2012 Fourth International Symposium on Information Science and Engineering

An Anti-Forensic Algorithm of JPEG Double Compression Based Forgery Detection

Feng Chunhui[1,a], Xu Zhengquan[2,b] and Zheng Xinghui[3,c]
[1,2,3]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing
Wuhan University
Wuhan, China
[a]feng@whu.edu.cn, [b]xuzq@whu.edu.cn, [c]eric@whu.edu.cn

***Abstract*— Nowadays, a number of forensic algorithms have been developed to detect the manipulations of digital images. However, little consideration has been given to anti-forensic schemes capable of fooling the forensic methods. Aiming at this situation, a novel anti-forensic algorithm is proposed to counter a forensic analysis that can detect double JPEG compression with the same quantization matrix; and the limitation of existing image forensics are illustrated accordingly. The proposed algorithm allows generating tampered images without overall recompression, therefore eliminates the double quantization artifacts which will generally exist in tampered JPEG images. Experiment results show that the tampered images generated by the presented method not only have a good disguise effect, but can also effectively counter the forgery detection method that is based on the detection of double JPEG compression with the same quantization matrix.**

Keywords- image forensics; anti-forensics; double quantization detection

I. INTRODUCTION

Image forensics and anti-forensics are two aspects that compete against each other. Existing image forensic methods have developed into wide variety of categories and reached a certain depth, while image anti-forensics is still in its infancy. However, image forensics and anti-forensics have a mutually reinforcing relationship. If one side is weak, the other will stagnate or develop blindly.

Forensic methods can be divided into two categories: the active forensics and the passive forensics [1]. Representative of active forensics is digital watermarking. However, as the watermark has to be inserted at the same time the image is recorded, it can only be applied in specially equipped digital cameras. In contrast, passive forensics only takes the statistic alternation left by the forgery activity as the evidence in the forensic process. Passive forensic methods mainly include pixel-based detection, format-based detection, physically based detection and geometric-based detection, et al [2].

Because of the wide application of JPEG compression standard, many forensic methods based on JPEG-format arose in the last ten years. Lukáš and Fridrich [3] and Popescu [4] pointed out that, if an image is doubly JPEG compressed with different quantization matrices, the histograms of the DCT coefficients will display certain periodical patterns, which is the so-called DQ effect. Lukáš proposed three approaches to estimate the primary quantization matrix of the tested JPEG image by using such patterns. Lin [5] pointed out that, a DCT histogram of a tampered JPEG image can be considered as the sum of two histograms, one has the special patterns and the other has a random distribution. Through analyzing the various patterns of the DCT histograms, Lin could determine whether the test image was a tampered one and track down the manipulated area.

However, most JPEG-format-based forensic methods depend on the distinguishing patterns resulted in double compression with different quantization matrices; if an image is recompressed using the same quantization matrix, there will not be any visible changes in the DCT histograms. In the year 2010, Huang [6] pointed out that, if an image is compressed repeatedly using the same quantization matrix, the number of the different DCT coefficients between two adjacently compressed JPEG images would gradually decrease. Via a perturbation strategy, the method in [6] could verify whether a test image was recompressed using the same quantization matrix.

For brevity, in the rest of this paper, double compression with different matrices is simplified as DCD; double compression with the same quantization matrix is simplified as DCS. Changes caused by DCD and DCS, being different though, are both referred to as DQ effect. In addition, the DCS detection is specifically referred to the forensic method proposed in [6], as well as any applications of the method.

Compared to image forensics, methods of anti-forensics reported in literature are rare. Kirchner [7] proposed a method that allowed resizing and rotating bitmap images without leaving a periodic pattern, which was an evidence of manipulation. Gloe [8] presented two counter forensic techniques that could perform resampling operations undetectably and forge traces of image origin. Kirchner [9] proposed a way to synthesize color filter array (CFA) pattern to conceal traces of manipulation from forensic techniques that analyze the CFA structure of images. Cao [10] proposed new variants of contrast enhancement operators which are undetectable by the existing contrast enhancement detectors.

As for the anti-forensic schemes based on JPEG format, Stamm [11] proposed a method that could remove the DQ effect caused by DCD by blending noise signal into the DCT

978-0-7695-4951-4/13 $26.00 © 2013 IEEE
DOI 10.1109/ISISE.2012.41

coefficients. However, as the method needed to distribute noise to the whole image, it would seriously degrade the image quality [12]. Moreover, Lai [13] proposed a forensic method which could detect this noise-adding action. To the best of our knowledge, no method that can be robust to the DCS detection has been reported in the literature.

Essentially, the DCS detection depends on the statistical changes of the DCT coefficients introduced by recompression. If the recompression process could be avoided, the forensic scheme would lose its efficacy. Based on this idea, we propose an anti-forensic method that can avoid the recompression process while manipulating a JPEG image.

What we have to clarify here is that, the method proposed in [6] is inherently not to detect specific image forgery, but double compression in general. However, as the authors of [6] claimed, one of the applications of their method was to detect tampered image generated by a smart forgery maker, who composites an image with DCS. Therefore, the main contribution of our proposed method is limited to counter such application of [6].

The reminder of this paper is organized as follows: Section II recalls the details of the novel JEPG-format based forensic scheme (the DCS detection) before the principle and specific realization steps of our corresponding anti-forensic method are presented in section III. The experimental results and discussions are provided in Section IV, and the conclusion is drawn in Section V.

II. DCS Detection and Its Limitation

Generally, disguising a JPEG image requires a double lossy JPEG compression which will introduce statistical changes of the DCT coefficients, and these changes can be utilized by forensic analysis. This section briefly reviews a novel forensic scheme by means of detecting such changes and analyzes its limitation.

Sometimes, experienced image forgers will use DCS to generate the manipulated image so that the DCT histograms will exhibit no distinguishing patterns compared to original ones, which leads to the failure of DCD detections. However, when a DCS happens, the DCT coefficients will still have an imperceptible yet regular change. According to [6], when an image is JPEG compressed over and over again with the same quantization matrix, the number of different DCT coefficients between sequential two compressed versions will monotonically decrease. The DCS detection method is conducted under the decreasing trend, altering a certain percentage (the *mpnc* value) of the DCT coefficients of each version to have the threshold τ. The number of different DCT coefficients between the test image and its recompressed version is denoted as D; the decision rule is

$$\begin{cases} \text{if } \tau \geq D, & \text{test image is a doubly compressed image} \\ \text{if } \tau < D, & \text{test image is a singly compressed image} \end{cases}.$$

The premise of DCS detection is that the test image has a *homogeneous* compression property. All the coefficients either went through a secondary compression, or a single compression. While a tampered image contains both singly and doubly compressed region, the variation tendency of the DCT coefficients will not completely follow the decreasing tendency mentioned above. Thus, the detection rate will be reduced.

III. Countermeasure Against DCS Detection

A. Principle of the Proposed Algorithm

From section II we can see that, having a JPEG image tested under DCS detection only has two results: the image is singly compressed or doubly compressed. Denote the detection operator by *Dete*, and the test image $F(x)$; the singly and doubly compressed images are denoted by $S(x)$ and $D(x)$, respectively. If $F(x)$ is determined as a singly compressed image, the detection result is signed as value 1; otherwise 0. Then the DCS detection can be abstractly presented as

$$Dete(F(x)) = \begin{cases} 1 & F(x) = S(x) \\ 0 & F(x) = D(x) \end{cases} \quad (1)$$

For most of the time, image forgery only has to modify part of the original image. Therefore, a tampered image $F(x)$ falls into the unchanged region $F_1(x)$, and the tampered region $F_2(x)$, so that

$$F(x) = \{F_1(x), F_2(x)\} \quad (2)$$

The traditional image forgery process is illustrated in Fig. 1(a). In this process, $F_1(x)$ does not have any other changes except for going through a double compression. Thus $F_1(x)$ can be denoted as $D_1(x)$, which means the doubly compressed region of $F(x)$; whereas $F_2(x)$ usually comes from an unknown source and goes through complex post-modification operation [5], it will most likely has the trace of single compression. Therefore, we can denote $F_2(x)$ as $S_2(x)$. Now a traditionally tampered JPEG image can be expressed as

$$F(x) = \{F_1(x), F_2(x)\} = \{D_1(x), S_2(x)\}. \quad (3)$$

By (3), it can be understood that if double compression can be avoided in the unchanged region, namely $F_1(x)=S_1(x)$, then

$$F(x) = \{F_1(x), F_2(x)\} = \{S_1(x), S_2(x)\} = S(x), \quad (4)$$

$$Dete(F(x)) = Dete(S(x)) = 1. \quad (5)$$

The detection result now is a singly compressed image. That is to say, the DCS detection fails at this point.

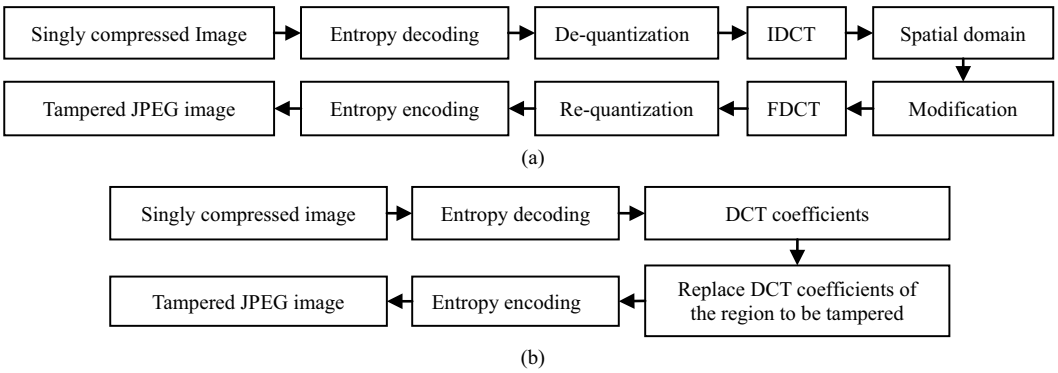

Figure 1. Comparison of the traditional JPEG image forgery and the proposed anti-forensic algorithm. (a) Flow chart of the traditional JPEG image forgery. (b) Flow chart of the proposed anti-forensic algorithm.

Based on the discussions above, we propose an anti-forensic algorithm that could counter the DCS detection. The flow chart of the proposed algorithm is shown in Fig. 1(b). Firstly, the original singly compressed JPEG image is decompressed to quantized DCT coefficient level, and then the DCT coefficients of the region needs to be modified are replaced. After that, the image is entropy encoded to form a tampered image.

During this process, $F(x)$ can be re-divided into the replaced region and the non-replaced region. The replaced region is the external rectangle of the tampered region $F_2(x)$, denoted by $F'_2(x)$; the region other than the replaced region is the non-replaced region, denoted by $F'_1(x)$. Now define the tampered image generated by the proposed algorithm as $F'(x)$, we have

$$F'(x) = \{F'_1(x), F'_2(x)\}. \tag{6}$$

Here the non-replaced region is kept the same with the corresponding region in the original image, which is singly compressed; so we can set $F'_1(x) = S'_1(x)$. Whereas the replaced region contains both singly and doubly compressed parts. Therefore, we have

$$F'_2(x) = \{D'_2(x), S'_2(x)\}. \tag{7}$$

Since $D'_2(x)$ in (7) can be fairly small, so that $D'_2(x) \ll S'_2(x)$, therefore $F'_2(x)$ can be approximated as $S'_2(x)$. Consequently, $F'(x)$ can be expressed as

$$F'(x) = \{F'_1(x), F'_2(x)\} = \{S'_1(x), S'_2(x)\} = S'(x). \tag{8}$$

From (8) and (1), we have

$$Dete(F'(x)) = Dete(S'(x)) = 1. \tag{9}$$

The detection result is a singly compressed image. From the derivation above, it can be inferred that the proposed algorithm could perfectly counter the DCS detection. The structure of the tampered images created by the proposed algorithm is illustrated in Fig. 2.

B. Implementations of the Proposed Algorithm

Specific steps of the proposed anti-forensic scheme are as follows:

1) Define the original singly compressed JPEG as J, the quanlity factor (QF) of J is Q_J. Suppose certain part of another image J' needs to be pasted into J for hiding the confidential information. The pasted part is denoted by ØS', as shown in Fig. 3(b).

2) Duplicate J to have J_d. Decompress J_d and J' to the spatial domain, duplicate and paste ØS' to J_d; use the tools in Photoshop to make ØS' and J_d blend naturally. Recompress J_d to a JPEG file using QF Q_J. The modified and recompressed version of J_d is denoted by J_m. (Fig. 3(c))

3) In J_m, locate the macroblocks in the external rectangle of ØS', as illustrated in Fig. 3(c), (e), and then obtain the quantized DCT coefficients of the located macroblocks. The collection of the obtained DCT coefficients is defined as ØD. (Fig. 3(e))

4) Entropy decoding the original image J to get the quantized DCT coefficients of the whole image; Use ØD to

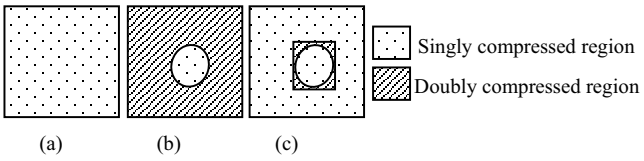

Figure 2. Structure of the tampered images produced by the anti-forensic algorithm. (a) Original singly compressed JPEG image. (b) Tampered JPEG image created by traditional forgery method. The circular tampered region is singly compressed; the rest of the image is the unchanged region with DQ effect; (c) Tampered JPEG image created by the anti-forensic scheme. DCT coefficients of the rectangular area (external rectangle of the tampered region) are replaced by other DCT coefficients. The non-replaced region outside the external rectangle is still singly compressed.

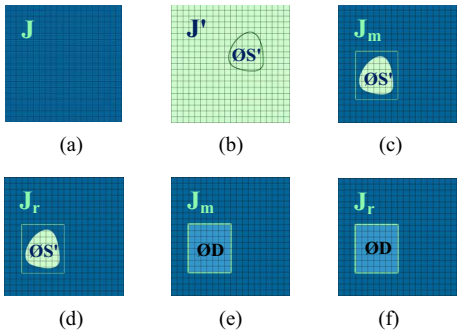

Figure 3. The proposed anti-forensic algorithm (a)Source image J. (b)Splicing souce image J'. (c)Modified and recompressed version of J's copy version J_d. (d)Final forged image J_r. (e) J_m is the source of the replacement coefficients ØD. (f) Replace ØD to the coefficients of J to get the tampered image J_r.

replace the DCT coefficients of the corresponding region in J. After that, entropy encoding the modified DCT coefficients to have the final forged image J_r. As shown in Fig. 3(f), the macroblocks outside the rectangular region are all singly compressed, which is the same with J, or virtually J itself. The replacement operation of the DCT coefficients can be realized by the IJG library [14].

IV. Experimental Results and Discussions

The experiments in this paper include two parts. The visual effects of the tampered images created by the anti-forensic algorithm are illustrated in the first part; the proposed algorithm against the DCS detection is evaluated in the second part.

A. Visual Effects of the Tampered Images

In the second step of our algorithm, we use the tools in Photoshop to blend ØS' and J_m; the area of the replaced rectangular region is moderately expanded, so that the edge of the replaced region can be fully overlapped with the original image J. The visual effects of the tampered images are illustrated in Fig. 4 (a) ~ (d), within which the tampered part is marked in dotted line. We can see that the tampered images have a quite natural look with no visually detectable traces.

B. Evaluation of the Proposed Algorithm Against DCS Detection

In this section, the robustness of the anti-forensic algorithm against the DCS detection is tested. The tampered test images are set to different groups; each group differs in the area of the replaced region, the content of the replaced region or in the quality factor. In this way, we can have the robust performance under variable conditions, so that the universality of our algorithm can be proved.

The tampered test images are generated as follows. 100 bmp-format images with the size 512×341 transcoded from RAW-format images taken by group members of our laboratory is denoted as the RAW library. All the 100 uncompressed images of RAW are singly compressed with

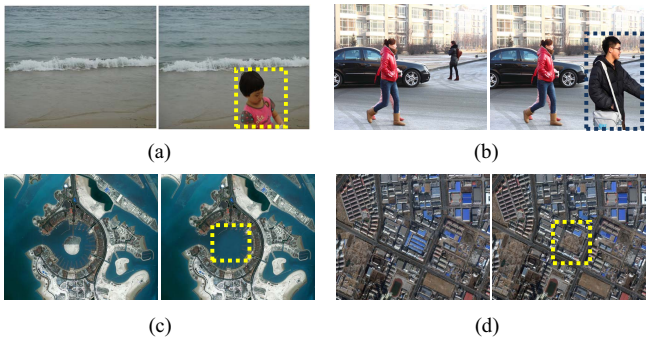

Figure 4. Visual effects of the tampered images. Images of (a) ~ (d) in the left-hand side are original versions, and the right-hand side are tampered versions.

QF=75, 85 and 95 respectively. These singly compressed JPEG images are used for the source of the original image J; then randomly choose 20 images from RAW, and JPEG compress them with different QFs, which are also randomly selected. The 20 JPEG images will be used as the source of J', which can be denoted by (J'_1, J'_2, ⋯, J'_{20}). Using J and J' as raw materials to generate two group of tampered images with different replaced area and different quality factors. In the first group, the replaced region accounts for 50% of the tampered images and in the second group 25%.

To be more explicit, we give a detailed production process of the first image group: the 100 original singly compressed images with QF 75 are denoted as J_i (i=1, 2, ⋯, 100), use J_i and J'_1 as the materials of the proposed anti-forensic method. The area of the replaced region takes 50% of J, and the content of the replaced region from J'_1 is identical for each J_i. In this way, we get 100 tampered images. Use the DCS detection method to detect the 100 tampered images, and then we have a percentage value Double Ratio (DR) that is the number of images that has been predicted as doubly compressed in 100 tampered images. As shown in Fig. 5 (a), the first dot in the blue broken line.

To finish the whole experiment, we in turn have the 100 original JPEG images and the other 19 J'_i (i=2,⋯,20) go through the same processing. Then we get the broken line which reflects the robustness of our anti-forensic algorithm against the DCS detection, under the conditions of QF=75, replaced region area 50% and different replace content. As shown in the blue broken line in Fig. 5 (a). Take the images with QF 85 and 95. Repeat the experiments above to get the red and green broken line in Fig. 5 (a).

Change the area of the replaced region to 25%, and repeat the experiment procedures above. Then we get Fig. 5 (b), which shows the robust performance of the proposed method against the DCS detection under a smaller replacement area.

What we have to point out again is that, as it is discussed in [5], in most cases, the tampered region in the forged images is singly compressed. So whether J' is singly compressed or doubly compressed will not be an influential issue to the detection results.

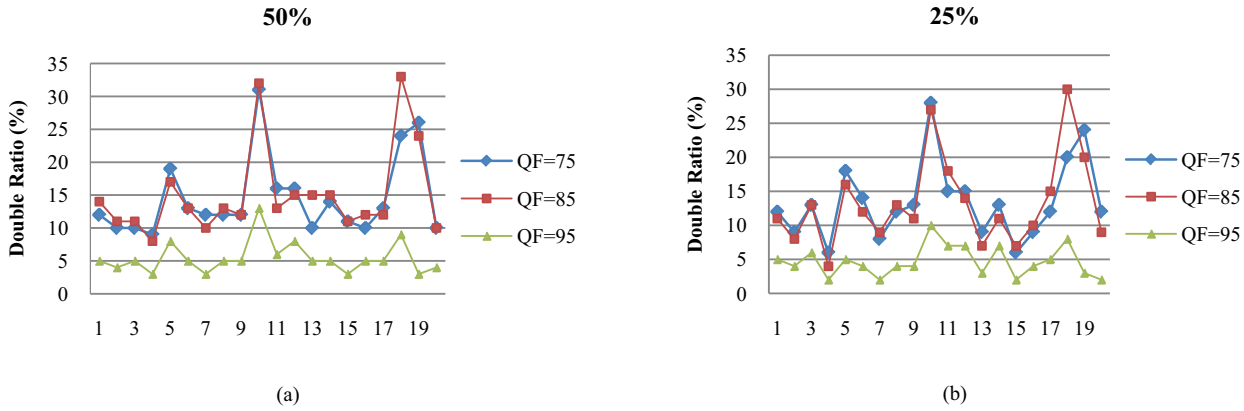

Figure 5. Robustness of the proposed algorithm under different replacement content and different QFs, as well as different replacement area. The horizontal axis represents the index of J'. (a) Robustness under 50% replacement area. (b) Robustness under 25% replacement area.

Compute the average value of DR under different replacement area and different QFs to get Table 1.

From Fig. 5 and Table 1 we can see that:

1. After testing thousands of tampered images with different replacement content, different replacement area and different QFs, the average value of the detection rate DR always stays in the margin error of the DCS detection. That is, the proposed anti-forensic algorithm is proved to be very robust against the DCS detection.

2. Under the condition of different QFs, the tendencies of the DR curves are almost the same. For instance, In Fig. 5, DR values in the 5th, 10th, and 18th J' under different QFs are all peak values; and in the 4th, 7th and 15th J', DR values are all valley values. This illustrates that when the original JPEG image J stays the same, detection rate will be different because of the different replacement content.

3. The DR curves of QF 75 and 85 are almost overlapped with each other in both Fig. 5 (a) and (b). Whereas the DR curve of QF 95 is obviously much lower than the other two. This result is consistent with the result drawn in [6]: a much higher quality factor can get a much accurate detection result.

V. Conclusion

This paper has taken a critical view on the reliability of JPEG-format-based forensic techniques that established on detecting the DQ effect of tampered images. In particular, we have presented and evaluated an anti-forensic algorithm to defeat a specific method that has been developed to unveil the operation of double compression with the same quantization matrix.

Through the experiment results obtained in our paper, double compression can be avoided in the unchanged region of the forged image, so that the DQ effect is mostly eliminated. Therefore, existing JPEG-format-based forgery detection method will lose effectiveness.

The robustness of the proposed algorithm against the DCS detection is demonstrated by numerous experiments in this paper. The experiments are carried out under the conditions of different replacement area, different quality factors as well as different replacement contents; the proposed scheme could counter the DCS detection effectively. Moreover, since our algorithm actually eliminates the DQ effect in the tampered image, it can be expect that the proposed algorithm could also counter other forgery detection algorithms based on detection of DQ effect.

Since our method is technically a splicing tampering, it can only be used in countering a specific application of [6], as declared in section I. However, we believe the proposed method can work together with other anti-forensic methods to achieve more comprehensive and powerful functionality.

TABLE I. Average value of DR under different replacement area and different QFs

Average value of DR (%)	QF=75	QF=85	QF=95
Replacement area: 50%	14.5	15.1	5.5
Replacement area: 25%	13.4	13.3	4.7

Acknowledgment

The work on this paper was supported by the State Key Development Program for Basic Research of China (973 Program) (Grant No. 2011CB302204).

References

[1] G. Cao, Y. Zhao and R.R. Ni, "Multimedia authentication,". CCCF. Vol. 7, No. 2, 2011, pp. 38-44.

[2] H. Farid, "Image forgery detection," IEEE Signal Processing Magazine. Vol. 26, No. 2, 2009, pp. 16-25.

[3] J. Lukáš, J. Fridrich, "Estimation of primary quantization matrix in double compressed JPEG images," The International Society for Optical Engineering. Vol. 6819, 2008.

[4] A. C. Popescu, H. Farid, "Statistic tools for digital forensics," Lecture Notes in Computer Science. Vol. 3200/2005, 2005, pp. 395-407.

[5] Z. C. Lin, J. F. He, and X. O. Tang, "Fast, automatic and fine-grained tampered JPEG image detection via DCT

coefficient analysis," Pattern Recognition. Vol. 42, 2009, pp. 2492-2501.

[6] F. J. Huang, J. W. Huang and Y. Q. Shi, "Detecting double JPEG compression with the same quantization matrix," IEEE Transactions on Information Forensics and Security. Vol. 5, No. 4, 2010, pp. 848-856.

[7] M. Kirchner, R. Böhme, "Tamper hiding: defeating image forensics," Proceedings of the 9th International Conference on Information Hiding. 2007, pp. 326-341.

[8] T. Gloe, M. Kirchner and A. Winkler, "Can we trust digital image forensics," Proceedings of the 15th International Conference on Multimedia. 2007.

[9] M. Kirchner and R. B¨ohme, "Synthesis of color filter array pattern in digital images," in Proc. SPIE-IS&T Electronic Imaging: Media Forensics and Security. Vol. 7254, 2009.

[10] G. Cao, Y. Zhao and R. R. Ni, "Anti-forensics of contrast enhancement in digital images," ACM Multimedia and Security Workshop. 2010, pp. 25-34.

[11] M. C. Stamm, K. J. R. Liu, "Anti-forensics of digital image compression," IEEE Transactions on Information Forensics and Security. Vol. 6, 2011, pp. 1050-1065.

[12] G. Valenzise, M. Tagliasacchi and S. Tubaro, "The cost of JPEG compression anti-forensics," 2011 IEEE International Conference on Acoustics, Speech and Signal Processing. 2011, pp. 1884-1887.

[13] S. Y. Lai, R. Böhme, "Countering counter-forensics: the case of JPEG compression," Lecture Notes in Computer Science. Vol. 6958, 2011, pp. 285-298.

[14] IJG (Independent JPEG Group). Available: http://www.ijg.org/