Forensics Investigation on Image Files using Quantization Value

Sabam Parjuangan[1], Muhammad Hafif[2], Suhardi[3]
[123]School of Electro Engineering and Informatics
Bandung Institute of Technology (ITB)
GaneshaStreet No. 10, Bandung-West Java, Indonesia
[1]33220021@std.stei.itb.ac.id, [2]23219353@office.itb.ac.id, [3]suhardi@stei.itb.ac.id

Abstract: Image files were one of the evidencesas a result of the forensics investigation on cybercrimes.The process of acquiring the evidence must have a high accuracy. The image files were obtained from social media that had incomplete metadata. This article outlined an in-depth approach for forensics investigation on image files. The method of this forensics examination conducted to the image files was through the investigative action procedures so that the incomplete metadata can still be analyzed. The data analysis technique used in this study was JPEG Analysis, ELA, and Noise methods. The result of this study was that the characteristics of the image files were able to be identified from social media through editting, and originating processes. The forensics measures described in this study were also able to prove the image filemodifications.

Keywords: Image Forensics, Cybercrimes, Quantization

1. INTRODUCTION

One of the pieces of evidence used in explaining an incident was an image file. Image files had a great opportunity to provide a complete chronology of events. The use of image files in various instances often occurred. Both events occurred in the real world and cyberspace so that the handling of image files were used as supporting material in explaining an event needs special attention. Image files that had multiple viewpoints were also modified in such a way as to provide misinformation about an event. Therefore, the image file must be able to review the events that occurred in it and the information contained therein. Image files were files that contained a description of an event. This file had several types of formats, such as JPG, JPEG, TIF, BMP, and PNG. These formats fell into the category of digital image files[1].

The image file processing or incident tracings was known as forensics. Forensics action had very broad meaning, especially for the forensics investigations in the digital field e.g., forensics investigation of mobile, network, hardware (hard disk/memory / other storage media), software, email, social media, text documents, video, sound, and image files. This article took an in-depth look at the forensics investigation of image files.

Image files were not separated from the environment e.g., the file storage media itself and the image file transmission media. Image files had several structures e.g., a collection of pixels arranged in a certain way and showing a description of an object/event. The smallest unit of an image file was described in pixels. Image files had two types e.g., pixel, and vector image files. Pixel image files usually in the form of in PNG, JPEG, TIFF, GIF, and BMP formats. Image file pixels were formed as a picture was taken and adjusted to the specifications of the device used in creating the image file. Meanwhile, vector type images were usually images used for animation. The types of images were AI, CDR, SVG, and EPS formats[1].

Based on the type of pixel image file, there were 4 types of images in this section, e.g., binary image files, grayscale image files, color image files, and multispectral image files. A binary image file was an image file consisting of two-bit dimensions. The color consisted of only two colors, e.g., black and white. The black color representation was 1 and the white representation was 0. A gray-scale image file was an image file that had pixels represented by two binary numbers. The black color representation for was 11, and that of white was 00. The color image file was in the form of RGB image file. This image file had various colors consisting of red, green, and blue. As it was the same as the multispectral image files, the each layer of the image files contained RGB.

Several previous studies that had been carried out for forensics investigations on image files revealed that forensics investigations on image files were an investigative procedure for gathering evidence of an incident using special tools and mathematical equations[2]. This investigation was equivalent to physical evidence presented in the form of a report that was able to be used as evidence in court. This research described everything from visual time analysis of software accessed through a hardware device, checking image files, and extracting the image files[1].

Other studies that described forensics investigations on image files were describing the process of identifying data compression events, localizing, and repairing image damage, image file header analysis, reconstruction of image file fragments, identification, and image viewing devices. The stages of the forensics investigation in this study showed that the image file was determined whether there was a file compression or not, and if there was a compression, the color changed, and the impact on the image file rose [3].On the other hand, the research that investigated image files forensically used resizing JPEG coefficient detection methods to detect compression or not and to analyze phase suitability to detect image splicing[4].

The entire method used in previous research was carried out to ensure that there was a compression event, there was an addition of an image, there was a change in the image file header, to ensure the device used in creating the image file. Existing forensics actions were focused on original evidence, which came from the main device, and even cryptographic hashes were carried out so thatthe image size did not change even though the transferring occurred. The novelty of this research was to carry out forensics investigations on image files that had been transmitted on social media platforms. Namely the social media platforms WhatsApp and Facebook. These two platforms became the focus because these two social media platforms were the most commonly used by Indonesians. Facebook users in Indonesia reached 82% of the social media platform, and WhatsApp users in Indonesia also reached 84% of Indonesian population – 272.1 million. In addition, several crime cases often involved these two social media[5].

This article describes the forensic action on image files sourced from social media Facebook and the WhatsApp application. The forensic investigation described in this article is the performance of using image quantization methods in conducting forensic investigations. This is important because the current level of cybercrime often involves image files sourced from social media. It is not uncommon for perpetrators or victims of social media crimes to use images on social media as evidence of crime. So that handling problems also involves image files sourced from social media. Therefore, a comprehensive handling of the evidence is required.

The structure of this article was the introduction explaining the background of this research and the motivation and the previous studies describing the forensics investigations on image files. The second part of this study was the method used in this research. The third part was the result and discussion of this study. The last part of this study was a part of the conclusions of this study

2. METHOD

The method used in this research is the JPEG Analysis method in which there are quantization and analysis of adjacent pixels with eight (8) neighbors. JPEG analysis is carried out on image pixels. The pixel adjunct operation with 8 neighbors uses an equation like in Figure 1 below and equation 1-6 [6].

T4	T3	T2
T5	P	T1
T6	T7	T8

Figure 1: Pixel Adjunct Operation -8 on Image Pixels

The caption in figure 1 where P: Coordinate point (b, K) and the equation derived from Figure 1 above to determine the value of each pixel as in the following equation.

$$T1: (b, K + 1) \tag{1}$$

$$T1: (b, K + 1) \tag{2}$$

$$T2, T4: (b - 1, K - 1) \tag{3}$$

$$T3, T5: (b, K - 1) \tag{4}$$

$$T6, T7: (b + 1, K - 1) \tag{5}$$

$$T8: (b + 1, K + 1) \tag{6}$$

The other methods used in this study were JPEG analysis. The purpose of using JPEG Analysis was analyzing the color changes in image files that had been compressed using algorithms on social media platforms. A kind of rising colors included red, green, and blue (RGB) in image files that had been transmitted through social media platforms. The contribution of this method in forensics investigations was to examine the transmitted image files that still had the same image description or experience different descriptions.

Another method used was the Error Level Analysis (ELA) method. The purpose of ELA was finding RGB color mixing in the image file display. The concept was that several things happened to the image file on condition that the image file was compressed. Image connection, image falsification, and image retouching occurred [8].Therefore, ELA was done to analyze the extent of the compression impact on an image file.

Another method used was noise analysis. Noise analysis was used to detect random variations in the color information in an image. The random variations that occurred were caused by several things including file compression, joining files, and also retouching. Some types of noise in images were photoelectronic, impulse, and noise in the image file structure. The complete method used in this research was shown in Figure 2 below.

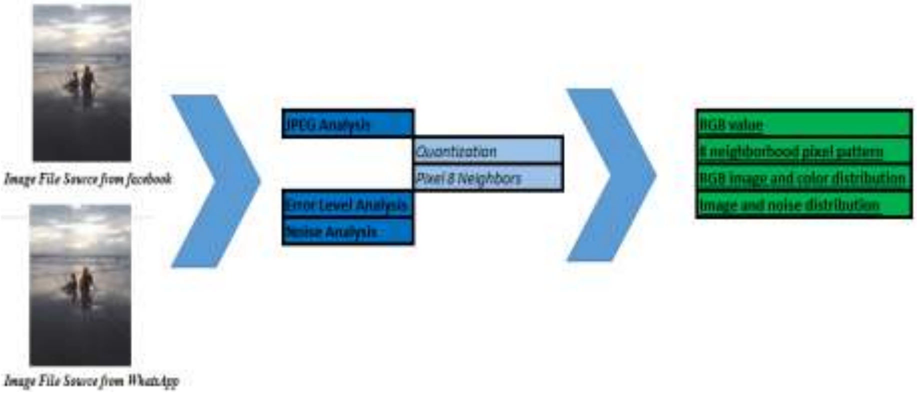


Image File Source from facebook
Image File Source from WhatsApp
JPEG Analysis
Error Level Analysis
Noise Analysis
Quantization
Pixel 8 Neighbors
RGB value
8 neighborbood pixel pattern
RGB image and color distribution
Image and noise distribution

Figure 2: Forensics Investigation Methods on Image File Pixels

3. RESULT AND DISCUSSION

The results of this study were divided into several parts. They were: the results from JPEG Analysis, Adjunct Pixel 8 Analysis, Error Analysis, and Noise Analysis. The results of the JPEG Analysis were shown in Figure 3. JPEG Analysis results showed the distribution of RGB colors in image files from Facebook, WhatsApp, and the original. The color spread was also labeled with the RGB value of the color so that it also referred to the RGB value. The distribution of RGB colors in image files was from the Facebook which was more diverse. This showed that the image file compression on the Facebook was higher than the WhatsApp platform. It was seen from the MAC file of the image file. The MAC image files from the Facebook and the WhatsApp were shown in Figure 4. The difference between the original image file size and image files from the Facebook and the WhatsApp was shown in Table 1. The compression percentage was determined by equation 7 below.

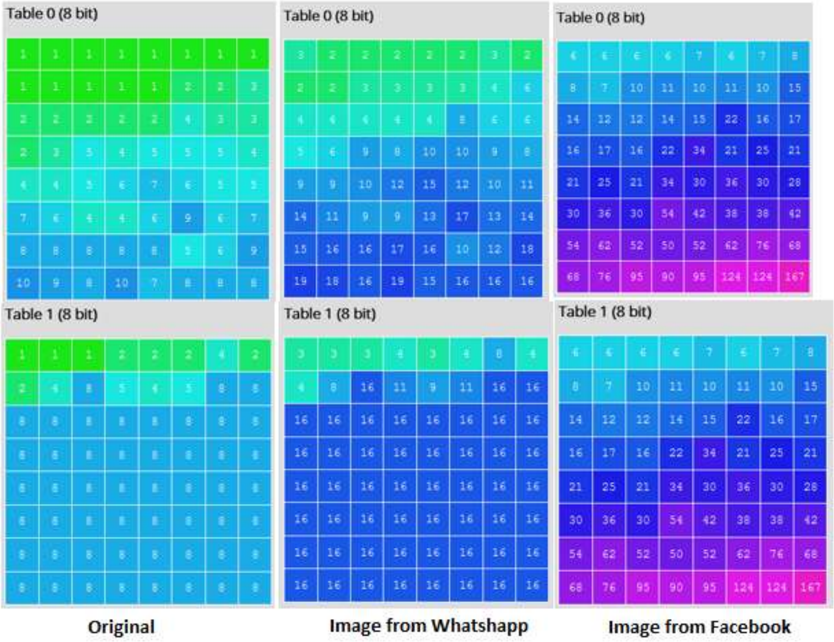
$$Compression(\%) = \frac{\mathrm{Ofs} - \mathrm{Crs}}{\mathrm{Ofs}} x\ 100\% \qquad (7)$$

Where:
Ofs: Original file size
Crs: Compression result size

Table 1: Image File Compression Result on Facebook and WhatsApp

Original (Kb)	**Facebook (Kb)**	**WhatsApp (Kb)**
2222.08	83.5	305
Compression(%)	***96%***	***86%***

The process of compressing an image file basically removed RGB values and changed the pixel density of an image file[9]. The higher the percentage of image file compression was, the higher the density level was removed and the RGB value was also increasingly spreading and mixing with other colors [10]. This process created a new color [11]. It was seen in Figure 3 that the compression results of the image files from the WhatsApp application did not have a new high color level – there was only a fewer color additions than the result of compression on the Facebook social media. Moreover, the image files from the Facebook social media showed more new color additions. Furthermore, it showed 2-time higher added image files from the WhatsApp application. This was in line with the result of the compression analysis on the Facebook social media which was higher than the WhatsApp application.



Original

Table 0 (8 bit)

1	1	1	1	1	1	1	1
1	1	1	1	1	2	2	3
2	2	2	2	2	4	3	3
2	3	5	4	5	5	5	4
4	4	5	6	7	6	5	5
7	6	4	4	6	9	6	7
8	8	8	8	8	5	6	9
10	9	8	10	7	8	8	8

Table 1 (8 bit)

1	1	1	2	2	2	4	2
2	4	8	5	4	5	8	8
8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8

Image from Whatshapp

Table 0 (8 bit)

3	2	2	2	2	2	3	2
2	2	3	3	3	3	4	6
4	4	4	4	4	8	6	6
5	6	9	8	10	10	9	8
9	9	10	12	15	12	10	11
14	11	9	9	13	17	13	14
15	16	16	17	16	10	12	18
19	18	16	19	15	16	16	16

Table 1 (8 bit)

3	3	3	4	3	4	8	4
4	8	16	11	9	11	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16

Image from Facebook

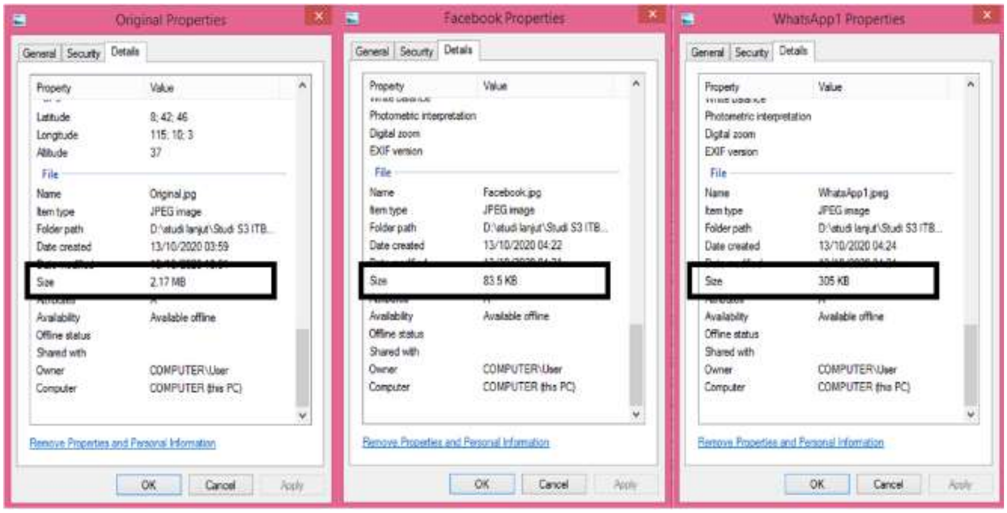
Table 0 (8 bit)

6	6	6	6	7	6	7	8
8	7	10	11	10	11	10	15
14	12	12	14	15	22	16	17
16	17	16	22	34	21	25	21
21	25	21	34	30	36	30	28
30	36	30	54	42	38	38	42
54	62	52	50	52	62	76	68
68	76	95	90	95	124	124	167

Table 1 (8 bit)

6	6	6	6	7	6	7	8
8	7	10	11	10	11	10	15
14	12	12	14	15	22	16	17
16	17	16	22	34	21	25	21
21	25	21	34	30	36	30	28
30	36	30	54	42	38	38	42
54	62	52	50	52	62	76	68
68	76	95	90	95	124	124	167

Figure 3: Quantization Values on Image File Pixels



Original Properties

Property	Value
Latitude	8; 42; 46
Longitude	115; 10; 3
Altitude	37
File	
Name	Original.jpg
Item type	JPEG image
Folder path	D:\studi lanjut\Studi S3 ITB...
Date created	13/10/2020 03:59
Size	2.17 MB
Availability	Available offline
Offline status	
Shared with	
Owner	COMPUTER\User
Computer	COMPUTER (this PC)

Remove Properties and Personal Information

Facebook Properties

Property	Value
Photometric interpretation	
Digital zoom	
EXIF version	
File	
Name	Facebook.jpg
Item type	JPEG image
Folder path	D:\studi lanjut\Studi S3 ITB...
Date created	13/10/2020 04:22
Size	83.5 KB
Availability	Available offline
Offline status	
Shared with	
Owner	COMPUTER\User
Computer	COMPUTER (this PC)

Remove Properties and Personal Information

WhatsApp1 Properties

Property	Value
Photometric interpretation	
Digital zoom	
EXIF version	
File	
Name	WhatsApp1.jpeg
Item type	JPEG image
Folder path	D:\studi lanjut\Studi S3 ITB...
Date created	13/10/2020 04:24
Size	305 KB
Availability	Available offline
Offline status	
Shared with	
Owner	COMPUTER\User
Computer	COMPUTER (this PC)

Remove Properties and Personal Information

Figure 4: Size Comparing Image Files

The further result was seen on the neighbor pixel test result on the image files. The result of this study was listed in Table 2.Table 2 contained the result for each neighbor of a pixel which referred to the quantization value in Figure 3. The 8-pixel neighborliness test was taken by sampling on the quantization value of the image files from WhatsApp. The purpose of this analysis was comparing the changing neighboring pixels in the compressed image files on each platform.

Table 2: Results of the 8 Neighboring Pixel Test

Item	Subscription -1	Subscription -2	Subscription - 3	Subscription - 4	Subscription - 5	Subscription - 6	Subscription - 7	Subscription - 8	Subscription - 9
(b,K)	b=2, K=7	b=5, K=7	b=8, K=7	b=2, K=4	b=5, K=4	b=8, K=4	b=2, K=1	b=5, K=1	b=8, K=1
T1	5.5	7	8.5	4	5.5	7	2.5	4	5.5
T2	7	10	13	4	7	10	1	4	7
T3	3.5	5	6.5	2	3.5	5	0.5	2	3.5
T4	7	10	13	4	7	10	1	4	7
T5	3.5	5	6.5	2	3.5	5	0.5	2	3.5
T6	4.5	6	7.5	3	4.5	6	1.5	3	4.5
T7	4.5	6	7.5	3	4.5	6	1.5	3	4.5
T8	11	14	17	8	11	14	5	8	11

Table 2 showed that this neighbor had a value that had the same pattern for each compression. The purpose of identifying neighboring 8 pixels was to be able to return the pixel value if the file was compressed so that the pixel value was found in the original file. When the original file pixel value was found, an RGB color representation was able to be created. This study was done to identify an image file's content that had been tampered with or not. In addition, it was done to test the similarity of the contents of the image files with the comparison image files by equalizing the pixel size of the image files[12].

The result of the next analysis was the Error Level Analysis (ELA) analysis. The result of this analysis showed that the image files from the Facebook social media and the WhatsApp application had the same error rate. This was shown in Figure 5. Figure 5 showed that the color distribution between the original image files and the image files from the WhatsApp application and Facebook social media was different but had the same color distribution pattern in the image files from the Facebook social media and the WhatsApp application. The ELA analysis was used to identify the edit level of image files. Identification process was changed in objects of the image through the distribution of colors in the results of the ELA analysis[8].

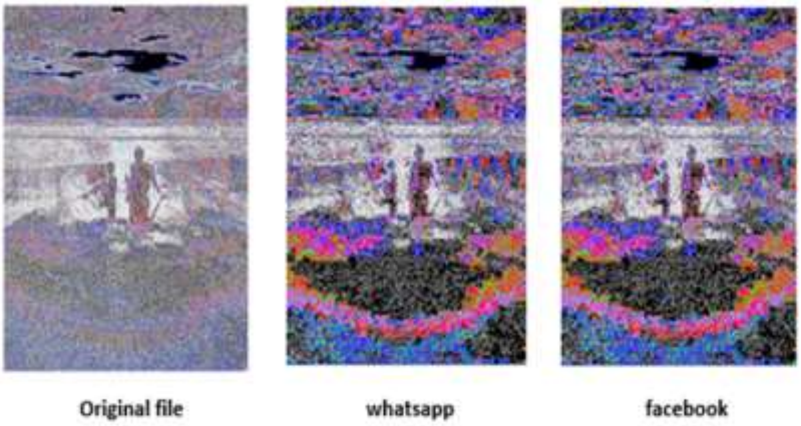


Figure 5: Result of ELA Analysis

The result of the Noise Analysis showed that the noise that occurred in files from the WhatsApp application and Facebook social media had the same noise. The result of this study was shown in

Figure 6. This noise showed that the result of ELA was in line with the results of the Noise Analysis. Therefore, the results of ELA analysis and Noise Analysis were always in line. The purpose of this analysis through ELA and Noise Analysis was identifying the noise in the image files. In other words, the evidence in the form of an image files was identified the noise level in the image files[13].

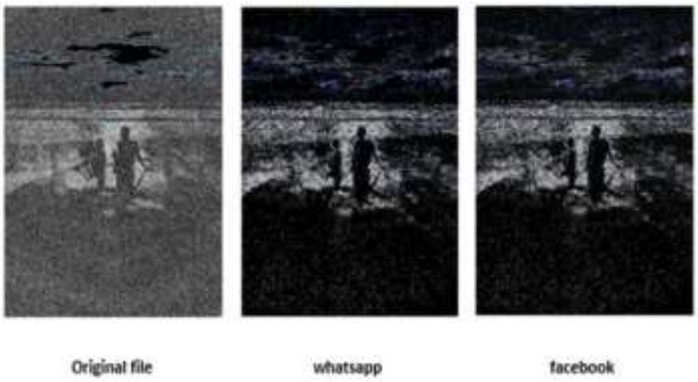


Figure 6: Result of Noise Analysis

4. CONCLUSIONS

Forensics investigation of original image files and that of Facebook and WhatsApp using the quantization method with the stages of JPEG Analysis, Noise Analysis, and Pixel 8 Neighboring is used in investigating image files. Things that are used for its investigation are for finding noise in the image files, the source of the image, the compression level of the image, and the comparison of the accuracy of the image content. In addition, it is also used to analyze the spread of pixel volume in the image so that it can be decrypted on the compressed image files. The compression rate for image files from Facebook social media was 86%; while, the image files from the WhatsApp application are 96%. The accuracy of the results of the analysis in carrying out forensics investigations on the original image files and those sourced from social media Facebook and WhatsApp is 100%.

REFERENCES

[1] R. S. Khalaf and A. Varol, “Digital forensics: Focusing on image forensics,” *7th Int. Symp. Digit. Forensics Secur. ISDFS 2019*, pp. 1–5, 2019.

[2] J. Fridrich, *Digital image forensics*, vol. 26, no. 2. 2009.

[3] M. F. Sabir and J. H. Jones, “A Non-Algorithmic Forensic Approach for Hiding Data in Image Files,” pp. 60–64, 2018.

[4] T. Gloe, M. Kirchner, and R. Böhme, “Can We Trust Digital Image Forensics ?,” pp. 78–86, 2007.

[5] A. L. L. T. H. E. Data *et al.*, “DIGITAL 2020,” 2020.

[6] S. Kim and H. J. Kim, “JPEG encryption with file format preservation and file size reduction,” *2019 IEEE 8th Glob. Conf. Consum. Electron. GCCE 2019*, pp. 215–216, 2019.

[7] C. M. Lewandowski, *Computer Forensics Investigating Data & Image Files*, vol. 1. 2015.

[8] N. B. A. Warif, M. Y. I. Idris, A. W. A. Wahab, and R. Salleh, “An evaluation of Error Level Analysis in image forensics,” *Proc. - 2015 IEEE Int. Conf. Syst. Eng. Technol. ICSET 2015*, pp. 23–28, 2016.

[9] M. Hernandez-Cabronero, V. Sanchez, I. Blanes, F. Auli-Llinas, M. W. Marcellin, and J. Serra-Sagrista, “Mosaic-based color-transform optimization for lossy and lossy-to-lossless compression of pathology whole-slide images,” *IEEE Trans. Med. Imaging*, vol. 38, no. 1,

pp. 21–32, 2019.

[10] C. Cai, L. Chen, X. Zhang, and Z. Gao, “Efficient Variable Rate Image Compression With Multi-Scale Decomposition Network,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 12, pp. 3687–3700, 2019.

[11] C. Cai, L. Chen, X. Zhang, and Z. Gao, “End-to-End Optimized ROI Image Compression,” *IEEE Trans. Image Process.*, vol. 29, no. 8, pp. 3442–3457, 2020.

[12] T. Zhang and R. Wang, “Doctored JPEG image detection based on double compression features analysis,” *2009 Second ISECS Int. Colloq. Comput. Commun. Control. Manag. CCCM 2009*, vol. 2, no. c, pp. 76–80, 2009.

[13] S. J. Cha, U. Kang, and E. J. Choi, “The image forensics analysis of jpeg image manipulation (lightning talk),” *Proc. - 2018 4th Int. Conf. Softw. Secur. Assur. ICSSA 2018*, pp. 82–85, 2018.