in [25]. In Table 4 we report the scaling factor and the corresponding cropping corner (upper-left corner along x and y axes) yielding the maximum PCE for each examined device. We consider the parameter search unsuccessful if the obtained maximum PCE is lower than 50 (denoted by "n.a." in Table 4). For instance, with the device $D11$ an image fingerprint should be scaled by a factor 0.59 and then cropped on the upper left side of 307 pixels along the y axis to match the video fingerprint (the right and down cropping are derived by the corresponding video size). $D13$ is a pretty unique case in which the full frame is applied for videos and the left (and right) cropping of 160 pixels is applied to capture images. We put a -160 meaning that the video frame is cropped by 160 pixel to capture images. Finally, we notice that we were not able to register the fingerprints for the devices $D21$ and $D26$ by means of the presented techniques. A deeper analysis of the registration techniques is still an open topic.

5 Conclusions

In this paper, we propose a new image and video dataset useful for benchmarking multimedia forensic tools. We collected thousands of images and videos from portable devices of most famous brands, including those featuring in-camera digital stabilization. We also prepared the "social version" of most contents, by uploading and downloading them to/from well-known social media platforms, namely Facebook, YouTube and WhatsApp.

We showed examples of some popular applications that would benefit from the proposed dataset, such as the video source identification, for which there are no sizeable benchmarks available in the research community. Furthermore, we showed how this dataset allows the exploration of new forensic opportunities such as comparing camera reference fingerprints estimated from still images and from videos. The whole dataset is made available[6] to the research community, along with a guide that clarifies its structure and several csv files containing technical information.

Although VISION is a huge collection of media contents, we believe that there is space for future improvements, indeed we are currently working to extend VISION with more videos, by means of a mobile application (MOSES [26]) from which videos can be captured and uploaded directly to our servers, following the main concept and guidelines of VISION. In conclusion, VISION

Table 4 Estimated Cropping and Scaling factors for non stabilized videos

ID	D01	D03	D07	D08	D09	D11	D13	D16	D17	D21
Scaling	0.5	0.48	0.27	1	0.61	0.59	1	0.46	0.59	n.a.
Cropping [x y]	[0 228]	[0 372]	[0 7]	[408 354]	[227 411]	[0 307]	[-160 0]	[8 396]	[0 1]	n.a.
ID	D24	D26	D27	D28	D30	D31	D32	D33	D35	D22
Scaling	0.5	n.a.	0.5	0.36	0.47	0.46	0.59	0.52	0.39	0.49
Cropping [x y]	[0 240]	n.a.	[0 228]	[0 0]	[39 10]	[9 397]	[0 0]	[464 693]	[0 306]	[0 246]

provides the contents needed to assess the performance of next generation image and video forensic tools.

Endnotes

[1] Not all the videos were exchanged through social media platforms. The technical details are explained in the Appendix

[2] Facebook website on March 2017.

[3] ClipGrab v3.6.3 - www.clipgrab.org

[4] youtube-dl v2017.03.10 - http://rg3.github.io/youtube-dl/

[5] We recommend downloading less than 20 videos at a time due to the YouTube policy.

[6] https://lesc.dinfo.unifi.it/en/datasets

[7] Note that the Rotation tag, related to the correspondent video standard, of each video is reported in the additional material released together with the dataset.

Appendix

In this section, we provide more detailed information on the VISION dataset. First of all, in Appendix Table 5, we review the version of the software/firmware of each device, along with the number of native images and videos (in columns #natOrigin and #vOrigin) and their social counterparts (in columns #natSocial, #vSocial). We would like the reader to notice that the amount of social media videos should be two times the original amount, since we uploaded all the videos in YouTube and WhatsApp. The reader can easily see that this is not true for some devices, indeed we removed 30 videos from the social folders because of issues due to the uploading system used by YouTube and WhatsApp. We acquired videos from each device using the landscape-mode, but a small amount of the overall acquisitions were captured in portrait mode. Unfortunately this was an issue with the YouTube uploading, because in this case, the video is modified by black padding: this raised a consistency problem and we decided to remove these contents. On the other hand, WhatsApp was not affected by this kind of problems, and the same portrait-mode videos were uploaded and downloaded correctly. Unfortunately, we could not upload 4 videos with the WhatsApp iOS application, and so far did not understand the reason causing this problem.

The devices affected by these problems were the D34 in which we exluded 14 videos, the D01 with 6 videos missing, the D08 with 4 files, the D06 with 2 files and one file from devices D04, D21, D22, and D33. Detailed information are given, in form of multiple CSV file, with the current Dataset.

In Appendix Table 6 we report for each device some statistics computed over all native images by means of *Exiftool 10.10*. In detail, the metadata tags that can be extracted with *Exiftool* are the *Lens Model*, *ISO*, *Aperture*, *Flash*, *Focal Length*, *Image Size* (represented in column Image Resolution) and *Orientation*.

All images were acquired in the JPEG format using the encoder *JPEG old-style*, and for most devices the color sub sampling is set to *YCbCr 4:2:0 (2,2)*, while D07 and D11 use *YCbCr 4:2:2 (2,1)* and only D28 supports both sub samplings.

The columns create/modify date and GPS present are counters: the first one counts the number of images in which the metadata *create date* and *modify date* are not identical. The second refers to the metadata *GPS Position* and counts the number of images in which this tag is not empty.

In Appendix Table 6, we collected information related to the Lens specifics such as the *focal length*, the *aperture value* or the *ISO*. It is worth noting that the *Lens model* tag is present only for Apple devices, and specifies which camera is used and some features of the lens. The *ISO* column contains a range of values for each device, that is the minimum and the maximum ISO value observed in the images metadata from that device. The Flash column lists all encountered values, indeed for some devices such as D05, we have images acquired with Flash in auto mode, off mode, fired, or not fired mode. In case of devices D13 and D20, namely iPad 2 and iPad mini, the flash function does not exists, since these devices are not equipped with a flash.

The *Image Size* tag is present in column *image resolution* and reports the resolution as width$\times$height; let us note that in devices D24 and D33 the image Orientation tag is not always present, when needed it is derived from their resolution in order to distinguish between landscape and portrait acquisitions such as 4608×2592 and 2592×4608 for D24. In all the remaining devices, each image is stored in *landscape* mode, that is horizontally (H); if the image was actually acquired with a different orientation, this is reflected in the Orientation tag, which may contain the values: Rotation 180 (R 180), Rotation 90 Clock-Wise (R 90 CW) or Rotation 270 Clock-Wise (R 270 CW).

From a forensics point of view the *create/modify date* is very interesting: in Appendix Table 6 the reader can see that for all Apple devices in this Dataset, the Create Date and Modify Date tags are different (when available), meaning that Apple devices store into Create Date the moment in which the *photo shoot* is computed and in Modify Date the moment in which the image is stored, that is, typically, a few seconds later. All the other devices set these tags to the same initial value.

In Appendix Table 7, we report for each device some statistical information computed over the recorded videos. The metadata tags gathered from the native videos by means of *Exiftool* were *file type*, *major brand*, *video frame*

Table 5 Devices featured in VISION

Brand	Model	ID	Software/firmware	#vOrigin	#vSocial	#natOrigin	#natSocial
Apple	iPad 2	D13	iOS 7.1.1	16	32	171	513
Apple	iPad mini	D20	iOS 8.4	16	32	159	477
Apple	iPhone 4	D09	iOS 7.1.2	19	38	217	651
Apple	iPhone 4S	D02	iOS 7.1.2	13	26	204	612
Apple	iPhone 4S	D10	iOS 8.4.1	15	30	178	534
Apple	iPhone 5	D29	iOS 9.3.3	19	38	224	672
Apple	iPhone 5	D34	iOS 8.3	32	50	204	612
Apple	iPhone 5c	D05	iOS 10.2.1	19	38	350	1050
Apple	iPhone 5c	D14	iOS 7.0.3	19	38	209	627
Apple	iPhone 5c	D18	iOS 8.4.1	13	26	204	612
Apple	iPhone 6	D06	iOS 8.4	17	32	132	396
Apple	iPhone 6	D15	iOS 10.1.1	18	36	227	681
Apple	iPhone 6 Plus	D19	iOS 10.2.1	19	38	259	777
Asus	Zenfone 2 Laser	D23*	–	19	38	210	630
Huawei	Ascend G6-U10	D33	–	19	37	155	465
Huawei	Honor 5C NEM-L51	D30	Android 6.0/NEM-L51C432B120	19	38	271	813
Huawei	P8 GRA-L09	D28	Android 6.0/GRA-L09C55B330	19	38	266	798
Huawei	P9 EVA-L09	D03	Android 6.0/EVA-L09C55B190	19	38	237	711
Huawei	P9 Lite VNS-L31	D16	Android 6.0/VNS-L31C02B125	19	38	235	705
Lenovo	Lenovo P70-A	D07	–	19	38	217	651
LG electronics	D290	D04	–	19	37	227	681
Microsoft	Lumia 640 LTE	D17	Windows Phone	10	20	188	564
OnePlus	A3000	D25	Android 7.0/NRD90M 15 dev-keys	19	38	287	861
OnePlus	A3003	D32	Android 7.0/NRD90M 138 dev-keys,	19	38	236	708
–	–	–	NRD90M 18 dev-keys	–	–	–	–
Samsung	Galaxy S III Mini GT-I8190	D26	I8190XXAMG4	16	32	150	450
Samsung	Galaxy S III Mini GT-I8190N	D01	I8190NXXAML1, I8190NXXALL6	22	38	205	615
Samsung	Galaxy S3 GT-I9300	D11	–	19	38	207	621
Samsung	Galaxy S4 Mini GT-I9195	D31	I9195XXUCNK1	19	38	216	648
Samsung	Galaxy S5 SM-G900F	D27	Android 6.0.1/G900FXXS1CQAA	19	38	254	762
Samsung	Galaxy Tab 3 GT-P5210	D08	P5210XXUBNK2	37	70	168	504
Samsung	Galaxy Tab A SM-T555	D35	T555XXU1AOE9	16	32	154	462
Samsung	Galaxy Trend Plus GT-S7580	D22	S7580XXUBOA1	16	31	163	489
Sony	Xperia Z1 Compact D5503	D12	14.5.A.0.270_6_f100000f	19	38	216	648
Wiko	Ridge 4G	D21	–	11	21	253	759
Xiaomi	Redmi Note 3	D24	Android 6.0.1/MMB29M	19	38	312	936
–	–	–	V8.1.1.0.MHOMIDI release-keys	–	–	–	–

rate, *media duration*, *audio channels*, *audio sample rate*, *image size*, and *rotation*[7].

In order to make Appendix Table 7 clearer, we used *video resolution* instead of the tag *Image Size*, and we did not report the tags related to the audio acquisition, although they will be described in the following paragraphs.

All videos were acquired using the video encoder *H.264/avc1* and *mp4a* for encoding audio. We remark that, for the D23 device videos were not captured at the maximum resolution available, as opposed to all other acquisitions. Similarly to Appendix Table 6, we included in Appendix Table 7 the columns create/modify date and

Table 6 Devices' image characteristics in VISION

ID	Lens model	ISO	Aperture	Flash	Focal length (mm)	Image resolution	Orientation	Create/modify date	GPS present
D01	–	50, 400	–	No flash, fired	3.5	2560 × 1920	H, R 180, R 270 CW, R 90 CW	-	-
D02	b.c 4.28mm f/2.4	50, 800	2.4	Off, did not fire	4.3	3264 × 2448	H, R 180	307	204
D03	–	50, 500	2.2	Auto, did not fire	4.5	3968 × 2976	H	-	272
D04	–	100, 1200	–	No flash	3.2	3264 × 2448	H	-	-
D05	b.c 4.12mm f/2.4	50, 3200	2.4	Auto, fired, off, did not fire	4.1	3264 × 2448	H, R 180	463	463
D06	b.c 4.15mm f/2.2	32, 800	2.2	Off, did not fire, auto	4.2	3264 × 2448	H, R 180, R 90 CW, R 270 CW	281	281
D07	–	95, 355	–	No flash	3.5	4784 × 2704	H	-	-
D08	–	50, 400	–	No flash	2.8	2048 × 1536	H, R 270 CW, R 90 CW	-	-
D09	b.c 3.85mm f/2.8	80, 1000	2.8	Auto, did not fire	3.9	2592 × 1936	H, R 90 CW, R 180	326	318
D10	S b.c 4.28mm f/2.4	50, 160	2.4	Auto, did not fire	4.3	3264 × 2448	H, R 180, R 90 CW, R 270 CW	311	311
D11	–	80, 640	2.6	No flash	3.7	3264 × 2448	H, R 270 CW, R 90 CW	-	301
D12	–	50, 800	–	On, fired, off, did not fire	4.9	5248 × 3936	H, R 270 CW, R 90 CW	-	174
D13	b.c 2.03mm f/2.4	40, 640	2.4	No flash function	2.0	960 × 720	H, R 90 CW, R 180, R 270 CW	330	-
D14	b.c 4.12mm f/2.4	50, 2000	2.4	Off, did not fire	4.1	3264 × 2448	H	339	-
D15	b.c 4.15mm f/2.2	32, 400	2.2	Off, did not fire, Auto	4.2	3264 × 2448	H	337	-
D16	–	50, 640	2	Auto, did not fire	3.8	4160 × 3120	H	-	-
D17	–	64, 3200	2.2	On, fired, off, did not fire	3.0	3264 × 1840	H, R 90 CW	-	285
D18	b.c 4.12mm f/2.4	50, 3200	2.4	Off, did not fire	4.1	3264 × 2448	H, R 180, R 270 CW, R 90 CW	305	305
D19	b.c 4.15mm f/2.2	32, 125	2.2	Off, did not fire	4.2	3264 × 2448	R 180	428	-
D20	b.c 3.3mm f/2.4	25, 640	2.4	No flash function	3.3	2592 × 1936	H, R 180, R 90 CW, R 270 CW	278	271
D21	–	100, 913	2.5	Off, did not fire	4.6	3264 × 2448	-	-	-
D22	–	50, 100	–	No flash	3.5	2560 × 1920	H, R 180, R 90 CW, R 270 CW	-	-
D23	–	50, 460	2	Off, did not fire	4.6	3264 × 1836	-	-	-
D24	–	100, 873	2	Off, did not fire	3.6	4608 × 2592, 2592 × 4608	-	-	466
D25	–	100, 1000	2	Off, did not fire	4.3	4640 × 3480	-	-	-
D26	–	50, 200	–	No flash, fired	3.5	2560 × 1920	H, R 180, R 270 CW, R 90 CW	-	-
D27	–	40, 400	2.2	No flash	4.8	5312 × 2988	H	-	-
D28	–	64, 1000	2	Auto, fired, did not fire	3.8	4160 × 2336	H	-	-
D29	b.c 4.12mm f/2.4	50, 800	2.4	Off, did not fire	4.1	3264 × 2448	H	324	-
D30	–	50, 2000	2	Auto, did not fire	3.8	4160 × 3120	H	-	-

Table 6 Devices' image characteristics in VISION

ID	Lens model	ISO	Aperture	Flash	Focal length (mm)	Image resolution	Orientation	Create/modify date	GPS present
D31	–	50, 1000	2.6	No flash	3.7	3264 × 1836	H, R 180, R 270 CW, R 90 CW	-	-
D32	–	100, 2500	2	Off, did not fire	4.3	4640 × 3480	-	-	58
D33	–	100, 2400	2	Auto, fired, did not fire	3.0	2448 × 3264, 2448x3264	H	-	236
D34	b.c 4.12mm f/2.4	50, 80	2.4	Off, did not fire	4.1	3264 × 2448	H, R 180, R 90 CW	310	-
D35	–	50, 160	2.2	No flash	3.3	2592 × 1944	H, R 180, R 90 CW, R 270 CW	-	280

Table 7 Devices' video characteristics in VISION

ID	File type	Video format	Frame rate	Media duration	Video resolution	Rotation	Create/modify date	GPS present
D01	MP4	MP4 Base Media v1	28.986, 30.233	0:01:08, 0:01:13	1280 × 720	0, 90	–	–
D02	MOV	Apple QuickTime (.MOV/QT)	24.009, 29.97	0:00:59, 0:01:12	1920 × 1080	0, 180	13	11
D03	MP4	MP4 v2 ISO 14496-14	30.011, 30.033	0:01:10, 0:01:17	1920 × 1080	0, 180	–	–
D04	MP4	MP4 Base Media v1	26.038, 30.024	0:01:10, 0:01:14	800 × 480	0	–	–
D05	MOV	Apple QuickTime (.MOV/QT)	25.008, 29.973	0:00:26.78, 0:01:18	1920 × 1080	0, 180	19	19
D06	MOV	Apple QuickTime (.MOV/QT)	30.006, 30.006	0:01:06, 0:01:11	1920 × 1080	0, 90	17	17
D07	3GP	3GPP Media (.3GP) Release 4	22.694, 30.004	0:01:11, 0:01:14	1280 × 720	0, 180	–	–
D08	MP4	MP4 Base Media v1	24.016, 29.686	0:01:01, 0:01:11	1280 × 720	0, 90, 270	–	–
D09	MOV	Apple QuickTime (.MOV/QT)	28.621, 29.969	0:01:09, 0:01:16	1280 × 720	0	19	17
D10	MOV	Apple QuickTime (.MOV/QT)	24.01, 29.97	0:01:10, 0:01:13	1920 × 1080	0	15	15
D11	MP4	MP4 Base Media v1	29.978, 30.006	0:01:11, 0:01:17	1920 × 1080	0	–	3
D12	MP4	MP4 v2 ISO 14496-14	29.904, 29.982	0:01:10, 0:01:24	1920 × 1080	0	–	–
D13	MOV	Apple QuickTime (.MOV/QT)	28.727, 29.967	0:01:09, 0:01:13	1280 × 720	0	16	–
D14	MOV	Apple QuickTime (.MOV/QT)	29.973, 29.973	0:01:12, 0:01:13	1920 × 1080	0	19	–
D15	MOV	Apple QuickTime (.MOV/QT)	29.983, 29.984	0:01:02, 0:01:05	1920 × 1080	0	18	–
D16	MP4	MP4 v2 ISO 14496-14	30.167, 30.205	0:01:10, 0:01:14	1920 × 1080	0, 180	–	–
D17	MP4	MP4 v2 ISO 14496-14	30.008, 30.008	0:01:10, 0:01:11	1920 × 1080	0	–	–
D18	MOV	Apple QuickTime (.MOV/QT)	24.003, 29.973	0:01:09, 0:01:12	1920 × 1080	0	13	13
D19	MOV	Apple QuickTime (.MOV/QT)	29.983, 30.0	0:01:10, 0:01:15	1920 × 1080	0, 180	19	–
D20	MOV	Apple QuickTime (.MOV/QT)	29.576, 29.972	0:01:10, 0:01:12	1920 × 1080	0	16	7
D21	MP4	MP4 v2 ISO 14496-14	18.893, 29.246	0:01:07, 0:01:26	1920 × 1080	0	–	–
D22	MP4	MP4 Base Media v1	30.03, 30.034	0:01:10, 0:01:20	1280 × 720	0	–	–
D23*	MP4	MP4 v2 ISO 14496-14	29.891, 29.903	0:01:09, 0:01:14	640 × 480	0	–	–
D24	MP4	MP4 v2 ISO 14496-14	20.215, 30.062	0:01:07, 0:01:13	1920 × 1080	0	–	11
D25	MP4	MP4 v2 ISO 14496-14	29.999, 30.01	0:01:10, 0:01:17	1920 × 1080	0, 180	–	–
D26	MP4	MP4 Base Media v1	29.262, 30.237	0:01:11, 0:01:14	1280 × 720	0	–	–
D27	MP4	MP4 v2 ISO 14496-14	29.97, 30.006	0:00:25.54, 0:01:16	1920 × 1080	0	–	–
D28	MP4	MP4 v2 ISO 14496-14	29.636, 29.886	0:01:09, 0:01:16	1920 × 1080	0, 180	–	–
D29	MOV	Apple QuickTime (.MOV/QT)	24.003, 29.973	0:01:10, 0:01:15	1920 × 1080	0	19	–
D30	MP4	MP4 v2 ISO 14496-14	30.14, 30.157	0:01:08, 0:01:17	1920 × 1080	0, 180	–	–
D31	MP4	MP4 Base Media v1	29.927, 30.013	0:01:12, 0:01:16	1920 × 1080	0	–	–
D32	MP4	MP4 v2 ISO 14496-14	29.898, 30.01	0:01:09, 0:01:16	1920 × 1080	0	–	2
D33	MP4	MP4 Base Media v1	24.967, 30.03	0:01:10, 0:01:14	1280 × 720	0, 90	–	19
D34	MOV	Apple QuickTime (.MOV/QT)	24.003, 29.973	0:01:08, 0:01:32	1920 × 1080	0, 90	32	–
D35	MP4	MP4 v2 ISO 14496-14	29.873, 30.012	0:01:10, 0:01:13	1280 × 720	0	–	16

GPS present. The former counts the number of videos in which metadata *create date* and *modify date* are not identical, while the latter counts the number of videos in which the *GPS location* tag is not empty.

We highlight that the values in the create/modify date column are different than 0 for Apple devices only: indeed, in these devices the create date differs from the modify date by the duration of the video.

All Apple devices store videos with the Apple QuickTime container (.MOV extension), and most of Android devices store video as MP4 using the H.264 encoder in two versions: *MP4 Base v1, ISO 14496-12* or *MP4 Base v2, ISO 14496-14*. It is worth mentioning that the Lenovo device (D07) encodes videos in H.264 but uses the container 3GP. Almost all videos last more than one minute, with the exception of a few in devices D02, D05,

and D27 where the shortest video duration is 25 seconds. Most devices record videos at a frame rate of 24 fps or more; as exceptions we have some videos from D07, D21 and D24.

As to the characteristics of the encoded audio, as a general distinction Apple devices acquire audio using one channel at sample rate of 44100 bit/s, whereas Android devices usually acquire two audio channels at 48000 bit/s. The Microsoft device (D17) uses one audio channel at 48000 bit/s, like some Samsung Devices, namely: Galaxy S III Mini (D01, D26) and the Galaxy Trend Plus (D22).

Abbreviations

Exif: Exchangeable image file format; HSI: Hybrid source identification; ISI: Image source identification; MF: Multimedia forensics; PCE: Peak to correlation energy; ROC: Receiver operating characteristic; SMP: Social media platform; SPN: Sensor pattern noise; VSI: Video source identification

Availability of data and materials

As stated in Section 5, the whole dataset will be made available at no cost to the research community at the following web address https://lesc.dinfo.unifi.it/en/datasets, with the goal of providing the content needed to assess the performance of next generation of image and video forensic tools. In addition to the dataset we release a guide to the dataset structure and several csv files containing names and technical information, such as metadata tags, of all the collected media.

Authors' contributions

DS contributed to acquisition of data (especially for obtaining the social-network version of videos), prepared all the statistics about data and corresponding metadata, and drafted part of the manuscript. MF took care of writing code, analyse and present results for the experimental section about image and video source identification. MI contributed to drafting the paper and writing code, and analysed the relationships between sensor pattern noise in images and videos. OAS carried out most of the data acquisition, checked the gathered material and contributed to obtaining the social-network version of images. AP conceived the study, participated to its design and coordination, and helped to draft the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details

[1]Department of Information Engineering, University of Florence, Via di S. Marta, 3, 50139 Florence, Italy. [2]FORLAB, Multimedia Forensics laboratory, PIN Scrl, Piazza G. Ciardi, 25, 59100 Prato, Italy. [3]Department of electronic Media, Saudi Electronic University, Abi Bakr As Sadiq Rd, Riyadh, 11673, Saudi Arabia.

Received: 10 May 2017 Accepted: 18 September 2017
Published online: 03 October 2017

References

1. Statista Inc., Statista. http://www.statista.com/statistics/263437/global-smartphone-sales-to-end-users-since-2007/. Accessed 22 Sept 2017
2. A De Rosa, A Piva, M Fontani, M Iuliani, in *2014 International Carnahan Conference on Security Technology (ICCST)*. Investigating multimedia contents (IEEE, Rome, 2014), pp. 1–6
3. A Piva, An overview on image forensics. ISRN Signal Proc. **2013**, 496701–22 (2013)
4. M Iuliani, M Fontani, D Shullani, A Piva, A hybrid approach to video source identification. arXiv:1705.01854[cs.MM] (2017)
5. J Lukas, J Fridrich, M Goljan, Digital camera identification from sensor pattern noise. IEEE Trans. Inf. Forensic Secur. **1**(2), 205–214 (2006)
6. W Van Houten, Z Geradts, Source video camera identification for multiply compressed videos originating from youtube. Digit. Investig. **6**(1), 48–60 (2009)
7. M Chen, J Fridrich, M Goljan, J Lukáš, in *Proc. of SPIE 6515 Electronic Imaging 2007*. Source digital camcorder identification using sensor photo response non-uniformity, (2007), pp. 65051–65051. International Society for Optics and Photonics
8. W-H Chuang, H Su, M Wu, in *IEEE International Conference on Image Processing (ICIP)*. Exploring compression effects for improved source camera identification using strongly compressed video (IEEE, Brussels, 2011), pp. 1953–1956
9. S Chen, A Pande, K Zeng, P Mohapatra, Live video forensics: source identification in lossy wireless networks. IEEE Trans. Inf. Forensic Secur. **10**(1), 28–39 (2015)
10. G Schaefer, M Stich, in *Proc. SPIE 5307 Electronic Imaging 2004*. Ucid: an uncompressed color image database, (2003), pp. 472–480. International Society for Optics and Photonics
11. T Gloe, R Böhme, The Dresden image database for benchmarking digital image forensics. J. Digit. Forensic Pract. **3**(2-4), 150–159 (2010)
12. T Gloe, R Böhme, in *Proceedings of the 25th Symposium On Applied Computing (ACM SAC 2010)*. The 'Dresden Image Database' for benchmarking digital image forensics, vol. 2 (ACM New York, Sierre, 2010), pp. 1585–1591
13. D-T Dang-Nguyen, C Pasquini, V Conotter, G Boato, in *Proceedings of the 6th ACM Multimedia Systems Conference*. MMSys '15. Raise: a raw images dataset for digital image forensics (ACM, New York, 2015), pp. 219–224
14. D Vázquez-Padín, F Pérez-González, in *2011 IEEE International Workshop on Information Forensics and Security*. Prefilter design for forensic resampling estimation (IEEE, Iguacu Falls, 2011), pp. 1–6
15. G Qadir, S Yahaya, ATS Ho, in *Proceedings of the IET IPR 2012, 3-4 July, London*. Surrey University Library for Forensic Analysis (SULFA), (2012)
16. L D'Amiano, D Cozzolino, G Poggi, L Verdoliva, in *Multimedia & Expo Workshops (ICMEW), 2015 IEEE International Conference On*. Video forgery detection and localization based on 3d patchmatch (IEEE, Turin, 2015), pp. 1–6
17. OI Al-Sanjary, AA Ahmed, G Sulong, Development of a video tampering dataset for forensic investigation. Forensic Sci. Int. **266**, 565–572 (2016)
18. F Bertini, R Sharma, A Iannı, D Montesi, MA Zamboni, in *The International Conference on Computing Technology, Information Security and Risk Management (CTISRM2016)*. Social media investigations using shared photos, (Dubai, 2016), p. 47
19. M Moltisanti, A Paratore, S Battiato, L Saravo, in *Image Analysis and Processing - ICIAP 2015 - 18th International Conference, Genoa, Italy, September 7-11, 2015, Proceedings, Part II*. Image manipulation on facebook for forensics evidence (Springer, Genoa, 2015), pp. 506–517
20. Z Wang, AC Bovik, HR Sheikh, EP Simoncelli, Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
21. D Bolton, youtube-dl documentation. github.com/rg3/youtube-dl/blob/master/README.md#readme. Accessed 22 Sept 2017
22. M Chen, J Fridrich, M Goljan, J Lukáš, Determining image origin and integrity using sensor noise. IEEE Trans. Inf. Forensic Secur. **3**(1), 74–90 (2008)
23. M Goljan, J Fridrich, T Filler, in *Publisher: Proc. SPIE 7254 IS&T/SPIE Electronic Imaging*. Large scale test of sensor fingerprint camera identification, (2009), pp. 72540–72540. International Society for Optics and Photonics
24. S Taspinar, M Mohanty, N Memon, in *2016 IEEE International Workshop on Information Forensics and Security (WIFS)*. Source camera attribution using stabilized video (IEEE, Abu Dhabi, 2016), pp. 1–6

25. M Goljan, J Fridrich, Camera identification from scaled and cropped images. Secur. Forensic Steganography Watermarking Multimedia Contents X. **6819**, 68190 (2008)
26. D Shullani, O Al Shaya, M Iuliani, M Fontani, A Piva, in *Proceeding of 2017 Tyrrhenian International Workshop on Digital Communications, Communications in Computer and Information Science, vol. 766, September 18–20, 2017, Palermo.* A Dataset for forensic analysis of videos in the wild, (2017), pp. 84–94