CrossMark

Forensics for partially double compressed doctored JPEG images

Gurinder Singh [1] · Kulbir Singh [1] iD

Abstract Digital image forensics is required to investigate unethical use of doctored images by recovering the historic information of an image. Most of the cameras compress the image using JPEG standard. When this image is decompressed and recompressed with different quantization matrix, it becomes double compressed. Although in certain cases, e.g. after a cropping attack, the image can be recompressed with the same quantization matrix too. This JPEG double compression becomes an integral part of forgery creation. The detection and analysis of double compression in an image help the investigator to find the authenticity of an image. In this paper, a two-stage technique is proposed to estimate the first quantization matrix or steps from the partial double compressed JPEG images. In the first stage of the proposed approach, the detection of the double compressed region through JPEG ghost technique is extended to the automatic isolation of the doubly compressed part from an image. The second stage analyzes the doubly compressed part to estimate the first quantization matrix or steps. In the latter stage, an optimized filtering scheme is also proposed to cope with the effects of the error. The results of proposed scheme are evaluated by considering partial double compressed images based on the two different datasets. The partial double compressed datasets have not been considered in the previous state-of-the-art approaches. The first stage of the proposed scheme provides an average percentage accuracy of 95.45%. The second stage provides an error less than 1.5% for the first 10 DCT coefficients, hence, outperforming the existing techniques. The experimental results consider the partial double compressed images in which the recompression is done with different quantization matrix.

Keywords Image forensics · Partially double compressed · First quantization matrix · DCT coefficient histogram · JPEG

✉ Kulbir Singh
  ksingh@thapar.edu

  Gurinder Singh
  gurinder.singh@thapar.edu

[1] Department of Electronics and Communication Engineering, Thapar University, Patiala, Punjab, India

Springer

1 Introduction

The reliability of images is significant in numerous regions, such as scientific examination, criminal examination, observation frameworks, knowledge administrations, medicinal imaging and news coverage. Nowadays, in the computerized scenario, the information related to an image can be easily manipulated [29]. Digital image forensics is used to find out if an image has undergone any kind of forgery or not.

The digital investigation of an image provides the historic information of an image related to the camera or device properties used for shooting. Based on this information, one can find that the considered image is a doctored or not [29]. The image is compressed once during a shooting in the camera device, and if this image is further decompressed and resaved with different compression quality, the resultant image becomes double compressed. The first quantization matrix is always lost from the image metadata in the case of double compressed images. Therefore, estimation of the first quantization matrix from double compressed images is a challenging task for the researchers due to its importance in the digital investigation [20].

In an advanced examination, various acquisition devices and systems generally utilize the JPEG compression for images [7]. The issues faced during image examination can be divided into two broad categories [16], the first regarding the legitimacy of the visual document, and the other concerning the identification of the device which is used for the image acquisition.

Numerous methodologies are available to discover the JPEG images manipulations [30] and [28]. Different methodologies, as portrayed in [1] and [25], deal with the analysis of the statistical distribution of the DCT coefficient values. It is conceivable to figure out whether an image is doubly compressed or not by analyzing the related histogram [20] and [23]. The probability of each block of an image is calculated by a technique [10] which is being subjected to double compression.

To find out the difference in the images, a Markov random process is utilized in [4] by exploiting the SVM. The JPEG ghost technique [8] is used to detect the double compressed regions in an image and in [31] a periodic function detection strategy is used to identify the altered regions. A scheme is recommended in [12] to analyze the double compression using the same quantization matrix. Based on the estimated primary quantization matrix, a new technique is opined in [27] for the detection of double compressed JPEG images. Furthermore, a technique is depicted in [32] to detect the double JPEG compression by utilizing the same quantization matrix. In [33], a technique is proposed to detect the in-camera JPEG compression for double compressed images. Moreover, a robust technique is presented in [13] to detect the copy-move forgery in images under high JPEG compression artifacts. A novel technique for image analysis based on three visualization learning strategies is revealed in [34] to identify double JPEG compression clues. In the case of a duplicate operation, the image recompression by utilizing same quantization table is distinguished in [19], concluding that the grid of the pasted part has a significant probability that it is not aligned with the existing one. A new technique based on the compression artifacts presented in [3] focuses on discovering the traces caused by recompression.

The recreation of the first quantization matrix utilized by the acquisition device which is lost during double compression is important in all issues expressed previously. The model of the camera devices utilizing the same quantization tables is identified by estimating some part of the first quantization matrix. For the estimation of the first quantization matrix, in [20] the author exposes a few thoughts taking into account the conduct of normalized histograms.

The works in [35] and [2] additionally point out the image falsifications but do not provide a proper estimation of the first quantization matrix. In a splicing or a cloning operation, the

DCT grid portion of an image pasted during duplication can or cannot be aligned to the original image which results in the occurrence of aligned and non-aligned traces. The first quantization matrix is essential among the parameters that are required to model the double compressed regions and accurately identify likelihood map [2]. A DCT coefficients histogram analysis is utilized in [24] to estimate the first quantization matrix from the double compressed JPEG image. In [9], a technique is proposed to estimate the first quantization matrix from double compressed JPEG images that cope with the case when the first quantization step is larger than the second. In the recent work [36], a novel distributed lossless coding technique is suggested for the hyperspectral images. Furthermore, a hyperspectral image compression and reconstruction technique based on the multi-dimensional or tensor data processing approach is presented in [37]. The hyperspectral images involve Distributive Source Coding (DSC) for the lossless compression. The artifacts left during the compression can be analyzed for the detection of forgery introduced in these images.

The work in this paper contributes towards rectifying the following problems:

- All the existing techniques generally concentrate only on the detection of the double compressed region in an image. This provides a wide scope for the extraction of this double compressed part from the image.
- Secondly, all the techniques in the related work are based on the estimation of first quantization matrix from the double compressed JPEG images. Therefore, estimation of first quantization matrix from partial double compressed JPEG images can increase the level of forensic investigation.
- Also, most of the techniques ignore the effect of the error introduced during color conversions (YCbCr to RGB and vice versa), rounding and truncation of the values to eight-bit integers, etc. The first quantization matrix estimation accuracy can be improved by removing the effects of this error.

In this paper, a technique is proposed to estimate the first quantization matrix from partial double compressed JPEG images. In the first stage, a technique is proposed not only to detect but also to automatically isolate the doubly compressed region from the image. Therefore, this isolation of double compressed region solves the problem to estimate the first quantization matrix from partial double compressed JPEG images. The second stage analyzes the isolated doubly compressed region to estimate the first quantization matrix. In the latter stage, a filtering technique is proposed to optimize the performance of the algorithm by reducing errors. The proposed approach is solely dedicated to detect the region which is recompressed using a different quantization matrix.

The paper is organized as follows. In Section 2, the detailed background of JPEG compression along with its effects is discussed. It also includes the brief overview of the JPEG ghost detection technique. Section 3 covers the proposed scheme in detail. The experiment results are provided in Section 4 and conclusion is discussed in Section 5.

2 Background of JPEG compression

The first step in the JPEG compression is the Discrete Cosine Transform (DCT). The DCT of the image is performed by dividing the whole image into non-overlapping blocks of 8×8 blocks. The DCT is performed to segregate the high frequency components from the low frequency

components of the image. Then a quantization is applied by using the 8 × 8 quantization matrix integer value for each DCT coefficient. An error is generated in this phase known as quantization error. This error is the key reason of loss of information in JPEG compressed images. After this, it is transformed into a data stream with the utilization of classic entropy coding [21]. The process of image compression is backtracked conversely while achieving the JPEG decompression as shown in Fig. 1. If the same scheme is employed on JPEG image, with a different quantization matrix, the result would yield a double compressed image.

The quantization is followed by the rounding and truncating of real values to transform the range of integers into the range of [0; 255]. In this process, rounding values and truncation error are produced. In addition to these errors, another type of error occurs due to the conversion between RGB and YCbCr color spaces [21]. The error analysis can be performed on the 8-bit gray scale images. After double compression, the value of each double quantized coefficient c_{DQ} can be modeled as [8]:

$$c_{DQ} = \left[\left(\left[\frac{c}{q_1} \right] q_1 + e \right) \frac{1}{q_2} \right] \qquad (1)$$

where c denotes the single DCT coefficient value, q_1 and q_2 are the first and second quantization steps. The operator [.] indicates the rounding function, and e includes the error due to several operations like color space conversions, rounding, and truncation of the values to 8-bit integers, etc. To infer the value of q_1, a further compression is performed in a proper range with a novel quantization step q_3 and this error is computed by using an error function as follows [8]:

$$d_e(c, q_1, q_2, q_3) = \left| \left[\left[\left[\frac{c}{q_1} \right] \frac{q_1}{q_2} \right] \frac{q_2}{q_3} \right] q_3 - \left[\frac{c}{q_1} \right] \frac{q_1}{q_2} \right] q_2 \right| \qquad (2)$$

If double compression is carried out on an image, only the last quantization steps are accessible but the first ones cannot be accessed because they are lost. The Fig. 2c indicates the DCT histogram of an image after double quantization. The primary quantization is carried out by taking a quantization step $q_1 = 11$ and then acquired values of DCT coefficients are de-quantized with same the quantization step. Conclusively, the values are again quantized using a quantization step $q_2 = 7$. Apparently, the distribution of the doubly quantized values contains periodic empty bins. This happens because during the second quantization, the coefficient values are re-distributed into more bins than in the first quantization [20].

2.1 The JPEG ghost detection

The JPEG ghost detection technique [8, 38] has the capacity to localize the parts of an image which have gone through double compression. Consider a DCT coefficient c_1 which is quantized by an amount q_1. Then the resultant coefficient from first quantization is subsequently quantized second time by quantization step q_2 to result coefficient c_2. With the
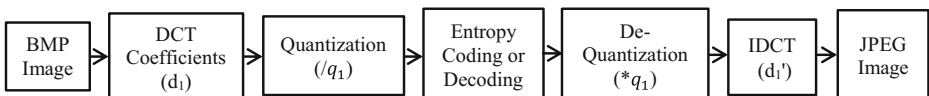

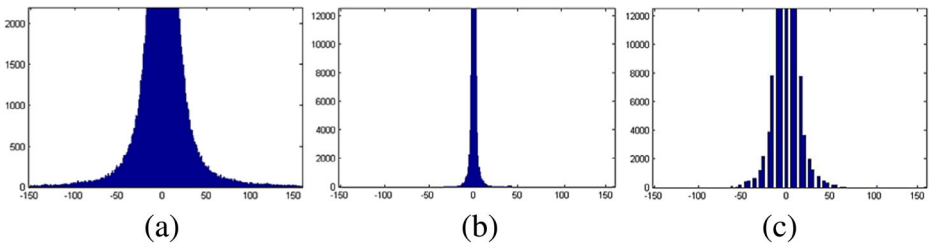Fig. 1 Single JPEG compression and de-compression

Fig. 2 (**a**) DCT histogram of an uncompressed image, (**b**) DCT histogram after the first compression with $q_1 = 11$, (**c**) DCT histogram after second compression with $q_2 = 7$

exception of $q_2 = 1$, the difference between c_1 and c_2 will be minimum when $q_2 = q_1$ and will increase as the difference between q_2 and q_1 increases. The JPEG ghost can be identified by considering each spatial frequency independently in each of the three luminance/color channels. However, various minima are possible when integer multiple quantization values are compared. The difference can be computed directly from the pixel values as follows:

$$d(x, y, q) = \frac{1}{3} \sum_{i=1}^{3} \left[f(x, y, i) - f_q(x, y, i) \right]^2 \tag{3}$$

where $f(x, y, i)$, $i = 1, 2, 3$, indicates each of three RGB color channels, and $f_q(.)$ is the result of compressing $f(.)$ at quality q. There is some disparity in the difference images due to the underlying image content within and outside of the tampered region which could possibly be confound a forensic analysis as shown in Fig. 3.

3 Proposed scheme

The estimation of first quantization matrix from the partial double compressed JPEG images has significant importance in the forensic investigation. Therefore, a two-stage scheme is proposed in this paper for the estimation of first quantization matrix from the partial double compressed JPEG images. In the first stage, a technique is proposed to the automatic isolation of double compressed region from an image, and then this region is analyzed in the second stage to
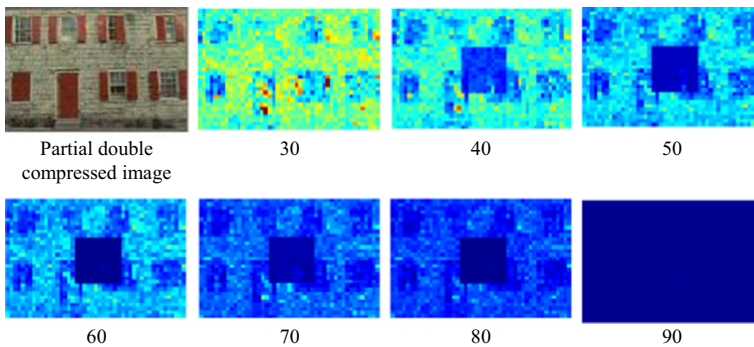

Fig. 3 Double compressed region detection through JPEG Ghost, along with the difference images corresponding to the different quality factor q

estimate the first quantization matrix. In the second stage, a filtering scheme is proposed to effectively reduce the effects of the error. The proposed scheme is for the scenarios when the image is recompressed with different quantization matrix than the first. The forgery introduced in an image through these partially double compressed JPEG images can also be analyzed.

In the proposed technique, the first stage as shown in Fig. 4 is the automatic isolation of double compressed region from an image by employing an enhanced JPEG Ghost detection technique. The conventional JPEG Ghost detection technique provides the gray scale difference image. The holes that occur in the gray scale difference image are then filled through morphology operation. Consequently, the adaptive median algorithm is applied after image complement which results in a binary image. The adaptive median algorithm classified the pixel values as noise by comparing each pixel value to its surrounding neighbored in the image. The pixel is considered as an impulse noise which is not structurally aligned with those pixels to which it is similar as well as which is different from a majority of its neighbors. These noise pixels are then replaced by the median pixel value of pixels in the neighborhood which have already passed the noise labeling test. The value of S_{max} (maximum allowed size of the neighborhood) is adjusted according to the intensity value of double compressed part in difference image to bring it in the reasonable range of $21 \leq S_{max} \leq 41$. The given doctored JPEG image is then masked with the binary image which provides the masked image. The desired part is then cropped from the masked image. The whole first stage of the proposed scheme is depicted in Figs. 5 to 8.
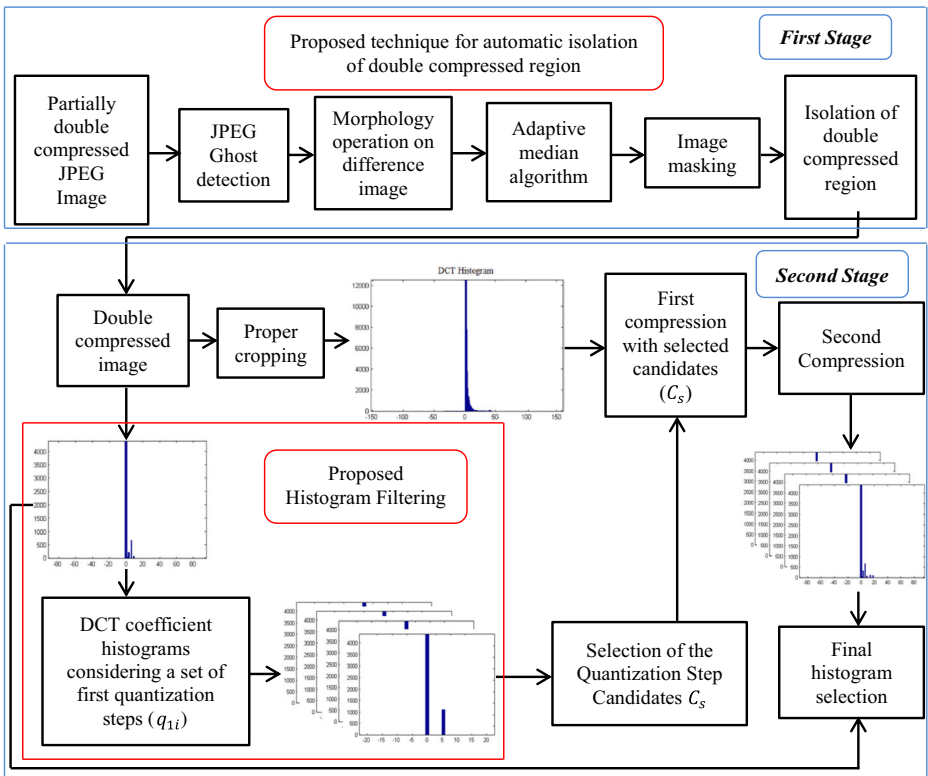

Fig. 4 Proposed scheme for the estimation of first quantization matrix from partial double compressed JPEG image
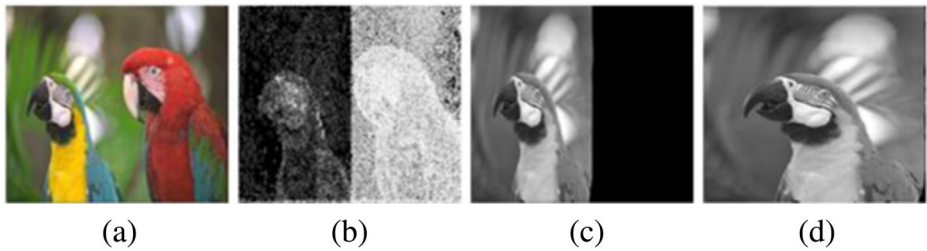
Fig. 5 (a) Partial double compressed image, Parrot (b) Difference image through JPEG ghost, (c) Image after morphology, adaptive filtering, and masking operation, (d) Isolated double compressed region of size 392 × 512 pixels

The second stage analyzes the double compressed region to estimate the first quantization matrix as shown in Fig. 4. The resultant image from the first stage is properly cropped in the second stage and then the DCT coefficients of the image are extracted. The reason behind this cropping operation is that the extracted double compressed part would show some irregularities at the edges. These irregularities can be observed in Figs. 5d, 6d, 7d and 8d. These irregularities are present only at the edges and not at any other region. Thus, the error transferred to the second step can be minimized by proper cropping of the extracted double compressed part at the edges as shown in Fig. 9. The merits of proper cropping have already been discussed in [15, 20].

Now the estimation of the first quantization matrix is carried out through the analysis of DCT coefficient histogram. The DCT histogram is filtered out through the proposed filtering scheme to reduce the effects of error e in (1). This filtering provides a set of filtered histograms. The rounding error e is modeled by approximating it as a Gaussian noise. This error proclaim itself by spreading peaks around the quantization step multiples. It will affect the second quantization step behavior along with the magnitude and statistics of the DCT coefficients. Due to this error two type of noises i.e. split and residual noise are encountered by the filtering strategy. The proposed histogram filtering scheme further reduces the effects of error e. By considering a set of first quantization steps in the range $q_{1i} \in \{q_{1min}, \ q_{1min} + 1, \ldots, q_{1max} \}$, several filtering operations are then performed.

The properties of successive quantizations are exploited more efficiently by function (4) as compared to the error function (2). Therefore, by exploiting the q_1 localization property of (4), limited first quantization candidates (C_s) are selected. The function d_{out} becomes nearly zero in
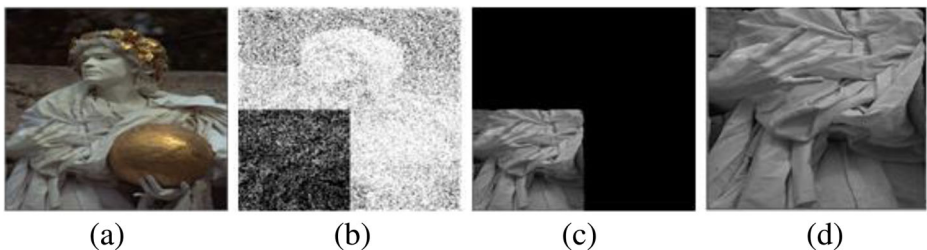

Fig. 6 (a) Partial double compressed image, Roman (b) Difference image through JPEG ghost, (c) Image after morphology, adaptive filtering, and masking operation, (d) Isolated double compressed region of size 260 × 385 pixels
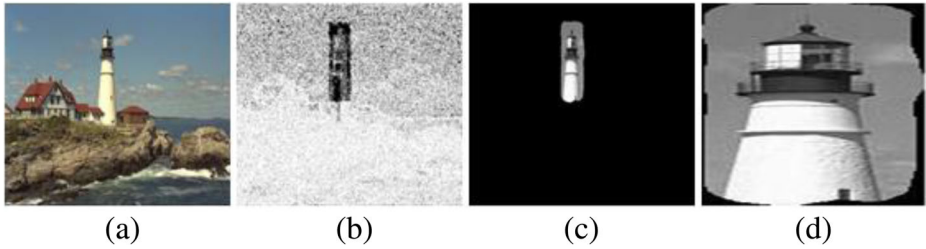
Fig. 7 (**a**) Partial double compressed image, Tower (**b**) Difference image through JPEG ghost, (**c**) Image after morphology, adaptive filtering, and masking operation, (**d**) Isolated double compressed region of size 82 × 205 pixels

the case $q_3 = q_{1i}$ as shown in Fig. 10, when the filtering is done with the right first quantization step by using the formulation as follows:

$$d_{out}(c, q_1, q_2, q_3) = \left| \left[\left[\left[\left[\frac{c}{q_1} \right] \frac{q_1}{q_2} \right] \frac{q_2}{q_3} \right] \frac{q_3}{q_2} \right] q_2 - \left[\left[\frac{c}{q_1} \right] \frac{q_1}{q_2} \right] q_2 \right| \tag{4}$$

Each d_{out} is evaluated for $q_3 = q_{1i}$ by considering a single DCT frequency. If this value approaches to zero, it is included to the limited candidates set C_s otherwise, discarded.

The double quantization process is simulated by considering the selected candidates. In this process, the first compression is performed by using selected candidates C_s and the second compression with known quantization step and finally a histogram is selected that best exploits the original histogram. Therefore, the desired first quantization step is selected which corresponds to the final selected histogram from a pool of selected candidates C_s.

3.1 Proposed DCT histogram filtering

Numerous methodologies [8, 9, 18, 23] generally do not consider the error e in (1) while employing the impacts of consecutive quantizations that take place after the de-quantizations. The performance of these methodologies reduces significantly when this source of error is neglected. But this simplification permits to handle the included mathematical equations easily. This error is introduced during the operations like color space conversions, rounding and truncation of the values. Based on the actual implementation, to examine each source of error separately is a challenging task. Therefore, the Gaussian distribution is used to model the overall effect of this error.
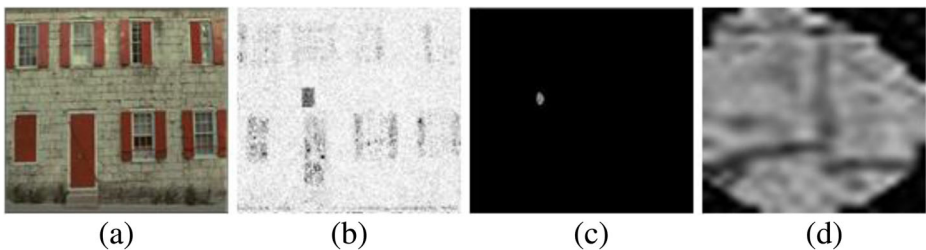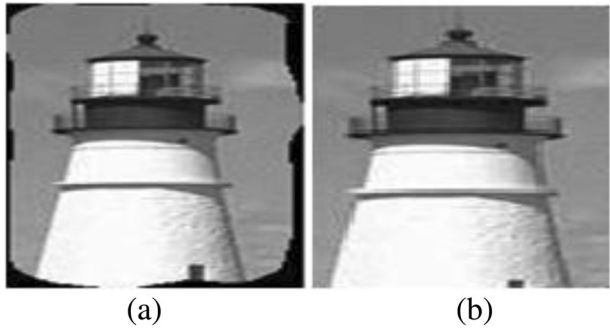

Fig. 8 (**a**) Partial double compressed image, Wall (**b**) Difference image through JPEG ghost, (**c**) Image after morphology, adaptive filtering, and masking operation, (**d**) Isolated double compressed region of size 23 × 30 pixels

Fig. 9 (**a**) Resultant double compressed region from the first stage, Tower image (**b**) Image after proper cropping in the second stage

Several situations that can arise when an image is compressed more than once depend on the magnitude of the primary and secondary quantization steps. In the first scenario, the wrong bin DCT coefficient elements are propagated in the resultant histogram when a small perturbation occurs as shown in Fig. 11a. In the second scenario, the bin containing original information in the histogram can be equally split into two adjoining bins and out of these two bins, one is a wrong bin, as shown in Fig. 11b. This undesired situation arises when a primary quantization bin in position uq_1 falls exactly halfway between two back to back bins in position vq_2 and $(v+1)q_2$ coming from the second quantization related as follows:

$$uq_1 = \frac{vq_2 + (v+1)q_2}{2}, \qquad u, v \in N^+ \tag{5}$$

Here a problem arises which has not been discussed previously, in which the bin from the second quantization $(v+1)q_2$ becomes common to the two different cases of split noise, as shown in Fig. 11c. The bin in position uq_1 of the primary quantization is situated exactly halfway between two back to back bins in position vq_2 and $(v+1)q_2$ coming from the second quantization. Similarly, the

Fig. 10 Error function (4) for a AC coefficient with $q_1 = 14$, $q_2 = 7$ and $q_3 \in \{1, 2, \ldots, 25\}$
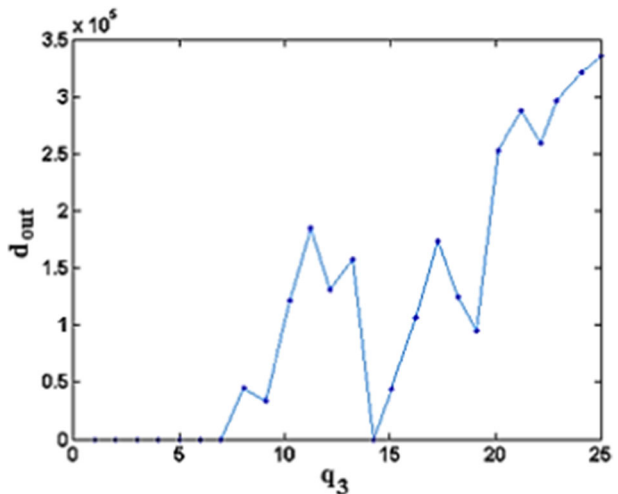
bin in position $(u+1)q_1$ is positioned absolutely in the middle of two bins coming from the second quantization in position $(v+1)q_2$ and $(v+2)q_2$. Thus, to cope with this problem a filtering algorithm has been proposed. The proposed algorithm first finds out the wrong bins in the double quantized histogram with the help of (5) by considering the particular values of quantization steps q_1 and q_2. The identified wrong bins are then moved to the right locations according to the proposed algorithm as shown in Fig. 12b. On the other hand, the residual noise is removed by setting the proper threshold as shown in Fig. 12c.

Filtering algorithm for the proposed split noise scenario

Parameters:

N_{bins} : Number of bins in DCT coefficients input histogram
bin_2: Bin height of the double compressed input histogram
uq_1: Bin centre position of single compressed histogram
vq_2: Bin centre position of double compressed histogram

begin
  for $i = q_{1min}: q_{1max}$ **do**
    for $j = 1: N_{bins}$ **do**
      for $k = 1: N_{bins}$ **do**
        $find(u(j)q_1(i) == [v(k)q_2 + (v(k)+1)q_2]/2)$
        if $bin_2(v(k)q_2) < bin_2((v(k)+1)q_2)$ **then**
          $bin_2((v(k)+1)q_2) = bin_2(v(k)q_2) + bin_2((v(k)+1)q_2)$
          $bin_2(v(k)q_2) = 0$
        else
          $bin_2(v(k)q_2) = bin_2(v(k)q_2) + bin_2((v(k)+1)q_2)$
          $bin_2((v(k)+1)q_2) = 0$
        end
      end
    end
  end
end

4 Experiment results

The efficacy of the proposed approach is examined by conducting several tests on different datasets images. The JPEG encoding is used with standard JPEG quantization tables suggested by the Independent JPEG Group [14]. The database is generated by considering the Kodak lossless true color image (PhotoCD PCD0992) dataset [5] and UCID (v2) dataset [26] images. A set of 560 partial double compressed images is obtained from both the datasets considering the quality factors ($QF1, QF2$) in the range 50 to 100. The images in both the datasets are aligned. The supporting idea for the arrangement of images is to consider an uncompressed JPEG image and to double compress the regions of variable sizes in descending order. The outputs are reported with respect to the quality factors rather than the particular quantization steps. This step simplifies the analysis of the results, since a single quality factor parameter
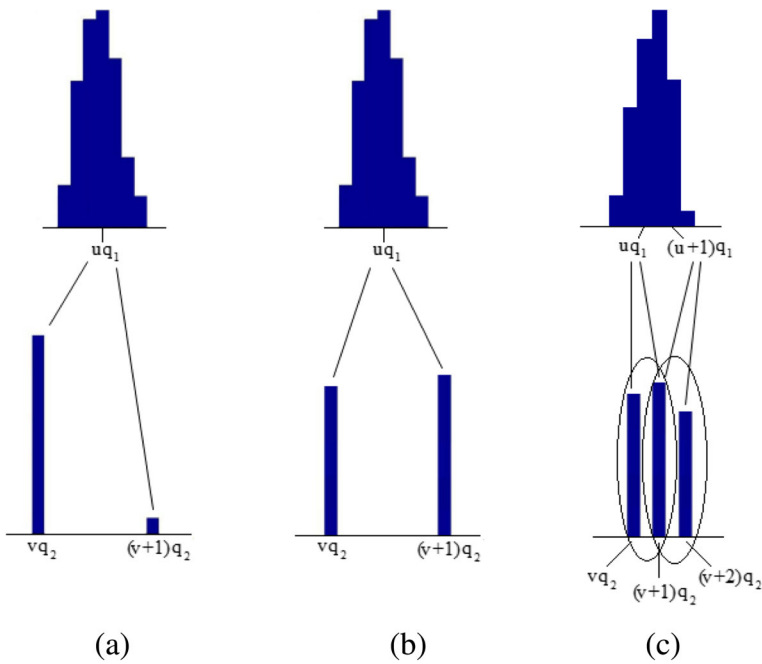
Fig. 11 (**a**) Represents the residual noise, (**b**) Represents split noise and (**c**) Represents the proposed split noise scenario

describes a quantization matrix with the 64 quantization steps that usually have different values corresponding to different frequencies as shown in Fig. 13. Since the most of the information is carried by the lower frequency DCT coefficients, therefore the experimental analysis is based on the first 15 components. The proposed partial double compression detection technique does not provide satisfactory results for the images in which recompression is performed with a cropping attack and the same quantization matrix. Since, after the cropping attack, if the image is recompressed with same quantization matrix, will lead to the desynchronization of DCT blocks [11, 22]. The main problem is then to evaluate the desynchronization of DCT blocks introduced into the image. The proposed detection scheme is unable to detect this desynchronization. Thus, the case of recompression with same quantization matrix is not considered in this paper.

Fig. 12 (**a**) Double quantized DCT histogram, (**b**) DCT histogram after split noise removal through proposed filtering algorithm, (**c**) DCT histogram after residual noise removal

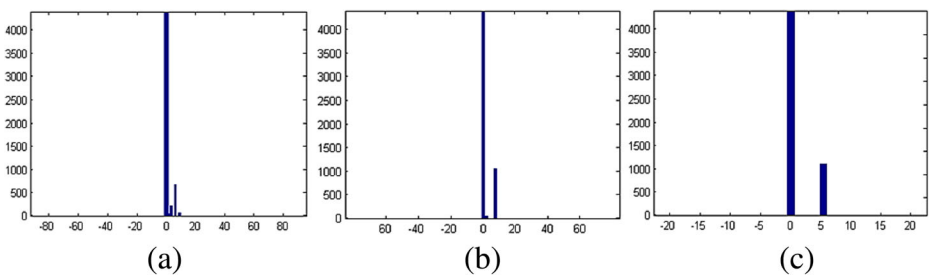**Fig. 13** Quantization matrix for quality factor (QF) =50, according to the JPEG standard

$$QF_{50} = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix}$$

The performance analysis of the first stage of proposed technique has been done by considering the different detection parameters on two different datasets. The blocking artifacts introduced in double compressed regions are computed using different detection parameters as shown in Tables 1 and 2. It is clear from the Tables 1 and 2 that the values of blockiness signature measure (K_F) [6], gradient aware blockiness measure $\left(B_{gr}^{\lambda}\right)$ [39], and the calibrated feature (K_L) [17], vary according to the present artifacts and the size of the double compressed region. The discussed detection parameters measure smoothness of the image, where for smooth images their values approach to zero. Therefore, these parameters also confirm that the isolated region is actually double compressed. These blocking artifacts detection parameters measure the effectiveness of the first stage in automatic isolation of double compressed region. The Tables 1 and 2 also shows the average values of various detection parameters by considering a set of 560 images build from the Kodak lossless true color image (PhotoCD PCD0992 dataset) as well as the UCID (v2) dataset. Therefore, the average values of various detection parameters further confirm the efficacy of first stage.

The percentage error of the first stage due to the edge irregularities of the isolated part increases with the reduction in the double compressed region size. The percentage error in isolation of double compressed region on the two different datasets i.e. Kodak lossless true color image (PhotoCD PCD0992) and UCID (v2) dataset is shown in Fig. 14. It is clear from the Fig. 14 that the error is larger in the case of UCID dataset due to their small image size which leads to the processing of fewer blocks as compared to the Kodak lossless true color image dataset.

The adaptive median algorithm is applied on the difference image generated from JPEG Ghost technique in which the boundary irregularities in the resultant binary image can be observed. The error due to the edge irregularities changes as the function of different JPEG compression factors. Tables 3 and 4 reports the average percentage of error in the first stage as a function of various JPEG compression factors on the Kodak lossless true color image

Table 1 Performance analysis of the first stage of proposed scheme based on various detection parameters by considering the Kodak lossless true color image (PhotoCD PCD0992 dataset)

Uncompressed Images	Double compressed region size	(K_F)	$\left(B_{gr}^{\lambda}\right)$	(K_L)
Parrot (768 × 512)	392 × 512 pixels	0.862	4.731	1.899
Roman (512 × 768)	260 × 385 pixels	0.486	3.773	0.356
Tower (768 × 512)	82 × 205 pixels	0.387	9.424	0.393
Wall (768 × 512)	23 × 30 pixels	1.428	4.003	1.319
Average values of detection parameters on the 560 images build from Kodak dataset		0.976	6.574	2.984

Table 2 Performance analysis of the first stage of proposed scheme based on various detection parameters by considering the UCID (v2) dataset

Uncompressed Images	Double compressed region size	(K_F)	$\left(B_{gr}^{\lambda}\right)$	(K_L)
Booth (384 × 512)	392 × 512 pixels	0.735	5.132	1.456
Car (512 × 384)	260 × 385 pixels	0.512	2.976	0.438
Flower (384 × 512)	82 × 205 pixels	0.262	8.313	0.381
Trophy (384 × 512)	23 × 30 pixels	1.107	4.747	1.421
Average values of detection parameters on the 560 images build from UCID (v2) dataset		0.814	5.921	2.138

(PhotoCD PCD0992 dataset) [5] and UCID (v2) dataset [26] respectively. It can be seen from the Tables 3 and 4 that the accuracy decreases in the case of UCID (v2) dataset as compared to the Kodak lossless true color image dataset due to their small image sizes.

Tables 5 and 6 reports the average percentage error of the second stage by considering 560 partial double compressed images created from Kodak lossless true color image (PhotoCD PCD0992 dataset) as well as the UCID (v2) dataset. The percentage of error is calculated by estimating the q_1 values at different quality factor values. Since the analysis is performed on the partial double compressed image datasets, therefore the isolated double compressed region is processed in the second stage to estimate the first quantization matrix whereas in the existing state-of-the-art techniques full double compressed images are used [2, 8, 9, 20, 24]. The DCT coefficients related to the low frequency in the DCT domain have small estimation error and do not essentially rely on the particular quality factor concerned with the first and second quantization. On the other hand, the estimation error for higher frequencies has significant correlation with the quality factors. The results obtained are usually better for higher $QF1$ and $QF2$ values as compare to the lower quantization. The comparative analysis of Tables 5 and 6 depicts that the percentage error in estimating the q_1 values at different quality factor corresponding to the first 15 DCT coefficients is more in case of UCID dataset as compare to the Kodak lossless true color image dataset due to the small size of images in UCID dataset.
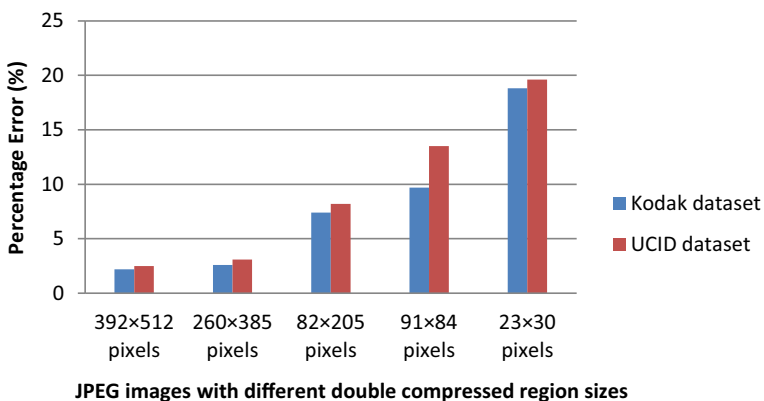

Fig. 14 Comparative analysis of percentage error (%) in the automatic isolation of the double compressed region through the first stage by considering Kodak lossless true color image (PhotoCD PCD0992) and UCID (v2) dataset

Table 3 Average percentage of error in the first stage as a function of various JPEG compression factors on the Kodak lossless true color image (PhotoCD PCD0992 dataset)

$Q_2 \setminus Q_1$	50	60	70	80	90
50	–	7.455	15.395	11.555	69.780
60	10.875	–	7.564	34.895	34.988
70	13.321	8.766	–	12.044	55.875
80	2.739	12.306	4.775	–	9.125
90	5.775	1.885	3.431	9.705	–

Table 4 Average percentage of error in the first stage as a function of various JPEG compression factors on the UCID (v2) dataset

$Q_2 \setminus Q_1$	50	60	70	80	90
50	–	9.775	17.635	12.550	71.703
60	12.766	–	8.414	36.750	36.999
70	14.975	10.650	–	13.120	57.845
80	3.886	14.285	5.745	–	11.535
90	8.746	4.250	6.455	12.548	–

Table 5 Average percentage of error in estimated $q1$ values at different quality factor corresponding to the first 15 DCT coefficients on the Kodak lossless true color image (PhotoCD PCD0992 dataset)

$QF_1 \setminus QF_2$	60	70	80	90	100
50	10.352	9.205	3.759	4.675	2.503
60	–	14.263	8.445	5.537	2.149
70	–	–	5.597	4.433	1.200
80	–	–	–	3.250	0.950
90	–	–	–	–	0.255

Table 6 Average percentage of error in estimated $q1$ values at different quality factor corresponding to the first 15 DCT coefficients on the UCID (v2) dataset

$QF_1 \setminus QF_2$	60	70	80	90	100
50	11.251	10.523	4.940	7.125	3.340
60	–	16.221	9.246	5.565	2.458
70	–	–	7.252	4.986	1.467
80	–	–	–	4.745	1.126
90	–	–	–	–	0.356

Table 7 Average percentage of error in estimated q_1 values corresponding to the DCT coefficient in zig-zag order considering several state-of-the-art approaches on the Kodak lossless true color image (PhotoCD PCD0992 dataset)

		Percentage Error (%)				
		Lukas [20]	Farid [8]	Bianchi [2]	Galvan [9]	Proposed Scheme
DCT Coefficient	1	0.41	0.25	0.31	0.14	**0.11**
	2	0.42	**0.21**	0.82	0.92	0.82
	3	0.53	**0.22**	1.09	0.95	0.93
	4	5.81	2.52	4.21	1.57	**1.42**
	5	3.77	1.83	2.33	0.84	**0.81**
	6	**1.26**	1.80	13.82	1.36	1.38
	7	8.00	12.91	5.97	2.51	**1.41**
	8	6.21	5.82	6.12	0.86	**0.80**
	9	4.01	6.01	8.23	**0.83**	1.07
	10	8.52	6.10	6.22	1.22	**1.13**
	11	7.33	20.06	11.08	8.57	**5.01**
	12	10.28	18.23	12.16	8.02	**6.02**
	13	12.02	18.07	7.02	4.02	**4.00**
	14	14.01	16.12	10.08	7.05	**5.02**
	15	24.42	29.23	16.20	13.01	**12.01**

The bold values correspond to the lowest percentage of error for a particular DCT coefficient

In order to study the effectiveness of the proposed approach, further analyzes have been led with respect to the specific DCT coefficient. The Tables 7 and 8 show the average percentage of error in the estimation of q_1 values that corresponds to the DCT coefficients on two different datasets. For the DCT coefficients corresponding to the high frequencies, the performance of the proposed scheme degrades but it provides high accuracy with an error less than 1.5% for the first 10 DCT coefficients as shown in Table 7.

Table 8 Average percentage of error in estimated q_1 values corresponding to the DCT coefficient in zig-zag order considering several state-of-the-art approaches on the UCID (v2) dataset

		Percentage Error (%)				
		Lukas [20]	Farid [8]	Bianchi [2]	Galvan [9]	Proposed Scheme
DCT Coefficient	1	4.71	5.27	5.05	3.52	**3.49**
	2	2.82	2.45	4.37	**1.81**	2.01
	3	3.01	2.66	5.80	2.07	**1.80**
	4	3.83	2.89	10.08	3.02	**2.07**
	5	3.57	2.62	7.81	2.01	**1.82**
	6	3.42	4.23	16.06	3.88	**3.06**
	7	5.52	7.07	8.07	5.07	**4.63**
	8	8.51	4.28	13.80	4.92	**3.04**
	9	4.03	4.07	11.81	**3.96**	6.05
	10	6.80	5.06	12.23	5.07	**5.03**
	11	15.02	8.07	8.07	10.91	**8.02**
	12	7.03	7.81	12.06	5.20	**4.01**
	13	8.04	10.67	8.01	**7.91**	10.02
	14	13.02	10.25	13.50	8.23	**7.04**
	15	20.01	19.83	12.02	14.50	**11.43**

The bold values correspond to the lowest percentage of error for a particular DCT coefficient

A comparative analysis of this second stage of proposed scheme is done with the various algorithms proposed in [2, 8, 9, 20]. The application of the second stage of the proposed scheme, when first 9 DCT coefficients are considered in a zig-zag manner, mixed results are obtained with few coefficients providing high error values. But, after the 9th DCT coefficient, the percentage error recorded is much low from the previously established techniques as shown in Table 7. If the Table 7 is closely examined, the percentage error can be conveniently justified. It can be inferred that the proposed scheme outperforms the method proposed by Bianchi [2] for all the considered 15 DCT coefficients. Whereas, the proposed scheme lag behind the schemes provided by Lukas [20] and Galvan [9] in only one DCT coefficient. Farid [8] surpass the proposed scheme for only two DCT coefficients.

The proposed filtering step considerably improves the performance of the second stage that copes with rounding and truncation error efficiently. The Tables 7 and 8 shows that the proposed scheme provides less percentage error as compared to the considered several state-of-the-art approaches in estimating the q_1 values corresponding to the DCT coefficient in zig-zag order on both datasets. It is clear from the Table 8 that the proposed scheme outperforms the existing techniques proposed by Lukas [20], Farid [8] and Bianchi [2] for all the considered 15 DCT coefficients. The technique provided by the Galvan [9] outperforms the proposed scheme for only two DCT coefficients.

The performances of the considered methods in the test on UCID (v2) dataset as shown in Table 8 are lower than on the Kodak lossless true color image dataset as shown in Table 7. The efficiency of considered schemes depends on the resolution of the image under analysis. The reliability of the analysis could be low as shown in Table 8, because of small size images in UCID (v2) dataset. Nevertheless, in today's scenario working with small images is not so common.

5 Conclusions

The unethical use of digital images emerges the forensic investigation techniques to find out the authenticity of the digital images. In this paper, a technique is proposed to detect and analyze the partial doubly compressed JPEG image which leads to the estimation of the first quantization matrix. The proposed technique is based on the concept that to estimate the first quantization matrix from the partial double compressed image, it is desired to detect and isolate the double compressed region efficiently. The experimental results show that the first stage of the proposed scheme has satisfactory performances with an average percentage accuracy of 95.45% even when the detected double compressed regions are of a small size. The proposed filtering strategy increases the accuracy to estimate the first quantization matrix in the second stage. For the first 10 DCT coefficients, an error less than 1.5% has been recorded. The experiment results depict that the performance of the proposed approach is better as compared to the considered state-of-the-art techniques for the partial double compressed images based on the two different datasets. The evaluation is based on the partial double compressed images in which the recompression is performed with different quantization matrix. The work can be extended to create a detection methodology for the partial double compressed images, where recompression is performed with same quantization matrix. The proposed scheme can also be generalized from the color images to the hyperspectral images.

References

1. Battiato S, Messina G (2009) Digital forgery estimation into DCT domain: a critical analysis. In: Proc. ACM Workshop Multimedia Forensics (MiFor), pp. 37–42
2. Bianchi T, Piva A (2012) Image forgery localization via block-grained analysis of JPEG artifacts. IEEE Trans. Inf. Forensics Security 7(3):1003–1017
3. Chen YL, Hsu CT (2011) Detecting recompression of JPEG images via periodicity analysis of compression artifacts for tampering detection. IEEE Trans. Inf. Forensics Security 6(2):396–406
4. Chen C, Shi YQ, Su W (2008) A machine learning based scheme for double JPEG compression detection. In: Proc. In: 19th Int. Conf. Pattern Recognit. (ICPR), pp, pp 1–4
5. Dataset Eastman Kodak [Online] (2014) Available: http://r0k.us/graphics/kodak/
6. Fan Z, De Queiroz RL (2003) Identification of bitmap compression history: JPEG detection and quantizer estimation. IEEE Trans Image Process 12(2):230–235
7. Farid H (2009a) Image forgery detection. IEEE signal process. Mag 26(2):16–25
8. Farid H (2009b) Exposing digital forgeries from JPEG ghosts. IEEE Trans. Inf. Forensics Security 1(4):154–160
9. Galvan F, Puglisi G, Bruna AR, Battiato S (2014) First quantization matrix estimation from double compressed JPEG images. IEEE Trans. Inf. Forensics Security 9(8):1299–1310
10. He J, Lin Z, Wang L, Tang X (2006) Detecting doctored JPEG images via DCT coefficient analysis. In: proc. 9th Eur. Conf Comput Vis (ECCV) 3953:423–435
11. Ho ATS, Li S (2015) Handbook of Digital forensics of multimedia data and devices. Wiley. doi:10.1002/9781118705773
12. Huang F, Huang J, Shi YQ (2010) Detecting double JPEG compression with the same quantization matrix. IEEE Trans. Inf. Forensics Security 5(4):848–856
13. Huang DY, Huang CN, Hu WC, Chou CH (2015) Robustness of copy-move forgery detection under high JPEG compression artifacts. Multimed Tools Appl. 1–22
14. Independent JPEG Group [Online] (2014) Available: http://www.ijg.org/
15. Jaiswal N, Meghrajani YK (2015) Saliency Based Automatic Image Cropping Using Support Vector Machine Classifier. In: Proc. Int. Conf. on Innovations in Information Embedded and Communication Systems, pp. 1–5
16. Kee E, Johnson MK, Farid H (2011) Digital image authentication from JPEG headers. IEEE Trans. Inf. Forensics Security 6(3):1066–1075
17. Lai S, Böhme R (2011) Countering counter-forensics: The case of JPEG compression. In: Proc. Int. Conf. Inf. Hiding, pp. 285–298
18. Lin Z, He J, Tang X, Tang CK (2009) Fast, automatic and fine-grained tampered JPEG image detection via DCT coefficient analysis. Pattern Recogn 42(11):2492–2501
19. Liu Q (2011) Detection of misaligned cropping and recompression with the same quantization matrix and relevant forgery. In: Proc. 3rd Int. ACM Workshop Multimedia Forensics Intell. (MiFor), pp. 25–30
20. Lukas J, Fridrich J (2003) Estimation of primary quantization matrix in double compressed JPEG images. In: Proc. Digit. Forensic Res. Workshop (DFRWS), pp. 5–8
21. Luo W, Huang J, Qiu G (2010) JPEG error analysis and its applications to digital image forensics. IEEE Trans. Inf. Forensics Security 5(3):480–491
22. Pevny T, Fridrich J (2008) Multi-class detector of current Steganographic methods for JPEG format. IEEE Trans. Inf. Forensics Security 3(4):635–650
23. Popescu AC, Farid H (2004) Statistical tools for digital forensics. In: Proc. 6th Int. Workshop Inform. Hiding, pp. 128–147
24. Puglisi G, Bruna AR, Galvan F, Battiato S (2013) First JPEG quantization matrix estimation based on histogram analysis. In: Proc. Int. Conf. on Image Processing, pp. 4502–4506
25. Redi JA, Dugelay JL (2011) Digital image forensics: a booklet for beginners. Multimed Tools Appl. 51(1):133–162
26. Schaefer G, Stich M (2004) UCID-an uncompressed colour image database. In Proc SPIE 5307:472–480
27. Shi F, Kang B, Li H, Zhu Y (2012) A new method for detecting JEPG doubly compression images by using estimated primary quantization step. Proc. Int. Conf. on Systems and Informatics, In, pp 1810–1814
28. Stamm MC, Liu KJR (2013) Information forensics: an overview of the first decade. IEEE Access 1:167–200
29. Swaminathan A, Wu M, Liu KJR (2008) Digital image forensics via intrinsic fingerprints. IEEE Trans. Inf. Forensics Security 3(1):101–117
30. Piva A (2013) An overview on image forensics. In: Proc. ISRN Signal Process, pp. 496–701
31. Thing VLL, Chen Y, Cheh C (2012) An improved double compression detection method for JPEG image forensics. In: Proc. IEEE Int. Symp. Multimedia (ISM), pp. 290–297
32. Yang J, Xie J, Zhu G, Kwong S, Shi YQ (2014) An effective method for detecting double JPEG compression with the same quantization matrix. IEEE Trans Inf Forensics Security 9(11):1933–1942
33. Zhang R, Wang RD (2015) In-camera JPEG compression detection for doubly compressed images. Multimed Tools Appl 74(15):5557–5575

34. Taimori A, Razzazi F, Behrad A, Ahmadi A, Zadeh MB(2016) A novel forensic image analysis tool for discovering double JPEG compression clues. Multimed Tools Appl. 1–35
35. Wang W, Dong J, Tan T (2011) Exploring DCT coefficient quantization effect for image tampering localization. In Proc. IEEE Int. Workshop Inform. Forensics Security, pp. 1–6
36. Zhang J, Li H, Chen CW (2015a) Distributed lossless coding techniques for hyperspectral images. IEEE Journal of Selected topics in Signal Processing 9(6):977–989
37. Zhang L, Zhang L, Tao D, Huang X, Du B (2015b) Compression of hyperspectral remote sensing images by tensor approach. Neurocomputing 147:358–363
38. Zach F, Riess C, Angelopoulou E (2012) Automated image forgery detection through classification of JPEG ghosts. In: Proc. of the joint 34th DAGM and 36th OAGM symposium, pp. 185–194
39. Ullerich C, Westfeld A (2008) Weaknesses of MB2. In: Proc. Int. Workshop Digital Watermarking, pp. 127–142

Gurinder Singh received B. Tech. degree in electronics and communication engineering in 2012 from Lovely Professional University, Punjab, and the M. Tech. degree in electronics and communication engineering in 2014 from Sri Guru Granth Sahib World University, Punjab. He is pursuing PhD degree in electronics and communication engineering from Thapar University, Patiala. His research interest includes the field of digital signal processing, digital image forensics and anti-forensics.

Kulbir Singh was born in Batala, Punjab, India. He completed his B. Tech. degree in 1997 at PTU, Jalandhar. He received his M.E. and Ph.D. degrees from Thapar University, Patiala, in 2000 and 2006, respectively. He is currently working as a professor in the Department of Electronics and Communication Engineering, Thapar University, Patiala. He has published more than 65 research papers in national and international journals/conference proceedings. He is a recipient of the Best Paper Award of the *IETE Journal of Education* for the year 2008. His research interests include signal processing, image processing, DSP processors based design and fractional transforms.