PHASE BASED FORGERY DETECTION OF JPEG ANTI FORENSICS

*Gamal Fahmy**

Electrical Engineering Dept.
Assiut University, Egypt
Prince Sattam University, KSA
gamal.fahmy@rub.de

Rolf Wurtz

Institut fur Neuroinformatik
Ruhr-Universitt Bochum, Germany
rolf.wuertz@ini.rub.de

ABSTRACT

Many recent techniques in forgery detection tried to counter noise dithering, which is well adopted in removing footprints of JPEG editing in any image tampering (anti forensics) process. In this paper we present a novel idea of detecting this noise dithering based on detecting spatial frequency phase variations for different blocks of suspect-able images. We try to measure the level of coherence of phase values for spatial frequencies of image pixels and detect threshold variations/distributions that would indicate some editing or tampering of images. The proposed technique is robust against noise dithering due to the fact that local homogeneous regions inherit distinctive phase values. These phase values are inconsistent with embedded or dithered noise signals that are considered to be out of phase and can be easily detected in noisy environments. Our proposed technique is compared with recent successful literature techniques for performance with degraded tampered/hacked images. Unlike other literature techniques, Our proposed approach does not require the knowledge of the quantization matrix that is typically adopted in original compression for forgery detection, as will be illustrated.

Index Terms— image forensics, JPEG edits, anti forensic

1. INTRODUCTION

The recent availability of photo editing software with the wide-spread use of digital imagery in different applications has made it easier than ever to edit, manipulate and tamper the content of digital images. While manipulating different uncompressed image versions (i.e. bmp) can be easier to tamper or hack, compressed images with the widely used JPEG standard, are harder to edit and tamper undetectably, due to traces left by the JPEG compression. Several JPEG forensics and anti-forensics detection techniques have appeared in the last decade to detect either forgery of compressed images [1–3], or removals of hacking footprints in counter forensics [4, 5]. We note here that the noise dithering technique presented in [4, 5], has been considered in later literature, [6–8], as a robust technique for countering forensics by introducing a dithering signal in the DCT domain that will match the distribution of tampered images to the original one and hide any possible hacking. This dithering signal has to be sufficient to fill in all gaps, but if it exceeded a certain limit it can also be detected. This detection can occur by re-compressing the dithered image with different compression parameters such as the JPEG quantization matrix (QM) [9], or quantization step size [6], or the QM size [8]. In [7], another novel idea was introduced by re-compressing decoded JPEG images until quantization noise saturates, and it has been reported to be the most robust technique in noisy environments as in [10, 11]. In [12, 13], it has been agreed upon that the amount of noise significantly affects the performance of the detection of forgery, although it has also been highlighted that there are some sweet spots where the sensitivity of this detection does not get affected with bad image quality [14] and also in [15].

In this paper we present a novel idea of exploiting spatial frequencies to find distinctive image features to detect any forgeries. Spatial frequency phase and magnitudes are simultaneously analyzed in local image regions. Spatial frequency phase and magnitudes have been long considered by human visual system and computer vision researchers [16, 17]. Marr in [18], designed several models to measure how humans perceive phase data and indicated that homogenous data samples tend to have close spatial frequency phase values, while noise samples and coefficients tend to be out of phase and can be easily detected from their phase values. In our proposed model we check for any spatial frequency phase values that are inconsistent in local regions but have their spatial frequency magnitude values consistent (skewed/matched) in same regions. This would indicate possible JPEG tampering and noise dithering and can be detected even in noisy graded image regions. This can locate possible noisy dithered regions of suspect-able images with outstanding performance compared to current literature techniques. This hypothesis came in light of recent video processing literature that focused on phase information for complex transform coeffi-

*Thanks for Alexander von Humboldt foundation, Germany for funding

978-1-5090-5844-0/16/$31.00 ©2016 IEEE

cients and highlighted that it may reveal useful information that might not have been detected before [19]. This analogy is based on the fact that different types of noise would have the same transform coefficient magnitude values with attached data, but different transform coefficient phase values due to the random nature of noise. Hence any amount of injected dithered noise would not corrupt the magnitude distribution of transform coefficients [4,5,20], but it can be detected from phase discrepancies.
We also tested our technique with different noise levels to measure its robustness against noise. Section 2 gives an overview about our proposed spatial frequency phase based detection technique; section 3 shows how we employed this technique for countering anti forensics. Section 4 shows our comparison with literature techniques in [6] and [7]. Some preliminary results were reported in [21], but with much less details, we here fully investigate the idea and the results.

2. SPATIAL FREQUENCY BASED IMAGE ANALYSIS

Any spatial domain (pixel domain) consists of different sinusoidal frequencies in the transform domain, each of these frequencies is represented as a spatial frequency magnitude (SFM) and a spatial frequency phase (SFP), as with any sinusoidal signal. If we take the Fourier transform of any image 8x8 block B_i, and note every Fourier coefficient in the k^{th} and j^{th} row and column as Z_{kj}, then a and b would correspond to the real and complex parts for that coefficient. R and θ would correspond the magnitude and phase of this complex coefficient as $R = \sqrt{a^2 + b^2}$, while the phase component is calculated as $\theta = (\frac{a}{b})$.
In the Human Visual System, the content of the perceived image is analyzed through multiple independent spatial frequency channels [18]. Although these channels do not interact substantially with each other, in smooth image regions spatial frequency phase values of these channels are coherent with each other. This level of spatial frequency phase coherence is perceived as added visual features against noise, even if this image region does not enjoy similar spatial frequency magnitude values. When the phase of the various spatial frequency components in an image region are not coherent in these same regions, this would indicate presence of noise, as noise is typically out of phase and does not correlate well with different image contents. It is our own belief, that phase data can not be adjusted with any type of dithering while adjusting the edited image region, as well be shown in next section.
We note here that quantization noise is typically added during any JPEG compression process, however quantization noise has fixed levels in every quantization range, unlike noise dithering, hence it can be distinguished as will be shown in section 4.
In noise dithering [4, 5, 20], where added noise is supposed to fill in quantization gaps of decoded image coefficients as follows: The histogram of the quantized DCT coefficients $\tilde{X}_i$ is like a comb-shap, a noisy signal N_i is supposed to fill in the gaps so that the distribution of the dithered signal Y_i approximates the original distribution of X_i before being quantized. Since X_i is assumed to have a Laplacian distribution [22], we give values to the AC zero coefficients according to:

$$f_{x_i}(x) = \frac{\lambda_i}{2} e^{-\lambda_i |x|}$$

where the decay parameter λ_i takes a value between 1 and 10^{-3} for natural images and is typically calculated from quantized coefficients $\tilde{X}_i$ using the maximum likelihood method [14, 22]. Hence the dithered signal N_i needs to be designed in such a way that its distribution depends on whether $\tilde{X}_i$ are zero or not. So for DCT coefficients quantized to zero:

$$f_{N_i}(n|\tilde{X}_i = 0) = \begin{cases} \frac{1}{c_0} e^{-\lambda_i |n|} & \text{if } -\frac{q_i}{2} \leq n < \frac{q_i}{2} \\ 0 & \text{otherwise} \end{cases}$$

where $c_0 = (2/\lambda_i)(1 - e^{\lambda_i q_i/2})$, similarly for other nonzero DCT coefficients

$$f_{N_i}(n|\tilde{X}_i = x) = \begin{cases} \frac{1}{c_1} e^{-sgn(x)\lambda_i (n+q_i/2)} & \text{if } -\frac{q_i}{2} \leq n < \frac{q_i}{2} \\ 0 \quad \text{otherwise} \end{cases}$$

As also shown in [13], for the DC DCT coefficient, we add uniform noise so that the overall distribution is Laplacian [22].
Even though noise is supposed to have similar spatial frequency magnitudes in all dithered image regions, for Fourier histogram coefficients (different frequencies), noise will not be phase consistent in this dithered region. Hence the degree of phase coherence in different image regions, can be used as an indication of the amount of added noise to this region of the image, which would correspond to whether it was edited or not.
We note here that noise dithering would add noise coefficients that match the decoded image coefficients in magnitude, hence the Fourier coefficient histogram will not detect any magnitude discrepancies, but there is a large amount of phase inconsistencies, which is an indication of added dithered noise. As noted before this phase inconsistency for dithered noise would be different than the uniform quantization noise as it is mostly random [13].

3. DETECTION OF SPATIAL FREQUENCY FORGERY

We propose the following procedure, we analyze both the local spatial frequency phase and magnitude values for 8x8 non overlapping image blocks B_i's, and a forgery flag is set if the distribution of phase values is inconsistent with normal Gaussian distribution, while magnitude values matches the Gaussian distribution, for the whole image. We propose the following steps:

1. Measure the local variance for each B_i of the original image, that corresponds to a specific 8x8 phase value block PHS_i.

2. If a PHS_i block has a large amount of edges (phase non smooth peaks) with a high variance value, we discard this block, this edge detection process is performed on each block at the pixel level.

3. Calculate the Fourier Transform for each 8x8 non overlapping pixel blocks of the original image, measure the distribution of phase values for every 8x8 Fourier block of decoded coefficients by measuring the histogram of θ for all block Fourier coefficients, Fig.3.

4. Measure the difference between the spatial frequency phase histogram and the Gaussian distribution, with histogram matching, [23], note γ as the histogram difference value(γ is experimentally chosen as the sum of the vector absolute difference between the two distributions).

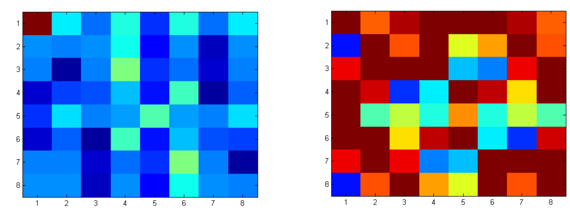
5. Measure the histogram of spatial frequency magnitudes of the same 8x8 block, Fig.2.

6. Determine the difference between the spatial frequency magnitude histogram and the Gaussian distribution, with the same histogram matching approach, [23], note η as the histogram difference value (similarly to γ, η is experimentally chosen as the sum of the vector absolute difference between the two distributions).

7. If the difference value γ-η is greater than a threshold value, then this would indicate that this is a hacked block.

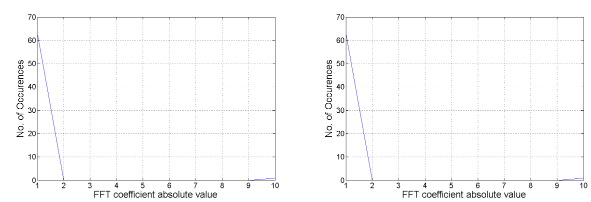
We note here that the local variance measured in step 2, is to determine the smoothness of the block to detect if this block of the image is likely to be edited or not. This is based on the hypothesis that only low variance regions would correspond to a homogenous region and they are likely to be attacked, while high variance regions (edge regions) are less likely to be attacked. Edge regions would consist of a large spectrum of spatial frequencies, hence most of the 8x8 JPEG-DCT frequencies would have large non zero values and the amount of injected dithered noise is almost zero. We also note here that in noise dithering (injection), we add a specified nonzero noise to DCT coefficients (real values domain, no imaginary values), hence the magnitude of the DCT coefficients can be adjusted, but the location of these newly added noise values is random within each 8x8 block. This will lead to the fact that the magnitude of the spatial frequencies can be adjusted in the Fourier domain as it is related to the value of the DCT coefficients for every 8x8 block, but the phase values can not as it correspond to the location for every DCT coefficient, that is chosen randomly. Fig.1, shows DCT values in an 8x8 block for both dithered and non dithered versions, fig. 2, shows the histogram for magnitude values of Fourier coefficients of that block, while fig.3 shows the histogram for phase values of Fourier coefficients of that same block. The horizontal axis in Fig.2 Fig.3 is labeled 1 to 10, just to indicate the full spectrum of Fourier magnitude or phase values. It can be clearly noticed the impact of noise dithering on FFT phase values, while FFT magnitude values remain unchanged.

We also note that quantization noise would have a semi Gaussian distribution, hence it can be easily distinguished from injected dithered noise that has a distribution far from being Gaussian, Fig.5. We based our flagging of suspectable blocks if phase values, that are typically harder and less likely to be altered, have an unexpected distribution that does not match the magnitude distribution, which is more likely to be modified



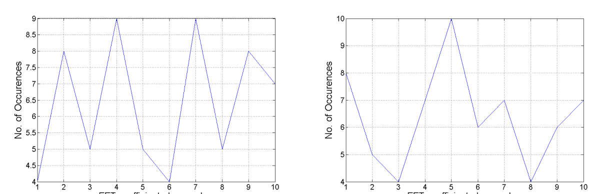
(a) Normal DCT values (b) DCT values with dithering

Fig. 1: Average DCT values for 8x8 blocks with and without dithering, it is clear that there is a difference in value distributions



(a) without dithering (b) with dithering

Fig. 2: Histogram of FFT absolute values for 8x8 block



(a) without dithering (b) with dithering

Fig. 3: Histogram of FFT phase values for 8x8 block

4. EXPERIMENTAL RESULTS

Table 1 lists the PSNR and the detection accuracy values obtained for different quality factors using our proposed tech-

Table 1: Detection accuracy along PSNR values

Q	Lena PSNR	Lena accuracy	Mandril PSNR	Mandril accuracy
30	28.28	%93	31.06	%92
40	31.2	%91	31.9	%89
50	32.28	%89	32.6	%84
60	34.8	%87	33.1	%81
70	36.1	%83	33.5	%78
80	37.4	%81	33.9	%76
90	38.6	%76	34.3	%73
95	39.23	%71	35	%71

nique. We based our assumption that original non-dithered distributions of phase values should follow a semi Gaussian distribution, as all tested images were nature outdoor images that follow that same distribution, hence prospective non attached phase or magnitude histogram distributions should follow the same model.

We adopted the successful technique of [4, 5] in injecting noise in different homogeneous and non-homogeneous image regions and tried to detect possible forgeries and hacking with our proposed technique. It was our own belief, that phase data can not be adjusted with any type of dithering while adjusting the edited image region. We compared our detection performance with several recent literature techniques as in [6, 7], as in Fig.4. In our tested data set of images we selected *different* 280 images that are JPEG compressed with different quality factors ranging from 30 to 95, which indicates different noise levels (Quantization noise). A quality factor of 100 was excluded due to the little amount of noise in it in general, hence our approach will not be applicable to it. For each image we also made a dithered version, where a counter forensic dithered noise signal is added in different regions of the image. This would make our test samples 560 images (half original, and half forensically countered). Fig.4 shows the average detection performance of our phase coherence based anti-forensic technique for JPEG images compressed with different Quality factors (30 to 95). Performance of the most recent successful literature techniques is also compared in Fig.4.

In our simulation results we measured the level of phase coherence of all images for both original and dithered versions. The resulting output was just a binary value of edited or not edited. While all original images were correctly classified as not edited images, different classifications were obtained for dithered images based upon region type (high/low variance) and JPEG quality factor. In our experiments we selected an accuracy metric as our performance metric for forgery detection. This accuracy metric is defined as the number of correctly identified noise dithered images as edited or not edited, to the total number of test sample images. This choice of accuracy is justified by the balanced database that we had that consisted of equal number of original and attached (edited or dithered) images. Fig.5 shows an example of the average phase distribution of image regions in Fourier domain with different quality factors (30-95), for original pixel regions, non injected (Non Dithered) reconstructed region and Dithered noise injected reconstructed region. This average distribution was for our entire 560 images.

It can be easily noticed that injected dithered noise has a distribution different than regular quantization noise and can be easily identified. This difference in distribution is due to the incoherence in spatial frequency phase that dithered noise enjoys with regular Fourier coefficients. Quantization noise is modeled and bounded in different pixel domain quantization levels, this makes noisy non dithered coefficients have a Gaussian phase distribution similar to original image pixels because of this coding modeling and bounding features. Hence quantization noise is different in distribution from dithered noise and can be categorized with original pixels in spatial frequency distributions. Fig.6 shows the difference in detection performance for dithered and non dithered image regions for different quality factors. It can be noticed that this difference is minimized with a higher quality factor.

Fig.7 shows an edited image with some dithering around the embedded object (door) in a homogenous region and the detected phase inconsistency in it. This is meant to show that the phase is inconsistency in Fourier domain would have the shown impact when it is reverted back to pixel domain.

5. CONCLUSIONS

In this paper a novel technique was presented for the detection of forgery or tampering of JPEG images, the proposed approach relied on detecting spatial frequency phase data and is primarily based on the assumption that injected dithered noise is typically out of phase in its embedded region, hence any noise dithering can be detected especially in homogenous regions where any phase discrepancy is easily recognizable. We make a distinction between quantization noise (that exist between pixel values) and dithered injected noise that is out of phase when analyzing spatial frequency phase distribution. A promising future direction of research can be investigating the tradeoff between forgery detection sensitivity and human sensitivity for the image in noisy environments.

6. REFERENCES

[1] A.J. Fridrich, B.D. Soukal, and A.J. Lukáš, "Detection of copy-move forgery in digital images," in *in Proceedings of Digital Forensic Research Workshop*. Citeseer, 2003.

[2] Z. Fan and R.L. de Queiroz, "Identification of bitmap compression history: Jpeg detection and quantizer estimation," *Image Processing, IEEE Transactions on*, vol. 12, no. 2, pp. 230–235, 2003.

[3] G. Valenzise, M. Tagliasacchi, and S. Tubaro, "The cost of jpeg compression anti-forensics," in *Acoustics,*

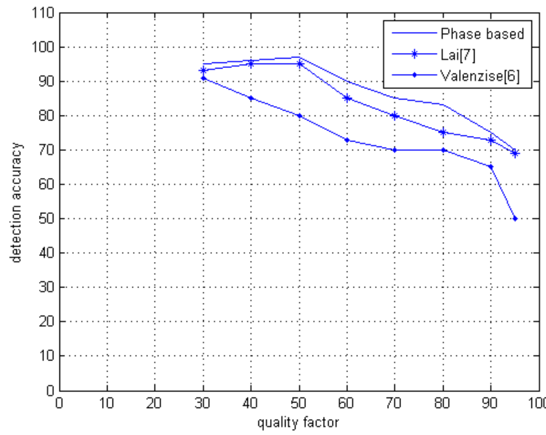


Fig. 4: Comparison of forgery detection performance

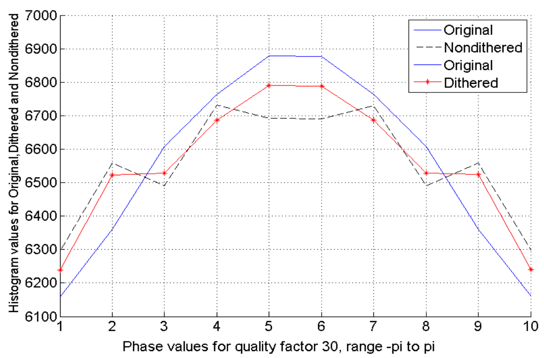


Fig. 5: Average phase distribution with Quality Factor 30-95

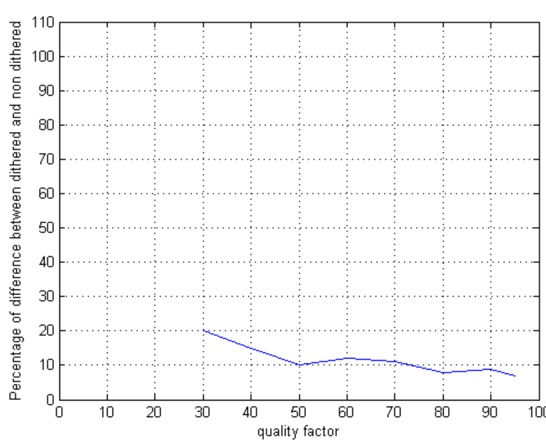


Fig. 6: Difference of detection performance between dithered and non dithered images

Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on. IEEE, 2011, pp. 1884–1887.

[4] M.C. Stamm, S.K. Tjoa, W.S. Lin, and K.J.R. Liu, "Undetectable image tampering through jpeg compression anti-forensics," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE, 2010, pp. 2109–2112.

[5] M.C. Stamm, S.K. Tjoa, W.S. Lin, and K.J.R. Liu, "Anti-forensics of jpeg compression," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1694–1697.

[6] G. Valenzise, V. Nobile, M. Tagliasacchi, and S. Tubaro, "Countering jpeg antiforensics," in *IEEE Int. Conf. on Image Processing, ICIP*, 2011.

[7] ShiYue Lai and Rainer Bhme, "Block convergence in repeated transform coding: Jpeg-100 forensics, carbon dating, and tamper detection.," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2013*, Vancouver, BC, Canada, 2013, pp. 3028–3032, IEEE.

[8] Gamal Fahmy, "Detectable tampering of jpeg anti-forensics," in *WIAR'2012; National Workshop on Information Assurance Research; Proceedings of*. VDE, 2012, pp. 1–4.

[9] H. Farid, "Exposing digital forgeries from jpeg ghosts," *Information Forensics and Security, IEEE Transactions on*, vol. 4, no. 1, pp. 154–160, 2009.

[10] Gamal Fahmy, "Nonorthogonal dct block convergence for jpeg-75 forensics," in *ISSPIT 2014;International Conference on Signal Processing and Information Technology; Proceedings of*. IEEE, 2014.

[11] G. Valenzise, M. Tagliasacchi, and S. Tubaro, "Detectability-quality trade-off in jpeg counter-forensics," in *IEEE Int. Conf. on Image Processing, ICIP*, 2014.

[12] Xiaoyu Chu, Matthew C Stamm, Yan Chen, and KJ Liu, "Concealability-rate-distortion tradeoff in image compression anti-forensics," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 3063–3067.

[13] Xiaoyu Chu, Matthew Christopher Stamm, Yan Chen, and KJ Liu, "On antiforensic concealability with rate-distortion tradeoff," *Image Processing, IEEE Transactions on*, vol. 24, no. 3, pp. 1087–1100, 2015.

[14] Giuseppe Valenzise, Marco Tagliasacchi, and Stefano Tubaro, "Revealing the traces of jpeg compression anti-forensics," *Information Forensics and Security, IEEE Transactions on*, vol. 8, no. 2, pp. 335–349, 2013.

[15] Wei Fan, Kai Wang, Francois Cayre, and Zhang Xiong, "Jpeg anti-forensics with improved tradeoff between forensic undetectability and image quality," *Information Forensics and Security, IEEE Transactions on*, vol. 9, no. 8, pp. 1211–1226, 2014.

[16] Eric Wharton, Karen Panetta, and Sos Agaian, "Human visual system based similarity metrics," in *Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on*. IEEE, 2008, pp. 685–690.

[17] Gamal Fahmy, John Black Jr, and Sethuraman Panchanathan, "Texture characterization for joint compression and classification based on human perception in the wavelet domain," *Image Processing, IEEE Transactions on*, vol. 15, no. 6, pp. 1389–1396, 2006.

[18] David Marr et al., "Vision: A computational investigation into the human representation and processing of visual information," 1982.

[19] Neal Wadhwa, Michael Rubinstein, Frédo Durand, and William T Freeman, "Phase-based video motion processing," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 4, pp. 80, 2013.

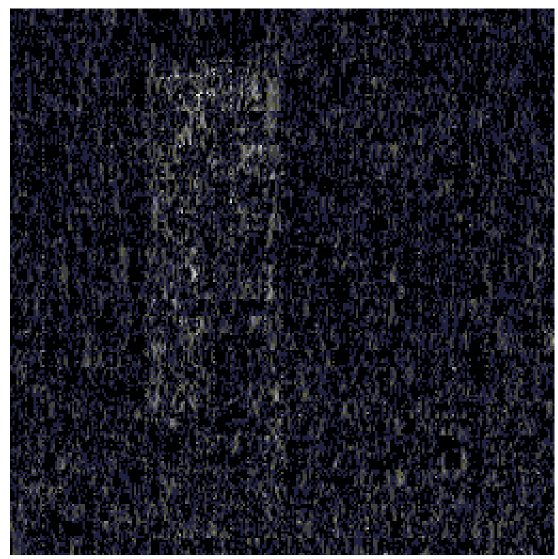
[20] Matthew C Stamm and KJ Liu, "Anti-forensics of digital image compression," *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 3, pp. 1050–1065, 2011.

[21] Gamal Fahmy, Abdullah Alqallaf, and Rolf Wurtz, "Phase based detection of jpeg counter forensics," in *IEEE International Conference on Electronics, Circuits, and Systems, Cairo Dec. 2015*, 2015.

[22] Edmund Y Lam and Joseph W Goodman, "A mathematical analysis of the dct coefficient distributions for images," *Image Processing, IEEE Transactions on*, vol. 9, no. 10, pp. 1661–1666, 2000.

[23] Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing (3rd Edition)*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.

(a) Normal Pixel values



(b) Phase inconsistency values

Fig. 7: A fake image with noise dithered around the embedded door with detected phase incoherence, white points indicate phase incoherence in Fourier domain