A Forensic Analysis of Images on Online Social Networks

Aniello Castiglione*, Giuseppe Cattaneo†, Alfredo De Santis‡
Dipartimento di Informatica "R.M. Capocelli"
Università degli Studi di Salerno
I-84084 Fisciano (SA), Italy
castiglione@ieee.org*, cattaneo@dia.unisa.it†, ads@dia.unisa.it‡

Abstract—The Web 3.0 is approaching fast and the Online Social Networks (OSNs) are becoming more and more pervasive in today daily activities. A subsequent consequence is that criminals are running at the same speed as technology and most of the time highly sophisticated technological machineries are used by them. Images are often involved in illicit or illegal activities, with it now being fundamental to try to ascertain as much as information on a given image as possible. Today, most of the images coming from the Internet flow through OSNs.

The paper analyzes the characteristics of images published on some OSNs. The analysis mainly focuses on how the OSN processes the uploaded images and what changes are made to some of the characteristics, such as JPEG quantization table, pixel resolution and related metadata. The experimental analysis was carried out in June-July 2011 on Facebook, Badoo and Google+. It also has a forensic value: it can be used to establish whether an image has been downloaded from an OSN or not.

Index Terms—Online Social Networks; OSN; Digital Forensics; Image Forensics; Quantization Table; Quality Factor; Facebook; Google+; Badoo; Pixel Resolution; Metadata.

I. INTRODUCTION

Online Social Networks (OSNs) are becoming more and more popular. Their growth is almost exponential. Facebook, launched in 2004, had over 100 million active users in July 2008, over 250 million in July 2009, over 500 million in July 2010 and over 750 million in July 2011 [1]. Facebook has been designed to easily share information such as messages and photos, and nowadays over 30 billion contents (photo albums, notes, web links, stories, etc.) are shared monthly on it. Above all, Facebook publishes a huge number of user photos which is growing at a rate of more than 3 billion of uploads per month. This produces a traffic of more than 1.2 million photos per second during peak time [2]. Similarly Badoo, launched in November 2006, had about 125 million active users in July 2011 and a rate of 1.8 million user photos and videos uploaded every day [3]. Finally, even more surprising, Google+ in the first month of its launch reached 25 million users registrations, while Twitter and Facebook took about 3 years [4].

These statistics show how fast OSNs are growing and how pervasive their use is in everyday activities. A consequence is that criminals runs at the same speed as technology and most of the time technological machineries are used by them.

Images are often involved in illicit or illegal activities, with it now being fundamental to try to ascertain as much as information on a given image as possible. Today, most of the images available on the Internet flow through OSNs. In [5], an analysis of the Lukas et al. [6] technique on images published by common OSNs was presented. This technique enables source camera identification by extracting the PRNU (Photo-Response Non-Uniformity) sensor noise from digital images.

In this paper, the following three characteristics of digital images published on some OSNs have been analyzed.

- *Image format.* Images can be encoded using different formats such as JPEG, BMP, and PNG. JPEG is the most used format. One of the analyzed aspects is the JPEG quantization tables selected by the OSN during the publishing process.
- *Metadata.* It provides information that supplements the primary content of digital documents such as file name, creation or modification date, orientation, creator, location or comments.
- *Pixel resolution.* Size of the image expressed in number of pixels for each row and each column.

An analysis of the characteristics focuses mainly on how OSNs process the uploaded images and what changes are made to their characteristics. The authors have not been able to find any documentation on this process either on the OSNs or in current literature. Information, mainly on the pixel resolution, can be found on forums and blogs. The following methodology was used:

1) First, a data set with images generated by several different brands of digital camera was created. It constitutes a heterogeneous data set with different characteristics, i.e. image format, metadata and pixel resolution.
2) Then, each image was uploaded and successively downloaded from each target OSN.
3) Finally, each input image was compared to the corresponding downloaded image in order to analyze how the OSN publication process modified the images with respect to the aforementioned characteristics.

The image analysis used the following tools: Exiftool, GIMP, IrfanView, JPEGSnoop, Matlab with the IPT package. Most of them run on both Windows and Linux OS.

However, the policies for the content management, and

particularly the publication process, may change over time depending on marketing issues as well as technical factors such as disk space or bandwidth availability. The analysis described in this work is based on experiments carried out in June-July 2011 accessing the target OSNs using authors and class students profiles according to the methodology described above.

The analysis focused on the following three OSNs: Facebook, Badoo and Google+. The experimental results show that all the target OSNs change the pixel resolution and metadata of the uploaded images to fixed values. Facebook and Badoo compress the images using predefined JPEG quantization tables. Therefore, every image downloaded by a given OSN presents known values for some characteristics. As a consequence, this analysis is useful in Image Forensics. If a given image matches all the predefined values of the relative characteristics of one of the OSNs, then it might have been downloaded from that OSN. Otherwise, it was not downloaded in its actual form.

The paper is organized as follows. Section II describes the image types published on the considered OSNs. The three aforementioned characteristics, i.e., image format, metadata and pixel resolution are analyzed in Sections III, IV, and V, respectively. Section VI presents some remarks about the Digital Forensics applications of the paper findings. Conclusions are made in Section VII.

II. The Online Social Networks and the Images

In this paper, three OSNs have been considered, namely Facebook, Badoo and Google+. Facebook is the most widely used OSN with about 750 million users all over the world. Badoo is mainly active in Latin America, France, Spain, and Italy [7]. Google+, although being the last to be launched and still in beta version, represents the most innovative and competitive OSN thanks to its integration with the Google services. The official opening to all users (not only to invited ones), the new features introduced, as well as the Facebook counter-moves, promise to produce a true revolution in the OSN scenario.

All of them offer features which allow their users to share images along with comments and user references.

It is possible to divide the published images into three types:

- *User supplied images* are uploaded with a "good" resolution and can be organized into albums or associated to user profiles. OSNs provide a publication service which lets the user upload their own images. This process defines some constraints for the images to be accepted for publication, such as image format and size. Some OSNs, during the upload process, let the user choose from different resolutions. For example, Facebook proposes two resolutions referred to here as *standard* and *high*. All these images are big enough to fit one browser page or are displayed as an album slideshow.
- *Thumbnails* that are the reduced-size version of the uploaded images used to help recognize and organize

them. They are produced using scaling/cropping operations on the user supplied images. These are mostly used as placeholders in the "walls" to identify the user or hypertext links to other contents.
- *Advertisement* images, supplied by the OSN's marketing services, on which the user has no control. This kind of images were not considered in the analysis.

Each OSN uses its own custom strategy to display images at the appropriate resolution according to the environment. For example, Badoo manages four thumbnail sizes to provide users with the best quality possible. In order to avoid overburdening the presentation, before starting the analysis, a shorthand notation was established for all the different kinds of images displayed on the three OSNs considered. In the case of Facebook, the following image types were considered:

FB_{hi} User supplied images with high resolution.
FB_{st} User supplied images with standard resolution.
FB_{pr} Profile pictures, i.e., the images associated with the user and generally displayed on its home page.
FB_{th} Small thumbnails at the lowest resolution.

Badoo does not give the possibility to choose among different resolutions, but derives four different thumbnails of different sizes from the uploaded image.

BD_{st} User supplied images with standard resolution.
BD_{th1} Thumbnails at the highest resolution.
BD_{th2} Thumbnails at medium resolution.
BD_{th3} Thumbnails at low resolution.
BD_{th4} Thumbnails at the lowest resolution.

Finally, Google+ has only one resolution for user supplied images and a fixed size thumbnail.

$G+_{st}$ User supplied images with standard resolution.
$G+_{th1}$ Thumbnails at the lowest resolution.

Google+ in the current beta version uses Picasa as its image repository. As a consequence, Google+ albums also include images previously uploaded to Picasa by a user with the same credentials. Picasa is an online photo-sharing service and has more options, with respect to images, than the OSNs previously analyzed. Picasa allows users to upload images in four ways according to their pixel resolution:

Pic_{hi} User supplied images uploaded without any processing and thus published in their original form.
Pic_{st} User supplied images with the suggested resolution useful to be printed or to be used as screensaver.
Pic_{me} User supplied images with medium resolution, best suited for fast download and sharing.
Pic_{lo} User supplied images with the lowest resolution to be included in blogs or web pages.

Finally, the methodology described in the Introduction was applied, by uploading and downloading images in the input data set for each OSN image type.

III. Image Format Analysis

Many image formats are available as a container of digital images with different characteristics. All the images published by Facebook and Badoo are only in the JPEG format, while

Google+ stores uploaded images in different formats such as JPEG, PNG, GIF and BMP, depending on the input image. Moreover, Picasa does not convert the input images. However due to its features, particularly the compression performance, JPEG is the most widely used file format to store digital images. Nevertheless, the three OSNs accept also images in other formats such as PNG, BMP and GIF. The test results show that there are also unaccepted formats such as TIFF. If the input image satisfies size constraints of the OSN, then the image is either published without modifying its encoding or is converted into another format preserving the pixel resolution. Otherwise, the OSN reduces the size of the image according to its policies and user supplied options using scaling operations.

A series of experiments were run on input images which were not scaled by the OSN. In order to have a more detailed understanding of the conversion process adopted by the OSNs, the same input images were converted using GIMP2 and IrfanView. These images have been compared to the ones downloaded from the OSN. The experimental results on Facebook showed that images converted (from GIF and PNG to JPEG) are identical to the ones converted using the GIMP2 tool. On the contrary, the conversion performed by IrfanView slightly differs. More precisely, for each RGB channel, corresponding pixel values differ by at most 1, that is, if one has value x then the other has value $x-1$, x, or $x+1$. In particular, the percentage of different values is 14,61% for the red channel, 17,46% for green and 17,35% for blue in case of PNG to JPEG conversion, while in the case of GIF to JPEG conversion, the percentages are 15,24%, 17,03% and 16,58%, respectively. On the other hand, the same experimental results on Badoo show that the conversion is different from the one performed by both GIMP2 and IrfanView. The results are reported in Table I for GIMP2 and in Table II for IrfanView. As for Facebook, for each RGB channel the values differ by at most 1.

Table I
DIFFERENCES BETWEEN IMAGES CONVERTED BY BADOO AND GIMP2.

format	Red	Green	Blu
PNG	23.17%	24.59%	26.93%
GIF	25.94%	27.10%	29.16%
BMP	12.96%	13.37%	14.36%

Table II
DIFFERENCES BETWEEN IMAGES CONVERTED BY BADOO AND IRFANVIEW.

format	Red	Green	Blu
PNG	10,20%	10,81%	13,27%
GIF	10.16%	11.37%	13.46%
BMP	5.19%	5.52%	6.47%

Google+ accepts and publish images in different formats. If the uploaded image, with a format JPEG, PNG, GIF, or BMP, has a resolution less than 2048 on the longest side, then the image is published as it is. Otherwise, the image is scaled down (see Section V) and the format is eventually converted to JPEG or PNG. In particular, images in the JPEG and PNG format are not converted, but images in BMP are converted to JPEG and GIF to PNG.

A consequence of the analysis is that OSNs do not add watermarks to the encoded data section of some images. In fact, whenever the image is published as it is or the format conversion process can be replicated obtaining the same output image, then no watermark is added by the OSN. This is the case of Google+ for images with a resolution smaller than 2048 on the long side. This is also true for images that are uploaded to Facebook in GIF or PNG and then converted to JPEG, since the OSN removes the entire EXIF section (see Subsection IV-A).

Badoo gives users the possibility to add a visible watermark to the published images. This is useful when preventing the unauthorized use of images stored on Badoo and makes it impossible for anyone to copy a photo and upload it back onto the Badoo web site [8]. Users can add the watermark by setting the appropriate option in the privacy section of their profile. This watermark consists of a strip located at the bottom of the image with the Badoo logo and the URL of the Badoo home page of the user who published the image. Its forensic value is clearly evident.

A. JPEG Quantization Tables

The JPEG standard defines a well known lossy compression algorithm. One of its most interesting features is the variable compression ratio. Specifically, it gives the user the possibility to choose the compression factor (namely the Quality Factor or QF) thus, optimizing the ratio quality/space. A detailed description of the JPEG format can be found in many text books. The encoding process is based on the Discrete Cosine Transform (DCT) of 8×8 pixels image blocks. The resulting DCT coefficients are then quantized by dividing each coefficient value by its corresponding entry in a predetermined quantization table (QT). Then, the resulting values are rounded to the nearest integer. Finally, the quantized DCT coefficients are ordered and losslessly encoded. The JPEG decompression process, first retrieves the quantized coefficients by a lossless decoding, than the DCT coefficient values are dequantized by multiplying the retrieved DCT coefficient values by their corresponding entries in the QT. For this reason, the 8×8 QT is part of the JPEG structure and is stored in a dedicated section of JPEG files.

For color images, this process is performed for both luminance and chrominance layers on distinct QTs resulting in two matrices, called luminance QT and chrominance QT. The JPEG standard suggests two standard QTs for the luminance and chrominance defined by the Independent JPEG Group in [9]. The Group also established a method that, given a Quality Factor (QF), computes a new QT matrix whose entries $C_{i,j}^*$ are computed from the corresponding entries $C_{i,j}$ of the

suggested matrices as follows:

$$C_{i,j}^* = \begin{cases} \left\lfloor \frac{C_{i,j} \cdot 5000/QF + 50}{100} \right\rfloor, & \text{if } 1 \leq QF < 50 \\ \\ \left\lfloor \frac{C_{i,j} \cdot (200 - QF \cdot 2) + 50}{100} \right\rfloor, & \text{if } 50 \leq QF \leq 99. \end{cases} \quad (1)$$

The value QF is a measure of the compression ratio as well as the perceived quality of the image. Higher values of QF correspond to a smaller compression ratio and better quality while lower values imply smaller file size and greater loss of image details. Using the aforementioned methodology, a set of images saved with different QFs ranging from 30 to 99, were uploaded to each OSN. Afterward, the images were downloaded to evaluate how the two QTs have been changed during the publication process.

The results of these experiments were very encouraging due to them showing that all the OSNs whenever modifying the input images, for example converting the input images to JPEG, compress them using fixed QTs. These QTs, that can be found in the resulting image encodings, are the same as those that can be derived by using 1 with fixed values of QF. These values are listed in Table III. For example, all the user-uploaded images published by Facebook have QTs corresponding to QFs=85. The luminance QT corresponding to QF=85, is reported in Table IV. The entries "No Mod" in Table III mean that the images of that type are not modified by the OSN, i.e. that the image is published without modifying its encoding.

Table III
QF VALUES FOR OSNs AND IMAGE TYPES

OSN	Image Type	QF_Lum	QF_Chrom
Facebook	FB_{hi}	85	85
	FB_{st}	85	85
	FB_{th}	95	95
Badoo	BD_{st}	91	91
	BD_{th1}	97	97
	$BD_{th2}, BD_{th3}, BD_{th4}$	94	94
Google+	$G+_{st}$	No Mod	No Mod
	$G+_{th1}$	81,45	88,78
Picasa	Pic_{hi}	No Mod	No Mod
	Pic_{me}	78,58	88,60
	Pic_{st}	81,45	88,78
	Pic_{lo}	78,58	88,60

Table IV
LUMINANCE QT CORRESPONDING TO QF=85

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

IV. METADATA ANALYSIS

When referring to digital images, metadata can be considered both internal and external to the file containing an image. Internal metadata are usually contained in the EXIF tags defined in [10] for both for digital image formats and audio files. External metadata are represented by the name of the file that OSNs use to store images on their technological infrastructure, or at least, the file name resulting after the download of an image from a given OSN.

A. EXIF extensions

After the publication of an image by an OSN, it is possible to notice that almost always the file size of the resulting image is less or equal to the image prior to being uploaded on the OSN. This is due to both the JPEG compression as well as the deletion of any EXIF metadata on the image

Facebook, during the JPEG compression process, applies some fixed parameters on the processed image, such as "Baseline DCT", "Huffman coding", fixed QTs (see Subsection III-A) and 24 bit encoding for each RGB pixel. Moreover, it uses a specific sub-sampling operation on the original size (YCbCr 4:2:0 (2 2)) and adds to the metadata information needed for the rendering of the image, the ICC Profile [11], to the FB_{hi} and FB_{st} image types. Even though Facebook removes the EXIF metadata, it is important to note that its thumbnail images FB_{th} have the EXIF field "Comment" set to the string "CREATOR: gd-jpeg v1.0 (using IJG JPEG v62), quality = 95". The GD Graphics Library is a freeware and open source graphic library often used for the creation of dinamically rich content on the Web.

Google+ manages metadata in a different way depending on the resolution of the image involved. If the image to be uploaded has a resolution of more than 2048 pixels (on the long side), then a resize operation is performed, and the EXIF metadata are removed. In this case, metadata associated to the image can be only seen on the OSN and are not present in the image when downloaded from the OSN. On the contrary, if the long side of the image has a resolution of less than 2048 pixels, then the image is not modified at all, including the original EXIF metadata which will be left untouched.

Badoo decreases the size of an image published thanks to the JPEG compression and the EXIF metadata deletion. During JPEG compression, Badoo uses the same fixed parameters of Facebook, with the only difference being the QFs value. Differently from Facebook, Badoo does not add information on the ICC Profile to be used. All the EXIF metadata are removed but one. In fact, a new EXIF "Comment" field is added to store important information regarding the user who uploaded the image. In details, the EXIF "Comment" is an hexadecimal string that can be decomposed into several parts. Here is an example:

```
zU0 9393951E2866 0D0D0FD5D5D5 9292941E286D 783F2CD7D7D7
8D8C911D2765 131416CECDC9 828284946750 121315C1C2C4
0E3758F0 0000C400
```

It is composed of $3 + 112$ characters divided as follows: the first 3 characters can be seen as a signature and are always zU0. The following 96 characters represent an internal color representation of the image and are grouped into 8 sections of 12-characters. The last characters can be

divided into 2 sections of 8-characters: "*0E3758F0*" that is the hex representation of the user identification number "*0238508272*", and "*0000C400*" i.e. the hex representation of the number of the image, "*50176*". Since it has been noted that Badoo uses images of at most 920 pixel wide, the resulting original file name is "*50176_920.jpg*" (where 920 is the Badoo standard resolution). Recalling that Badoo publishes an image by using the following scheme, *http://badoo.com/[user_identification/p[photo_number]*, the resulting URL of the image from the previous example will be *http://badoo.com/0238508272/p50176*.

B. Image File Names

In this subsection, the file names of the images coming from an OSN are analyzed. Some file names give interesting information which is useful for a Digital Forensics analysis. When a user asks to download an image from an OSN, the browser prompts as the default file name to save the image on the user computer the same name used for that image on the OSN. Therefore, the analysis took into account also the policy used by each OSN to assign a name to images after the user upload.

Facebook and Badoo give no possibility to choose the image file name or to keep the source file name while Google+/Picasa leave/keep the source file name and, when a user download an image from these OSNs, is prompted with the source file name as the default name to save that image. Facebook assigns an unique "Image IDentifier" to each image along with the "Album IDentifier" and the "User IDentifier". Therefore, it is possible to download the image *2387802023252* in the album *a.2387759502189.141733* of the user *1496850761* by setting the fields *fbid* and *set* in the URL obtaining the following HTTP query string:

```
http://www.facebook.com/photo.php?fbid=2387802023252\
    &set=a.2387759502189.141733.1496850761\&type=1
```

Whereas this image will be downloaded using the download action, the following file name `314951_2387802023252_1496850761_2715967_378561091_n.jpg` will be proposed to the user having in the file name the "Image ID" and the "User ID".

Badoo does not explicitly allow the download of images. To perform the same test of Facebook, the HTML source code was analyzed, extracting the URL in the tag "SRC" of the image under study. For example, visiting the profile of user *172329121*, the image *t1285234502/813151_300.jpg* will be displayed using the following URL:

```
http://77.67.26.43/167/3/8/8/172329121/696068/
    t1285234502/813151_300.jpg
```

where 77.67.26.43 is the IP address of the host *p34.badoo.com*. In the same directory, the thumbnails with a lower resolution are also stored. As a consequence, the image downloaded from this URL will have a height of 300 pixels, while the BD_{th1}, BD_{th2} and BD_{th3} versions can be downloaded using as file names 813151_48.jpg, 813151_96.jpg, and 813151_192.jpg respectively, at the end of the previous URL.

As previously stated, Google+ and Picasa do not change the source file name and therefore the images downloaded from this OSN are not discussed.

V. PIXEL RESOLUTION ANALYSIS

The pixel resolution of an image is usually described with a pair of two positive integer numbers, where the first number is the number of pixel columns and the second is the number of pixel rows. This is one of the indicators of the appearance quality of the image as well as its size. The larger the numbers, the better the quality and the greater its size. An OSN with many published images can be interested in limiting the size of images to save on bandwidth and total storage needed. Therefore, upper bounds on the pixel resolution are established and any image with a greater resolution is converted to the upper bound resolution. The upper bound is big enough to allow a good appearence quality while saving on storage. The upper bound varies according to some image classification: for example, thumbnails have a smaller pixel resolution than other images. Fixing upper bounds $UP_N \times UP_M$ on both the number of pixel rows $N \leq UP_N$ and the number of pixel columns $M \leq UP_M$ implies that an image with greater values $N \times M$ has to be resized, while images with pixel resolution smaller than the upper bound are published without resizing. Rescaling the original resolution to the bound resolution $UP_N \times UP_M$ causes distortion in the picture if the resolution ratios N/M and UP_N/UP_M are different. However, if the bound does not preserve the resolution ratio $N/M = UP_N/UP_M$ and distortion should be avoided, then the resizing process has two possibilities:

- The image is resized by a factor which is the maximum value among the ratios N/UP_N and M/UP_M. Namely, the resized image has pixel expansion $N/\alpha \times M/\alpha$ where $\alpha = \max\{N/UP_N, M/UP_M\}$.
- Another possibility would be cropping the image.

The first approach is followed by the OSNs analyzed in this paper for almost all images. While the latter approach is used for thumbnails and profile images on Facebook. Facebook lets the user upload images with two options on the resolution: standard FB_{st} or high FB_{hi}. The published image will be at most 720×720 in the standard resolution and 2048×2048 in the high resolution. These are upper bounds: the values of pixel resolution $N \times M$ have to be $N \leq 720$ and $M \leq 720$ in the standard case ($N \leq 2048$ and $M \leq 2048$ in the high case). Images with pixel resolution values $N \times M$ both smaller than the limit, are published without resizing. Otherwise, the image is resized by a factor of $\alpha = \max\{N/720, M/720\}$ for the standard or $\alpha = \max\{N/2048, M/2048\}$ for the high case. The resulting image will have its greater dimension equal to 720 pixel in the standard case or 2048 pixel in the high resolution case. Each Facebook user can upload his profile picture FB_{pr}. To be accepted each image has to be at most 4MB and with a pixel resolution of at least 180 pixel as the number of columns. Facebook thumbnail images FB_{th} have pixel expansion 50×50. They are derived from the FB_{pr}

uploaded by users, i.e., the $N \times M$ user image is resized and cropped near the center, to get a 50×50 image.

Images published by Badoo (BD_{st}) have a pixel resolution of at most 920×920. This is an upper bound: the pixel resolution values $N \times M$ have to be $N \leq 920$ and $M \leq 920$. The resizing of images with a greater resolution is done analogously to Facebook. There is also a constrain on the allowed images: users can upload images with a pixel resolution $N \geq 200$ and $M \geq 200$, otherwise they are not accepted by Badoo for publishing. Badoo associates thumbnails to published images. Thumbnails can have different pixel resolutions constraints: $M = 48$ (BD_{th1}) , $M = 96$ (BD_{th2}), $M = 192$ (BD_{th3}), and $M = 300$ (BD_{th4}), with no limitation on N. According to the four different cases, the image is resized preserving the apect ratio.

Images published by Google+ have a pixel resolution of at most 2048×2048 ($G+_{st}$). As for the previous cases, this is an upper bound: the values of pixel resolution $N \times M$ have to be $N \leq 2048$ and $M \leq 2048$. Images with pixel resolution values $N \times M$ smaller than the limit are published without resizing or any processing. The experiments show that the MD5 and SHA-1 values of the input user image and those of the downloaded image are the same. The resizing of images with a greater resolution is done similarly to Facebook and Badoo.

Picasa allows users to upload images in four ways, according to their pixel resolution. The "original dimension" manages the image at the original image resolution (for example 4288×2848). The "recommended" (1600 pixels) resolution is mostly used for prints, sharing online albums, or for use as a screensaver. The third resolution, "medium (1024 pixels) is the one preferred for sharing online albums with friends and family. The last resolution, "small" (640 pixels) is mainly used for publishing images on blogs and web pages.

VI. OSN Image Forensics

The science of Image Forensics has been widely discussed in many scientific publications and provides the forensics analysts several tools and methodologies to investigate different ways an image can be involved in a digital investigation. Traditionally, previous studies and findings investigate how an image has been created, modified, or "used" during a crime or an illegal act. The analysis presented in this paper gives some advice on the image "fingerprints", useful in reconstructing social networking activities. This data, could be correlated to further information coming from other sources of evidence to allow investigators to reconstruct illegal or illicit activities on some of the most common OSNs. The analysis presented so far can be relevant in an Image Forensic analysis to establish whether an image has been uploaded to a particular OSN and then published or not.

The analysis was carried out on images published on the three considered OSNs. Clearly, if a user has modified the image then some forensically relevant fingerprints may disappear. However, there are many forensic techniques which are capable of detecting a variety of standard image manipulations. Recompressing an image which has previously been JPEG compressed, also known as "double JPEG" compression, can be detected [12], [13] and the quantization table used during the initial application of JPEG compression can be estimated. If anti-forensic techniques, such as those in [14], are used then forensically significant compression fingerprints (i.e., forensically detectable fingerprints) are removed from the image and the mentioned techniques will not work.

It is worth pointing out that Facebook and Google+ do not add watermarks to the encoded data section of different kinds of image. Badoo gives users the possibility to add a visible watermark to the published images.

VII. Conclusions

The paper has analyzed the characteristics of images published on some OSNs. The analysis has mainly focused on how the OSN processes the uploaded images and what changes are made to some characteristics, such as JPEG quantization table, pixel resolution and related metadata. The experimental analysis was carried out in June-July 2011 on Facebook, Badoo and Google+. Due to the rapidly approaching changes, the experimental analysis presented in this paper should be updated following the OSN changes in the publication process. It could be interesting to repeat the authors analysis when Google+ is released (it is still in a beta version) and Facebook replies.

References

[1] Facebook, August 2011. [Online]. Available: http://www.facebook.com/press/info.php?statistics

[2] T. P. Blog. (Jun 2010) Exploring the software behind facebook, the world's largest site. [Online]. Available: http://royal.pingdom.com/2010/06/18/the-software-behind-facebook/

[3] Badoo. (Jul 2011). [Online]. Available: http://corp.badoo.com/company

[4] G. Inc. (2011, July) Social network user statistics as of july 2011. [Online]. Available: http://google-plus.com/598/social-network-user-statistics-as-of-july-2011/

[5] U. F. Petrillo, A. Castiglione, G. Cattaneo, and M. Cembalo, "Experimentations with source camera identification and online social networks," *J. Ambient Intelligence and Humanized Computing*, vol. http://dx.doi.org/10.1007/s12652-011-0070-2, 2011.

[6] J. Lukás, J. J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, 2006.

[7] (Jul 2011). [Online]. Available: http://trends.google.com/websites?q=badoo.com

[8] (2011, Jul) Badoo help page. [Online]. Available: http://badoo.com/help

[9] I. J. Group. (2011, Jan). [Online]. Available: http://www.ijg.org/

[10] Joint Photographic Experts Group, "Jpeg standards: Iso/iec is 10918-1, itu-t recommendation t.81," http://www.jpeg.org/jpeg/index.html, 2004.

[11] International Color Consortium, "ICC specifications," http://www.color.org/icc_specs2.xalter, Feb 2010.

[12] A. C. Popescu and H. Farid, "Statistical tools for digital forensics," in *Information Hiding*, ser. Lecture Notes in Computer Science, J. J. Fridrich, Ed., vol. 3200. Springer, 2004, pp. 128–147.

[13] T. Pevný and J. J. Fridrich, "Detection of double-compression in jpeg images for applications in steganography," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 2, pp. 247–258, 2008.

[14] M. Stamm and K. Liu, "Anti-forensics of digital image compression," *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 3, pp. 1050 –1065, 2011.