

# Analysis of Non-Aligned Double JPEG Artifacts for the Localization of Image Forgeries

T. Bianchi <sup>#\*1</sup>, A. Piva <sup>#2</sup>

<sup>#</sup> Dept. of Electronics and Telecommunications, University of Florence, Via S. Marta 3, 50139 Firenze, Italy

<sup>\*</sup> National Inter-University Consortium for Telecommunications, Via S. Marta 3, 50139 Firenze, Italy

<sup>1</sup> tiziano.bianchi@unifi.it

<sup>2</sup> alessandro.piva@unifi.it

**Abstract**—In this paper, we present a forensic algorithm to discriminate between original and forged regions in JPEG images, under the hypothesis that the tampered image presents a non-aligned double JPEG compression (NA-JPEG). Unlike previous approaches, the proposed algorithm does not need to manually select a suspect region to test the presence or the absence of NA-JPEG artifacts. Based on a new statistical model, the probability for each  $8 \times 8$  DCT block to be forged is automatically derived. Experimental results, considering different forensic scenarios, demonstrate the validity of the proposed approach.

## I. INTRODUCTION

The availability of easy-to-use image processing tools allowing to modify the content of digital images is today so large that the diffusion of fake contents through the digital world is becoming increasing and worrying. Such a possibility raises several problems in all the fields in which the credibility of images should be granted before using them as sources of information, like insurance, law and order, journalism, medical applications.

In the last years many image forensic techniques have been proposed as a means for revealing the presence of forgeries in digital images through the analysis of statistical and geometrical features, JPEG quantization artifacts, interpolation effects, demosaicing traces, feature inconsistencies, etc. [1].

Since the majority of digital images is stored in JPEG format, several forensic tools have been designed to detect the presence of tampering in this class of images. The forgery is revealed by analyzing some artifacts introduced by JPEG recompression occurring when the forged image is generated; in particular, such artifacts can be categorized into two classes, according to whether the second JPEG compression uses a DCT grid aligned with the first compression or not. The first case will be referred to as aligned double JPEG (A-DJPG) compression, whereas the second case will be referred to as non-aligned double JPEG (NA-DJPG) compression. Approaches belonging to the first category include [2], where the author proposes to detect areas which have undergone a double JPEG compression by recompressing the image at different quality levels and looking for the presence of so-called *ghosts*, and [3],[4], where double JPEG compression

is detected analyzing the statistics of blockwise DCT coefficients. The presence of non-aligned double JPEG compression has been investigated in [5],[6] and [7], that detect particular distortions in blocking artifacts, in [8], where the shift of the primary JPEG compression is determined via a demixing approach, and in [9], where the periodicity of blockwise DCT coefficients is studied.

However, the above algorithms rely on the hypothesis to know the right location of the forgery area, for example by applying a segmentation of the image under test before the forensic analysis [6], or they are just designed to decide if the whole image has been doubly JPEG compressed [5][7], so that the correct localization of the forgery in a tampered image is still an open issue. To the best of our knowledge, only some forensic algorithms designed to work in the presence of aligned double JPEG compression are able to localize a tampered area: in [3] and [4] two methods are proposed for the automatic localization of tampered regions with a fine-grained scale of  $8 \times 8$  blocks.

In this paper, we propose then the first forensic tool that, differently from previous works, can reveal a tampering at a local level, without any prior information about the location of the manipulated area, in the presence of non-aligned double JPEG compression. The output of the algorithm is a map that gives the probability, or the likelihood, for each  $8 \times 8$  image block to be tampered. The proposed algorithm can be applied in different forensic scenarios, in which either the presence or the absence of NA-JPEG artifacts at a local level can be interpreted as evidence of tampering.

## II. FORENSIC SCENARIOS

In order to correctly interpret the presence or the absence of artifacts due to double compression, in the following analysis we will consider two different scenarios.

A first scenario is that in which an original JPEG image, after some localized forgery, is saved again in JPEG format. We can assume that the forger disrupts the JPEG compression statistics in the tampered area: examples could be a cut and paste from either a non compressed image or a resized image, or the insertion of computer generated content. In this case, 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 DQ artifacts. In the following, we will refer to this case as the single compression forgery (SCF) hypothesis.

A second scenario is that of image splicing. In this kind of forgery, it is assumed that a region from a JPEG image is pasted onto a host image that does not exhibit JPEG compression statistics, and that the resulting image is JPEG recompressed. In this case, the forged region will exhibit double compression artifacts, whereas the non manipulated region will present no such artifacts. In the following, we will refer to this second case as the double compression forgery (DCF) hypothesis.

Under the SCF hypothesis, NA-DJPG artifacts will be present if the original image is randomly cropped before being recompressed in JPEG format. Under the DCF hypothesis, assuming that the forged region is randomly pasted in the new image, there is a probability of 63/64 that the  $8 \times 8$  block grids of the host image and of the pasted region will be misaligned, and thus that the forged region will exhibit NA-DJPG artifacts.

### III. SINGLE AND DOUBLE JPEG COMPRESSION MODELS

In this section, we will describe the statistical model used to characterize NA-DJPG artifacts. We will also introduce some simplifications that will be useful in defining the proposed detection algorithm, as well as some modifications needed to take into account the effects of rounding and truncation errors between the first compression and the second compression.

#### A. JPEG Compression Model

The JPEG compression algorithm can be modeled by three basic steps [10]:  $8 \times 8$  block DCT of the image pixels, uniform quantization of DCT coefficients with a quantization matrix whose values depend on a quality factor  $QF$ , entropy encoding of the quantized values. The image resulting from decompression will be obtained by the inverse of each step in reverse order: entropy decoding, dequantization, inverse block DCT. In the following analysis, we will consider that quantization is achieved by dividing each DCT coefficient by a proper quantization step  $Q$  and rounding the result to the nearest integer, whereas dequantization is achieved by simply multiplying by  $Q$ .

Let us then assume that an original uncompressed image  $\mathbf{I}$  is JPEG compressed with a quality factor  $QF$ , and then decompressed. Since entropy encoding is perfectly reversible, the image obtained after JPEG decompression can be modeled as follows:

$$\mathbf{I}_1 = \mathbf{D}_{00}^{-1} \mathcal{D}(\mathcal{Q}(\mathbf{D}_{00}\mathbf{I})) + \mathbf{E}_1 = \mathbf{I} + \mathbf{R}_1. \quad (1)$$

In the above equation,  $\mathbf{D}_{00}$  models an  $8 \times 8$  block DCT with the grid aligned with the upper left corner of the image,  $\mathcal{Q}(\cdot)$  and  $\mathcal{D}(\cdot)$  model quantization and dequantization processes, respectively, and  $\mathbf{E}_1$  is the error introduced by rounding and truncating the output values to eight bit integers. The last quantity  $\mathbf{R}_1$  can be thought of as the overall approximation error introduced by JPEG compression with respect to the original image. In the above chain, if we neglect rounding/truncation

(R/T) errors, the only operation causing a loss of information is the quantization process  $\mathcal{Q}(\cdot)$ .

Let us now analyse the artifacts that appear in the presence of a double non-aligned JPEG compression, due to the interaction of successive quantization and dequantization processes.

#### B. NA-DJPG Compression

In the case of NA-DJPG compression, we can assume that the original image  $\mathbf{I}$  has been JPEG compressed with a quality factor  $QF_1$  using a DCT grid shifted by  $(r, c)$ ,  $0 \leq r \leq 7$  and  $0 \leq c \leq 7$ , with respect to the upper left corner, so that the image obtained after JPEG decompression can be represented as:

$$\mathbf{I}_1 = \mathbf{D}_{rc}^{-1} \mathcal{D}_1(\mathcal{Q}_1(\mathbf{D}_{rc}\mathbf{I})) + \mathbf{E}_1 \quad (2)$$

where  $\mathbf{D}_{rc}\mathbf{I}$  are the unquantized DCT coefficients of  $\mathbf{I}$  and  $\mathcal{Q}_1$ ,  $\mathcal{D}_1$  denote that a proper quantization matrix corresponding to the quality  $QF_1$  was used.

We then assume that the image has been again JPEG compressed with a quality factor  $QF_2$ , but now with the block grid aligned with the upper left corner of the image. If we consider the DCT coefficients of the second compression after entropy decoding, no noticeable artifacts are present. However, if we consider the image after the second decompression, i.e.,  $\mathbf{I}_2 = \mathbf{I}_1 + \mathbf{R}_2$ , and we apply a block DCT with alignment  $(r, c)$ , we have

$$\mathbf{D}_{rc}\mathbf{I}_2 = \mathcal{D}_1(\mathcal{Q}_1(\mathbf{D}_{rc}\mathbf{I})) + \mathbf{D}_{rc}(\mathbf{E}_1 + \mathbf{R}_2). \quad (3)$$

Since the JPEG standard uses 64 different quantization steps, one for each of the 64 frequencies within a  $8 \times 8$  DCT, the DCT coefficients will be distributed according to 64 different probability distributions. According to the above equation, each unquantized DCT coefficient obtained by applying to the doubly compressed image  $\mathbf{I}_2$  a block DCT with alignment  $(r, c)$  (i.e. the same alignment of the first compression) will be distributed as

$$p_Q(x; Q_1) = p_1(x) * g_Q(x) \quad (4)$$

where  $Q_1$  is the quantization step of the first compression,  $g_Q(x)$  models the distribution of the overall approximation error, i.e. the term  $\mathbf{D}_{rc}(\mathbf{E}_1 + \mathbf{R}_2)$ ,  $*$  models convolution, and

$$p_1(v) = \begin{cases} \sum_{u=v-Q_1/2}^{v+Q_1/2} p_0(u) & v = kQ_1 \\ 0 & \text{elsewhere} \end{cases} \quad (5)$$

models the distribution of the DCT coefficients after quantization by  $Q_1$  and dequantization, being  $p_0(u)$  the distribution of the original unquantized coefficients.

If we model the approximation error as the sum of the R/T error in the DCT domain plus the quantization error due to uniform quantization with quantization step  $Q_2$ , by invoking the central limit theorem we can assume that the R/T error is Gaussian distributed with mean  $\mu_e$  and variance  $\sigma_e^2$ , and thus the approximation error is Gaussian distributed with mean  $\mu_e$  and variance  $\sigma_e^2 + Q_2^2/12$ , i.e.,

$$g_Q(x) = \frac{1}{\sqrt{2\pi(\sigma_e^2 + Q_2^2/12)}} e^{-(x-\mu_e)^2/(\sigma_e^2 + Q_2^2/12)} \quad (6)$$

In the absence of NA-DJPG compression, that is if the image did not undergo a first JPEG compression with alignment  $(r, c)$ , the unquantized DCT coefficients obtained by applying a shifted block DCT can be assumed distributed approximately as the original unquantized coefficients, that is

$$p_{\text{NQ}}(x) = p_0(x) \quad (7)$$

since a misalignment of the DCT grids usually destroys the effects of quantization [11].

### C. Simplified Model

Although the model in (4) is quite accurate, it requires the knowledge of the distribution of the unquantized coefficients  $p_0(u)$ , which may not be available in practice. However, it is possible to make some simplifications in order to obtain a model less dependent from the image content.

Indeed, if we can assume that the histogram of the original DCT coefficients is locally uniform, that is  $p_0(u)$  is smooth, we can simplify

$$p_1(x) \approx \begin{cases} Q_1 p_0(x) & x = kQ_1 \\ 0 & \text{elsewhere} \end{cases} \quad (8)$$

Hence, if we assume that the JPEG approximation error due to the last compression is smaller than  $Q_1$ , and thanks to (7), we have that (4) can be simplified to

$$p_Q(x; Q_1) \approx n_Q(x) \cdot p_{\text{NQ}}(x), \quad x \neq 0. \quad (9)$$

where  $n_Q(x) = n_{Q,0}(x) * g_Q(x)$  and

$$n_{Q,0}(x) \triangleq \begin{cases} Q_1 & x = kQ_1 \\ 0 & \text{elsewhere} \end{cases} \quad (10)$$

In Fig. 1 the models proposed in (4), (9), and (7) are compared with the histograms of unquantized DCT coefficients of a NA-DJPG compressed and a singly compressed image: in both cases there is a good agreement between the proposed models and the real distributions.

## IV. FORGERY LOCALIZATION ALGORITHM

In the following, we will assume that for each DCT coefficient  $x$  of an image, we know both the probability distributions of  $x$  conditional to the hypothesis of being tampered, i.e.,  $p(x|\mathcal{H}_1)$ , and the probability distributions of  $x$  conditional to the hypothesis of not being tampered, i.e.,  $p(x|\mathcal{H}_0)$ .

The above conditional distributions are given by (4) and (7), according to whether we are considering the SCF or the DCF hypothesis. For example, under the DCF hypothesis we have  $p(x|\mathcal{H}_1) = p_Q(x; Q_1)$  and  $p(x|\mathcal{H}_0) = p_{\text{NQ}}(x)$ . In the following, for the sake of simplicity, we will always assume the DCF hypothesis, i.e.  $p(x|\mathcal{H}_0)$  denotes the distribution of singly compressed coefficients, and  $p(x|\mathcal{H}_1)$  is the distribution of doubly compressed coefficients.

Given  $p(x|\mathcal{H}_1)$  and  $p(x|\mathcal{H}_0)$ , a DCT coefficient  $x$  can be classified as belonging to one of the two models according to the value of the likelihood ratio

$$\mathcal{L}(x) = \frac{p(x|\mathcal{H}_1)}{p(x|\mathcal{H}_0)}. \quad (11)$$

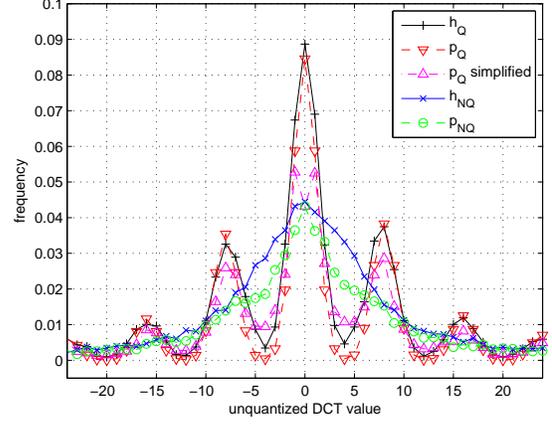


Fig. 1. Example of NA-DJPG compression model:  $h_Q$  and  $h_{\text{NQ}}$  denote the histograms of unquantized DCT coefficients of a NA-DJPG compressed and a singly compressed image, respectively. The distributions obtained according to equations (4), (9), and (7) are in good agreement with these data.

If multiple DCT coefficients within the same  $8 \times 8$  block are considered, by assuming that they are independently distributed we can express the likelihood ratio corresponding to the block at position  $(i, j)$  as

$$\mathcal{L}(i, j) = \prod_k \mathcal{L}(x_k(i, j)) \quad (12)$$

where  $x_k(i, j)$  denotes the  $k$ th DCT coefficient within the block at position  $(i, j)$ <sup>1</sup>. Such values form a likelihood map of the JPEG image with resolution  $8 \times 8$  pixel, which can be used to localize possibly forged regions within the image.

By using the simplified model described in Section III-C, it is possible to approximate the likelihood ratio as either  $\mathcal{L}(x) = 1/n_Q(x)$  (in case of SCF hypothesis) or  $\mathcal{L}(x) = n_Q(x)$  (in case of DCF hypothesis). The likelihood map obtained using such simplifications can be expressed as

$$\mathcal{L}(i, j) \approx \prod_k n_Q(x_k(i, j))^b \quad (13)$$

where  $b = -1$  (SCF) or  $b = 1$  (DCF), and depends only on compression parameters, i.e.,  $Q_1$ ,  $Q_2$ , having removed any dependencies from the image content. Hence, even if approximated, the adoption of the simplified models can lead to a more robust localization of possibly forged regions.

### A. Estimation of Model Parameters

The models described in Section III require the estimation of some parameters in order to be applied in practice. Among these parameters,  $p_0(u)$ ,  $Q_2$ ,  $\mu_e$ , and  $\sigma_e$  are common to both  $p(x|\mathcal{H}_1)$  and  $p(x|\mathcal{H}_0)$ , whereas  $Q_1$  is required only to characterize the distribution of doubly quantized coefficients. Moreover, we should determine the shift  $(r, c)$  between the

<sup>1</sup>With a slight abuse of notation, we use the same symbol  $\mathcal{L}(x)$  even if for different  $k$  we have different likelihood functions. The same convention is used in (13) when referring to  $n(x)$ .

first compression and the last compression in order to compute the unquantized DCT coefficients as in (3).

As to  $Q_2$ , we will assume that it is available from the JPEG image header. As to the shift  $(r, c)$ , we will assume that it has already been estimated, e.g. using the methods described in [9][8]. As to the other parameters, they are estimated according to the following procedures.

1) *Estimation of  $Q_1$* : The estimation of the quantization step of the primary compression is crucial for the correct modeling of doubly compressed regions. When dealing with a possibly forged image, usually there is no prior knowledge regarding the location of such regions. An image block could include an original area, as well as a tampered one. Thus, the distribution of the DCT coefficients of a tampered image can be modeled as a mixture of  $p(x|\mathcal{H}_1)$  and  $p(x|\mathcal{H}_0)$ , i.e.,

$$p(x; Q_1, \alpha) = \alpha \cdot p(x|\mathcal{H}_0) + (1 - \alpha) \cdot p(x|\mathcal{H}_1; Q_1) \quad (14)$$

where  $\alpha$  is the mixture parameter and we have highlighted the dependence of  $p(x|\mathcal{H}_1)$  from  $Q_1$ . Based on the above model, the maximum likelihood estimate of  $Q_1$  can be obtained as

$$\hat{Q}_1 = \arg \max_{Q_1} \sum_x \log[\alpha_{opt} p(x|\mathcal{H}_0) + (1 - \alpha_{opt}) p(x|\mathcal{H}_1; Q_1)] \quad (15)$$

where  $\alpha_{opt}$  is the optimal mixture parameter. For each  $Q_1$ , the optimal mixture parameter can be estimated using an expectation-maximization (EM) algorithm.

Since  $Q_1$  is a discrete parameter with a limited set of possible values, the minimization in (15) can be solved iteratively by trying every possible  $Q_1$  and using the corresponding  $\alpha_{opt}$ . 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.

2) *Estimation of  $p_0(u)$* : Following the observations in [11], we propose to approximate the distribution of the unquantized DCT coefficients using the histogram of the DCT coefficients of the decompressed image computed after the DCT grid is suitably shifted with respect to the upper left corner. In particular, we will use a shift of  $\pm 1$  with respect to the estimated shift  $(r, c)$  of the primary compression, where the sign of the increment is chosen so as to keep the shift values between 0 and 7 and to avoid the case  $(0, 0)$ .

3) *Estimation of  $\mu_e$  and  $\sigma_e$* : The true values of both  $\mu_e$  and  $\sigma_e$  should be estimated by relying on the primary JPEG compression, which in general is not available when observing the tampered image. In practice, we found that they can be well approximated by measuring the R/T error on the tampered image. The rationale is that both  $\mu_e$  and  $\sigma_e$  are mainly determined by the coarse-grained statistics of the image content, which usually are little affected by tampering.

By looking at equation (1), given as input the quantized DCT coefficients of the observed image  $\mathbf{C}_2$  we can think to compute the term  $\mathbf{E}_2$  by reconstructing the image with infinite precision as  $\mathbf{D}_{00}^{-1} \mathcal{D}(\mathbf{C}_2)$ , which can be approximated by using floating point arithmetic, and taking the difference with the image  $\mathbf{I}_2$  which is obtained by rounding and truncating to 8 bit precision the floating point values.

For the experimental validation of the proposed work, we have built an image dataset 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; each test has been performed by cropping a central portion with size  $1031 \times 1031$ : this choice allows us to still have a  $1024 \times 1024$  image after randomly cropping a number or rows and columns between 0 and 7.

Starting from this dataset, we have created manipulated images exhibiting NA-DJPG artifacts, following both SCF and DCF hypotheses. As to the NA-DJPG SCF dataset, each original image is JPEG compressed with a given quality factor  $QF_1$  (using the Matlab function `imwrite`); the image is randomly cropped by removing a number of rows and columns between 0 and 7; the central portion of size  $256 \times 256$  is replaced with the corresponding area from the original TIFF image; finally, the overall “manipulated” image is JPEG compressed with another given quality factor  $QF_2$ . In this way, the image will result NA-DJPG compressed everywhere, except in the central region where it is supposed to be forged.

The creation of the DCF datasets is dual with respect to the above procedure. Each original image is JPEG compressed with a given quality factor  $QF_1$ ; the central portion of size  $256 \times 256$  is cut with a random shift with respect to the JPEG grid and pasted onto the TIFF image so as to respect both the alignment of the image content and the alignment with the DCT grid; finally, the overall “manipulated” image is JPEG compressed with another given quality factor  $QF_2$ . In this way, the central region of the image, which is supposed to be forged, will result NA-DJPG compressed.

In all the above datasets,  $QF_1$  and  $QF_2$  are taken from the sets  $[50, 60, \dots, 90]$  and  $[50, 60, \dots, 100]$ , respectively, achieving 30 possible combinations of  $(QF_1, QF_2)$  for each of the 100 tampered images.

The selection of a proper performance metric is fundamental for evaluating the performance of the method. Our algorithm provides as output, for each analyzed image, a map that represents the likelihood of each  $8 \times 8$  block to be forged. After a thresholding step, a binary detection map is achieved, that locates which are the blocks detected as tampered. By assuming to have for each analyzed image the corresponding binary mask whose  $32 \times 32$  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 scheme, measured as false alarm probability  $P_{fa}$  and missed detection probability  $P_{md}$ . These two probabilities can be computed by measuring the following parameters:  $n_{\text{NMF}}$ : number of blocks not manipulated, but detected as forged;  $n_{\text{MNF}}$ : number of blocks manipulated, but not detected as forged;  $n_1$ : 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

TABLE I  
AUC ACHIEVED BY THE PROPOSED ALGORITHM USING THE STANDARD MODEL UNDER THE SCF HYPOTHESIS.

		$QF_2$					
		50	60	70	80	90	100
$QF_1$	50	0.58	0.79	<b>0.95</b>	<b>0.99</b>	0.99	0.99
	60	0.51	0.61	0.87	<b>0.98</b>	0.99	0.99
	70	0.48	0.50	0.62	0.92	0.98	0.99
	80	0.48	0.48	0.49	0.61	0.95	0.99
	90	0.48	0.48	0.48	0.48	0.55	0.98

TABLE II  
AUC ACHIEVED BY THE PROPOSED ALGORITHM USING THE SIMPLIFIED MODEL THE UNDER SCF HYPOTHESIS.

		$QF_2$					
		50	60	70	80	90	100
$QF_1$	50	<b>0.71</b>	<b>0.85</b>	0.94	0.98	<b>0.99</b>	<b>0.99</b>
	60	<b>0.59</b>	<b>0.71</b>	<b>0.89</b>	0.97	<b>0.99</b>	<b>1.00</b>
	70	<b>0.52</b>	<b>0.56</b>	<b>0.71</b>	<b>0.94</b>	<b>0.99</b>	<b>1.00</b>
	80	<b>0.52</b>	<b>0.52</b>	<b>0.54</b>	<b>0.69</b>	<b>0.97</b>	<b>0.99</b>
	90	<b>0.51</b>	<b>0.51</b>	<b>0.52</b>	<b>0.51</b>	<b>0.57</b>	<b>0.99</b>

are given by:

$$P_{fa} = \frac{n_{\text{NMF}}}{n_I - n_M} \quad P_{md} = \frac{n_{\text{MNF}}}{n_M}$$

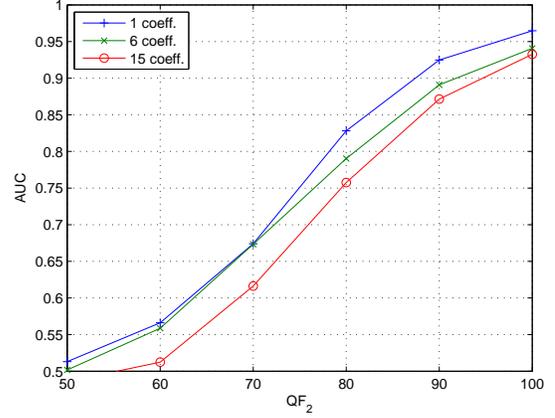
and the correct detection probability is:  $P_d = 1 - P_{md}$ .

For depicting the tradeoff between the correct detection rate  $P_d$  and the false alarm rate  $P_{fa}$  the receiver operating characteristic (ROC) curve is considered. 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 of the detector with a unique scalar value.

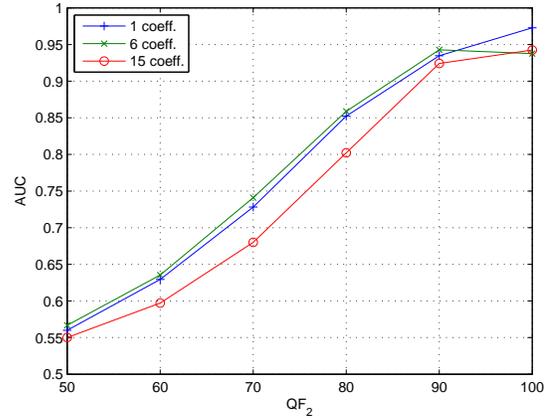
In the following, we will compare the AUC values obtained using the standard map in (12) and the simplified map in (13): to the best of our knowledge, these are the first methods that permit to localize possibly forged areas by relying on non-aligned double JPEG compression, so other methods can not be compared with our schemes. In all cases, likelihood maps are obtained by cumulating different numbers of DCT coefficients for each block, starting from the DC coefficient and scanning the coefficients in zig-zag order.

The AUC values achieved for different  $QF_2$  under the SCF hypothesis are shown in Fig. 2: when  $QF_2$  is sufficiently high ( $> 80$ ), NA-DJPG artifacts can be effectively used to localize traces of tampering. When comparing the two approaches, the simplified map appears more robust than the standard map for lower  $QF_2$  values. As to the effects of the cumulation of different DCT coefficients, the best results are obtained by considering the first 6 coefficients with the simplified map: when considering a higher number of coefficients the AUC values decrease, suggesting that NA-DJPG artifacts can not be reliably detected at the higher frequencies.

In order to assess the effects of different  $QF_1$  values, the AUC values obtained for different combinations of  $(QF_1, QF_2)$ , using the first 6 DCT coefficients to compute the likelihood map, are reported in Tables I-II. For ease of



(a)



(b)

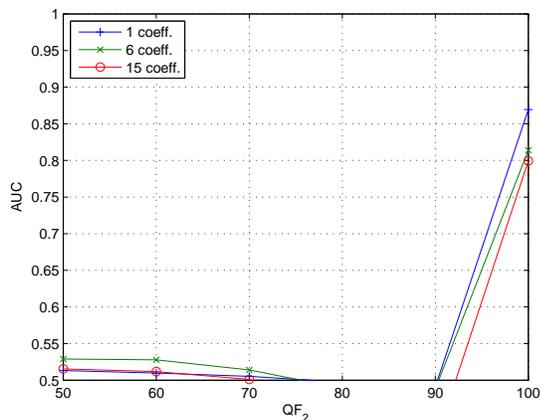
Fig. 2. AUC achieved for different  $QF_2$  using different numbers of DCT coefficients in the SCF scenario: (a) proposed algorithm with standard map; (b) proposed algorithm with simplified map.

reading, for each combination of  $(QF_1, QF_2)$  the highest AUC value between the two considered approaches is highlighted in bold. In this case, the simplified map achieves always the best performance except in three cases. Noticeably, it is not possible to achieve AUC values significantly greater than 0.5 when  $QF_2 < QF_1$ . However, it suffices  $QF_2 - QF_1 \geq 10$  to achieve an AUC value very close to one, which means that in this case forged areas can be localized with great accuracy.

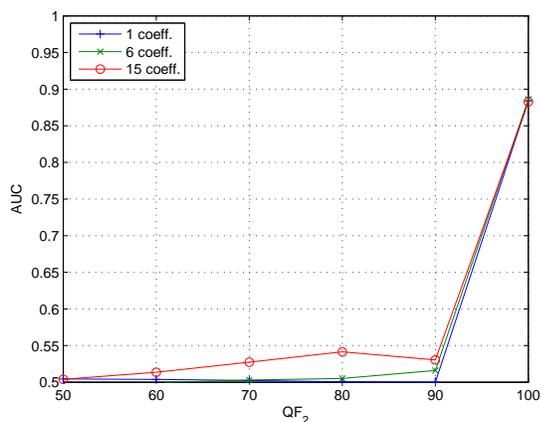
In Fig. 3, we provide the AUC values under the DCF hypothesis. In this case the performance of forgery localization is much lower than under the SCF hypothesis, allowing to localize traces of double compression only when  $QF_2$  is very high ( $> 90$ ).

#### A. Examples

The algorithm has also been tested on a set of images representing realistic cases of forgery; in Figure 4 an example of a tampered image is shown: the likelihood map clearly reveals that the pyramid is a tampered object, and it also shows some false alarms in the background, due to the low intensity variance in this area that does not allow a correct estimation



(a)



(b)

Fig. 3. AUC achieved for different  $QF_2$  using different numbers of DCT coefficients in the DCF scenario: (a) proposed algorithm with standard map; (b) proposed algorithm with simplified map.

of the statistics of DCT coefficients. However, these errors can be eliminated by a proper postprocessing on the map, before a final decision is made.

## VI. CONCLUSIONS

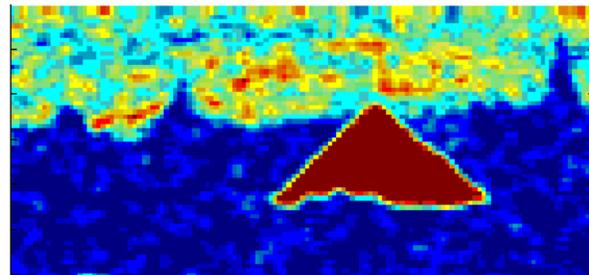
In this work, a forensic algorithm to detect and localize a tampered area into a digital image by exploiting the presence of non-aligned double JPEG compression artifacts has been proposed. The method is based on the derivation of a statistical model of DCT coefficients of NA-JPEG images, which is used, following an approach similar to the one proposed in [3] and in [4], for the generation of a likelihood map that shows the probability of each  $8 \times 8$  block of being tampered. The validity of the proposed system has been demonstrated by computing the ROC curves for the forgery detector based on thresholding the probability map. Results are also confirmed by tests carried on realistic forgeries.

## ACKNOWLEDGMENT

This work was partially supported by the REWIND Project funded by the Future and Emerging Technologies (FET)



(a)



(b)

Fig. 4. Application to realistic forgeries: (a) image under analysis; (b) likelihood map obtained using the simplified model. Red/blue areas correspond to high/low probability of being tampered. The proposed algorithm clearly shows that there is a high probability of the pyramid to be tampered.

programme within the 7FP of the European Commission, under FET-Open grant number: 268478, and by the project LivingKnowledge - Facts, Opinions and Bias in Time funded by the European Commission under contract no. 231126.

## REFERENCES

- [1] H. Farid, "A survey of image forgery detection," *IEEE Signal Processing Mag.*, vol. 2, no. 26, pp. 16–25, 2009.
- [2] —, "Exposing digital forgeries from JPEG ghosts," *IEEE Trans. Inf. Forensics Security*, vol. 4, no. 1, pp. 154–160, Mar. 2009.
- [3] Z. Lin, 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.
- [4] T. Bianchi, A. D. Rosa, and A. Piva, "Improved DCT coefficient analysis for forgery localization in JPEG images," in *Proc. of ICASSP 2011*, May 2011, pp. 2444–2447.
- [5] W. Luo, Z. Qu, J. Huang, and G. Qui, "A novel method for detecting cropped and recompressed image block," in *Proc. of ICASSP 2007*, vol. 2, 2007, pp. II-217–II-220.
- [6] 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, pp. 1687–1690.
- [7] Y.-L. Chen and C.-T. Hsu, "Detecting recompression of JPEG images via periodicity analysis of compression artifacts for tampering detection," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 2, pp. 396–406, June 2011.
- [8] Z. Qu, W. Luo, and J. Huang, "A convolutive mixing model for shifted double JPEG compression with application to passive image authentication," in *Proc. of ICASSP 2008*, 2008, pp. 1661–1664.
- [9] T. Bianchi and A. Piva, "Detection of non-aligned double JPEG compression with estimation of primary compression parameters," in *Proc. of ICIP 2011*, Sept. 2011.
- [10] A. C. Popescu and H. Farid, "Statistical tools for digital forensics," in *In 6th International Workshop on Information Hiding*. Springer-Verlag, Berlin-Heidelberg, 2004, pp. 128–147.
- [11] J. Lukáš and J. Fridrich, "Estimation of primary quantization matrix in double compressed JPEG images," in *Digital Forensic Research Workshop*, 2003.