IMPROVED DCT COEFFICIENT ANALYSIS FOR FORGERY LOCALIZATION IN JPEG IMAGES

Tiziano Bianchi, Alessia De Rosa, Alessandro Piva

Università di Firenze, Dip. di Elettronica e Telecomunicazioni
Via S. Marta 3, 50139, Firenze - Italy

ABSTRACT

In this paper, we propose a statistical test to discriminate between original and forged regions in JPEG images, under the hypothesis that the former are doubly compressed while the latter are singly compressed. New probability models for the DCT coefficients of singly and doubly compressed regions are proposed, together with a reliable method for estimating the primary quantization factor in the case of double compression. Based on such models, the probability for each DCT block to be forged is derived. Experimental results demonstrate a better discriminating behavior with respect to previously proposed methods.

Index Terms— multimedia forensics, JPEG artifacts, Bayesian inference, forgery localization

1. INTRODUCTION

The large amount of digital images spreading everywhere and the ease of use of commercial image processing tools, make the diffusion of fake contents through the digital net increasing and worrying. An image resulting from the splicing of different photos conveys information distant from the original event captured by the original source: the information conveyed by the tampered image is fake and influences the opinion of viewers believing the credibility of the image they are watching. Such a problem has important effects in several fields in which the credibility of images should be verified before using them as sources of information, for example insurance, law and order, journalism, medical applications.

During the last years many image forensic algorithms have been proposed for exposing digital forgeries by means of different approaches based on the analysis of statistical and geometrical features, JPEG quantization artifacts, interpolation effects, demosaicing traces, feature inconsistencies, etc. [1]. Even thought such forensic tools seem to be helpful for the verification of image integrity, on the other hand their application in real scenarios has not been properly validated yet. In fact, most of the approaches proposed so far look for the presence/absence of some characteristics within a given region and rely on the hypothesis to know the right location of the forgery area. This issue should be carefully considered when forensic techniques are applied in practical cases, where images are provided without any a-priori information.

Since a large amount of images today is compressed in JPEG format, several forensics tools have explored the DCT coefficient characteristics for revealing some incoherence within the image [2] [3], but even in this case the local tampering detection remains an open issue. Some attempt has been done in such a direction, e.g. in [4] authors propose an algorithm for the detection of image regions that have been transplanted from another image and in [5] authors presented a method for the automatic localization of tampered regions in JPEG images, with a fine-grained scale of 8×8 blocks.

It is the aim of this paper to move one step more in this direction, by proposing an improved version of [5], where authors compute a probability map (described in Section 2) to differentiate tampered and unchanged regions. The model used to build the map has some main limitations that will be deeply analyzed and overcome by the new approach (Section 3), thus leading to a significant improvement of the accuracy of the probability map estimation and consequently of the algorithm performance, as shown in Section 4.

2. PREVIOUS WORK

In many scenarios, an original JPEG image, after some localized forgery, is saved again in JPEG format without resizing. Hence, DCT coefficients of unmodified areas will undergo a double JPEG compression thus exhibiting double quantization (DQ) artifacts, while DCT coefficients of forged areas will result from a single compression and will likely present no artifacts. In the following, we will refer to this scenario as the single compression forgery (SCF) hypothesis. In [5], it is firstly proposed to use the statistics of DCT coefficients of doubly compressed JPEG images to discriminate between tampered and original areas under the SCF hypothesis. The idea is to use Bayesian inference to assign to each DCT coefficient a probability of being doubly quantized. Such probabilities, accumulated over each 8×8 block, will provide a DQ probability map allowing us to tell original areas (high DQ probability) from tampered areas (low DQ probability).

Bayesian inference is based on the probability distribution of DCT coefficients conditional to the hypothesis of being tampered, i.e., $p(x|\mathcal{H}_b)$, where x is the value of the DCT coefficient and $\mathcal{H}_0$ ($\mathcal{H}_1$) indicates the hypothesis of being tampered (original). In [5], such conditional probabilities are derived by observing that the histogram of the DCT coefficients having same position within a 8×8 block exhibits a periodic pattern after double quantization. Namely, the number of bins of the original histogram that are mapped in the bin corresponding to the value x in a DQ histogram are given by:

$$n(x) = R(x) - L(x) \tag{1}$$

where $L(x) = Q_1 \left(\left\lceil \frac{Q_2}{Q_1} \left(x - \frac{1}{2} \right) \right\rceil - \frac{1}{2} \right)$, $R(x) = Q_1 \left(\left\lfloor \frac{Q_2}{Q_1} \left(x + \frac{1}{2} \right) \right\rfloor + \frac{1}{2} \right)$, Q_1, Q_2 are the quantization steps used in the first and second compression, respectively, and $n(x)$ is a periodic function, with period $P = Q_1 / \gcd(Q_1, Q_2)$. Once P has been estimated from the histogram of the DCT coefficients (see [5]

This work was sponsored by the European Commission under the Project LivingKnowledge (IST-FP7-231126).

for details) the authors of [5] propose to estimate the conditional probabilities as

$$p(x|\mathcal{H}_0) = 1/P \tag{2}$$

and

$$p(x|\mathcal{H}_1) = h(x)/\sum_{k=0}^{P-1} h(\lfloor x/P \rfloor P + k) \tag{3}$$

where h is the histogram of the DCT coefficients[1]. Under the hypothesis that the DCT coefficients within a block are mutually independent and that $\mathcal{H}_0$, $\mathcal{H}_1$ are equiprobable, the probability of a block being tampered can be estimated as

$$p = p(\mathcal{H}_0|x_0, \ldots, x_{63}) = \frac{\prod_i p(x_i|\mathcal{H}_0)}{\prod_i p(x_i|\mathcal{H}_0) + \prod_i p(x_i|\mathcal{H}_1)} \tag{4}$$

In [5], some features extracted from the probability map given by (4) are fed to a classifier in order to automatically detect whether the image is tampered or not.

3. PROPOSED DCT COEFFICIENT ANALYSIS

A limitation of the above described method is that the conditional probability $p(x|\mathcal{H}_1)$ is estimated according to the observed histogram of x. In the case of a tampered image, however, such a histogram will actually be a mixture of $p(x|\mathcal{H}_1)$ and $p(x|\mathcal{H}_0)$. Hence, for large forgeries we expect the histogram of x to be a poor estimate of $p(x|\mathcal{H}_1)$.

In order to overcome this limitation, we should be able to separate the two conditional probabilities from the observed mixture. By assuming that the histogram $h_0(x)$ of the DCT coefficients before the first JPEG compression is available, a better estimate of $p(x|\mathcal{H}_1)$ could be obtained as

$$p(x|\mathcal{H}_1) = \sum_{L(x) \le u < R(x)} h_0(u). \tag{5}$$

Unfortunately, (5) is difficult to use in practice, since it would require a reliable estimate of both $h_0(x)$ and Q_1. Hence, we propose to introduce the following approximation

$$\frac{1}{n(x)} \sum_{L(x) \le u < R(x)} h_0(u) \approx \frac{1}{Q_2} \sum_{L'(x) \le u < R'(x)} h_0(u) \triangleq \tilde{h}(x) \tag{6}$$

where $L'(x) = Q_2 x - \frac{Q_2}{2}$ and $R'(x) = Q_2 x + \frac{Q_2}{2}$. The above approximation holds whenever $n(x) > 0$ and the histogram of the original DCT coefficient is locally uniform. In practice, we found that for moderate values of Q_2 this is usually true, except for the center bin ($x = 0$) of the AC coefficients, which have a Laplacian-like distribution. Note that the right hand side of (6) is obtained by resampling $h_0(x)$ with step Q_2, i.e., $\tilde{h}(x)$ can be viewed as the histogram of the DCT coefficients after a single compression with quantization step Q_2. A simple technique for estimating $\tilde{h}(x)$ is to consider the DCT coefficients obtained by recompressing with the second quantization matrix a slightly cropped version of the tampered image [6].

According to the SCF hypothesis, we can estimate the conditional probabilities as

$$p(x|\mathcal{H}_0) = \tilde{h}(x) \tag{7}$$

and

$$p(x|\mathcal{H}_1) = n(x) \cdot \tilde{h}(x), \quad x \neq 0. \tag{8}$$

Based on the above equations, the probability of a block being tampered can be simply estimated as

$$p = 1/\Big(\prod_{i|x_i \neq 0} n_i(x_i) + 1 \Big) \tag{9}$$

where $n_i(x_i)$ indicates the function $n(x)$ related to the ith DCT coefficient within a block. Note that since both (7) and (8) may not be accurate when $x = 0$, the actual computation of the probability map does not take into account DCT coefficients equal to zero. Moreover, due to the fact that most of DCT coefficients are zero at high frequencies, only the first low frequency coefficients are used in practice.

3.1. Estimation of Q_1

As can be seen, in order to compute the probability map with (9) we need a reliable estimate of the quantization table used by the first JPEG compression. A method for estimating Q_1 in doubly compressed JPEG images is proposed in [6]. Unfortunately, the method in [6] does not take into account the fact that under the SCF hypothesis the observed histogram of the DCT coefficients is a mixture. Hence we propose a simple yet effective method, partly inspired by [6], to cope with the mixture model. If we assume that the DCT coefficients of the tampered and original areas have similar distributions, then the probability distribution of the observed coefficients for $x \neq 0$ can be modeled as

$$p(x; Q_1, \alpha) = \alpha \cdot n(x; Q_1) \cdot \tilde{h}(x) + (1 - \alpha) \cdot \tilde{h}(x) \tag{10}$$

where α is the mixture parameter and we have highlighted the dependence of both $p(x)$ and $n(x)$ from Q_1. Based on the above model, the actual value of Q_1 can be estimated as

$$\hat{Q}_1 = \arg\min_{Q_1} \sum_{x \neq 0} [h(x) - p(x; Q_1, \alpha_{opt})]^2 \tag{11}$$

where, for each Q_1, α_{opt} is the optimal mixture parameter in the least square sense and is given by

$$\alpha_{opt} = -\frac{\sum_{x \neq 0} \tilde{h}(x)[n(x; Q_1) - 1] \cdot [\tilde{h}(x) - h(x)]}{\sum_{x \neq 0} \tilde{h}(x)^2 [n(x; Q_1) - 1]^2}. \tag{12}$$

Since Q_1 is a discrete parameter with a limited set of possible values, the minimization in (11) can be solved iteratively by trying every possible Q_1 and using the corresponding α_{opt}. In order to estimate the complete quantization matrix, the above minimization problem is separately solved for each of the 64 DCT coefficients within a block.

3.2. Effects of Rounding and Truncation Errors

After the first JPEG compression, pixel values are usually rounded to the nearest integer and truncated to eight bit values. This introduces rounding and truncation (R/T) errors between successive JPEG compressions, which are not taken into account by the function $n(x)$ as defined in (1). This fact may introduce severe inaccuracies in the estimation of the probability map. For example, when $Q_2 < Q_1$ the function $n(x)$ is zero for some values of x. If a single DCT coefficient in a block assumes one of these values, according to (9) the

[1] Note that the above probabilities are also conditional to x belonging to a specific period of the histogram.

block is assigned a probability equal to one of being tampered. However, due to R/T errors, even DCT coefficients belonging to original areas may assume values for which $n(x) = 0$, causing a large number of false alarms.

In order to overcome the above problems, we propose to use a slightly modified version of $n(x)$. Even if the exact behavior of a DQ histogram affected by R/T errors is difficult to model, we can assume that R/T errors will cause every bin of $h_0(x)$ to spread over the adjacent bins proportionally to σ_e, where σ_e is the standard deviation of the R/T error. Hence, after quantization by Q_2, such a spread will be proportional to σ_e/Q_2. As a first approximation, this can be modeled by convolving $n(x)$ with a kernel having standard deviation σ_e/Q_2. According to this model, (9) and (10) are modified by replacing $n(x)$ with

$$n'(x) = n(x) * g(x) \tag{13}$$

where $g(x)$ is a Gaussian kernel having standard deviation σ_e/Q_2.

Besides these effects, truncation often introduces a bias on R/T errors, which will affect the statistics of DC coefficients. If the bias is equal to b, the relationship between the unquantized DC coefficient u and the doubly quantized DC coefficient x becomes $x = \left\lfloor \left(\left\lfloor \frac{u}{Q_1} \right\rfloor Q_1 + b \right) \frac{1}{Q_2} \right\rfloor$. Hence, in the case of DC coefficients $n(x)$ must be conveniently modified by redefining $L(x) = Q_1 \left(\left\lceil \frac{Q_2}{Q_1} \left(x - \frac{b}{Q_2} - \frac{1}{2} \right) \right\rceil - \frac{1}{2} \right)$, $R(x) = Q_1 \left(\left\lfloor \frac{Q_2}{Q_1} \left(x - \frac{b}{Q_2} + \frac{1}{2} \right) \right\rfloor + \frac{1}{2} \right)$.

The exact values of both σ_e and b should be computed on the first compressed image, which is not available together with the tampered one. However, we found that in practice they can be well approximated by measuring the R/T error on the tampered image.

4. PERFORMANCE ANALYSIS

In this section we firstly describe the experimental methodology we followed in order to evaluate the performance of the improved forgery detector and to provide a proper comparison with the previous one; then, we show the experimental results coming from such an analysis.

The image dataset used for testing the algorithms is composed by 100 non-compressed TIFF images, having heterogeneous contents, coming from three different digital cameras (namely Nikon D90, Canon EOS 450D, Canon EOS 5D) and each acquired at its highest resolution; the central portion of size 1024×1024 is then extracted from each image to form the original dataset. Starting from this, we create the corresponding manipulated images following the SCF hypothesis. Namely, each original image is JPEG compressed with a given quality factor QF_1 (using the Matlab function `imwrite`); the central portion of size 256×256 is then replaced with the corresponding area from the original TIFF image; finally, the overall "manipulated" image is JPEG compressed (again with Matlab) with another given quality factor QF_2. In this way, the image will result doubly compressed everywhere, except in the central region where it is supposed to be forged. Both QF_1 and QF_2 are taken from the set $[50, 60, \ldots 100]$ achieving 36 possible combinations of (QF_1, QF_2) for each of the 100 tampered images.

The selection of a proper performance metric is fundamental at this stage. Both the considered algorithms provide as output, for each analyzed image, a probability map that represents the probability of each 8×8 block to be forged (i.e. for each 1024×1024 image a 128×128 probability map is given). After a thresholding step, a binary detection map is achieved, that locates which are the blocks detected as tampered. We now assume to know for each analyzed image the position of forged areas; since in our experimental tests a central 256×256 region is modified to be singly compressed, it is possible to associate to any manipulated image a corresponding 128×128 binary mask whose 32×32 central portion indicates forged blocks. A comparison between the algorithm output detection map and the known tampering mask will allow to estimate the error rates of the forensic schemes, measured as false alarm probability P_{fa} and missed detection probability P_{md}. These two probabilities can be computed by measuring the following parameters: n_{NMF}: number blocks not manipulated, but detected as forged; n_{MNF}: number of blocks manipulated, but not detected as forged; n_{I}: number of blocks in the image (16384 in our tests); n_{M}: number of manipulated blocks (1024 in our tests). Starting from these figures, the error probabilities are given by:

$$P_{fa} = \frac{n_{\text{NMF}}}{n_{\text{I}} - n_{\text{M}}} \qquad P_{md} = \frac{n_{\text{MNF}}}{n_{\text{M}}}$$

and the correct detection probability is: $P_d = 1 - P_{md}$.

For depicting the tradeoff between the correct detection rate P_d and the false alarm rate P_{fa} we refer to the receiver operating characteristic (ROC) curve and consider it as an appropriate means for measuring the performances of the forgery detector. In particular, since the ROC curve is a two dimensional plot of P_d versus P_{fa} as the decision threshold of the detector is varied, we adopt the area under the ROC curve (AUC) in order to summarize the performance with a unique scalar value representing the general behavior of the detector. It is known that AUC should assume values between 0.5 and 1 for realistic and effective (i.e. no random) detectors.

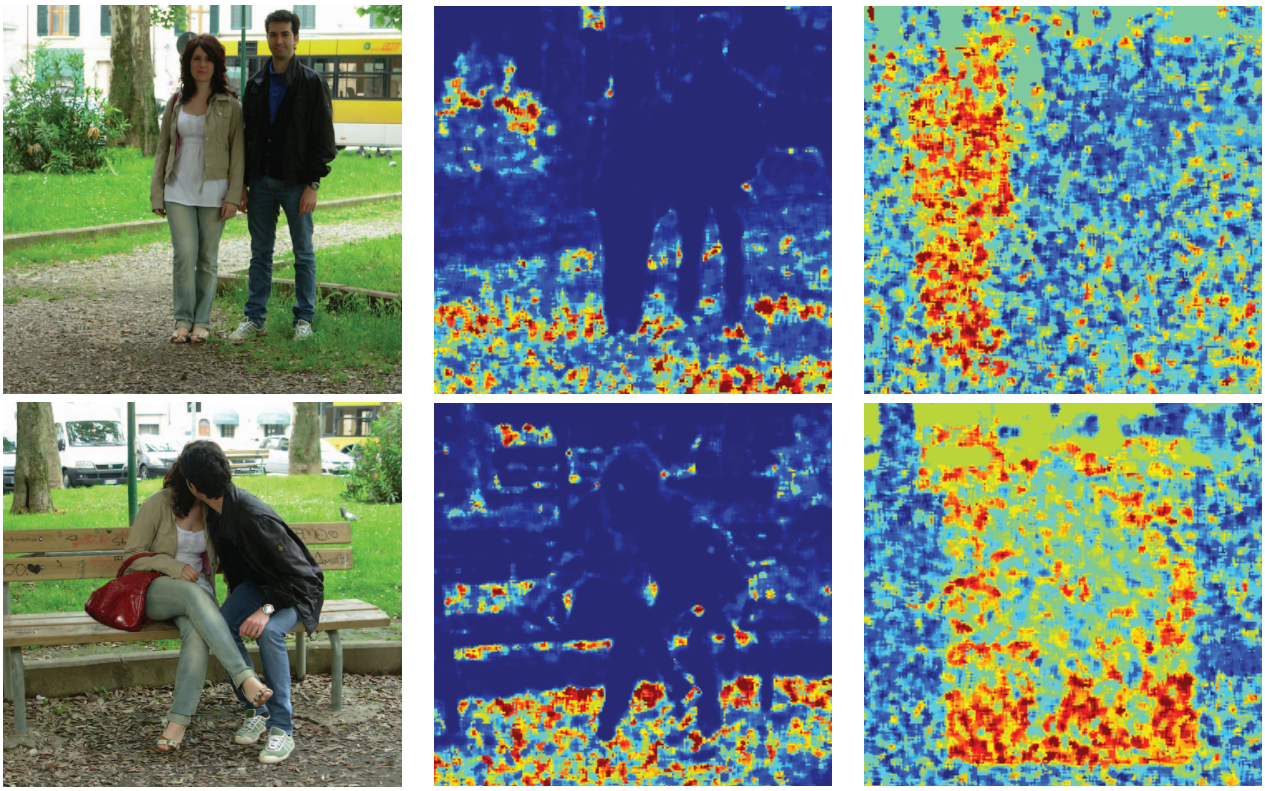
In practice, we moved the decision threshold from 0 to 1 with step 0.01 and for achieving the AUC corresponding to any combination of (QF_1, QF_2) we averaged the results of P_{fa} and P_d computed on all the 100 tampered images. The values for AUC achieved by the algorithm in [5] and by the improved method are shown in Tables 1 and 2 respectively. The best results for each combination (QF_1, QF_2) are highlighted in bold and demonstrate that the new probability map has an improved accuracy that helps in discriminating forged and unchanged regions. In particular, the proposed

		QF_2					
		50	60	70	80	90	100
QF_1	50	**0.50**	0.60	0.83	0.94	0.98	0.99
	60	0.49	**0.50**	0.67	0.86	0.98	0.98
	70	0.50	0.51	**0.50**	0.81	0.98	0.99
	80	0.51	0.49	0.52	**0.51**	0.95	0.99
	90	0.50	0.50	0.51	0.50	**0.51**	0.99
	100	**0.50**	0.50	**0.50**	**0.51**	**0.51**	**0.52**

Table 1. AUC achieved by the algorithm in [5].

		QF_2					
		50	60	70	80	90	100
QF_1	50	0.50	**0.90**	**1.00**	**1.00**	**0.99**	**0.99**
	60	**0.78**	0.50	**0.95**	**1.00**	**0.99**	**0.99**
	70	**0.77**	**0.83**	0.49	**1.00**	**0.99**	**0.99**
	80	**0.69**	**0.81**	**0.85**	0.49	**0.99**	**0.99**
	90	**0.58**	**0.63**	**0.70**	**0.78**	0.50	**0.99**
	100	0.50	**0.50**	0.50	0.50	0.50	0.50

Table 2. AUC achieved by the proposed algorithm.



(a) (b) (c)

Fig. 1. *Application to realistic forgeries: (a) images under analysis; (b) probability maps of [5]; (c) probability maps of the proposed algorithm. Red/blue areas correspond to high/low probability of being tampered. The proposed algorithm clearly shows an elongated area (above) and a rectangular region (below) having a high probability of being tampered, which are not visible in the maps of [5].*

method is able to reveal tampering even if $QF_2 < QF_1$, which often occurs in practical cases.

The algorithm has also been tested on a set of images representing realistic cases of forgery; in Figure 1 two examples of suspected images are shown: our probability map clearly reveals an area where a third person could have been removed and a rectangular region that can have been pasted from another image, whereas the probability maps of the previous method only provide information strongly dependent on the image content.

5. CONCLUDING REMARKS

In this paper, we presented an improved statistical test to discriminate between forged and original areas in JPEG images. The proposed test is based on the hypothesis that original areas will undergo a double compression while forged areas will only have single compression. New models for the probability distributions of DCT coefficients in the case of single and double compression are proposed, from which the probability for each 8×8 DCT block to be forged is derived. The validity of the improved system has been demonstrated by computing the ROC for the forgery detector based on thresholding the probability map. The AUC values obtained for different combinations of (QF_1, QF_2) have been compared with a previously proposed method [5], showing a better discriminating performance, especially when $QF_2 < QF_1$. Results are also confirmed by tests carried on realistic forgeries. Future research will focus on the automatic interpretation of the probability map, as well as on the combination of such a result with the output of other multimedia forensics tools.

6. REFERENCES

[1] H. Farid, "A survey of image forgery detection," *IEEE Signal Processing Mag.*, vol. 2, no. 26, pp. 16–25, 2009.

[2] W. Luo, Z. Qu, J. Huang, and G. Qui, "A novel method for detecting cropped and recompressed image block," in *Proc. of ICASSP 2007*, 2007, vol. 2, pp. II–217–II–220.

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

[4] M. Barni, A. Costanzo, and L. Sabatini, "Identification of cut & paste tampering by means of double-JPEG detection and image segmentation," in *Proc. of ISCAS 2010*, 2010.

[5] Z. Lina, J. He, X. Tang, and C.-K. Tang, "Fast, automatic and fine-grained tampered JPEG image detection via DCT coefficient analysis," *Pattern Recognition*, vol. 42, no. 11, pp. 2492–2501, Nov. 2009.

[6] J. Lukáš and J. Fridrich, "Estimation of primary quantization matrix in double compressed JPEG images," in *Digital Forensic Research Workshop*, 2003.