First Quantization Matrix Estimation for Double Compressed JPEG Images Utilizing Novel DCT Histogram Selection Strategy

Nandita Dalmia
Department of Electronics and Communication Engineering
National Institute of Technology Rourkela
Rourkela, India-769008
dalmianandita@gmail.com

Manish Okade
Department of Electronics and Communication Engineering
National Institute of Technology Rourkela
Rourkela, India-769008
okadem@nitrkl.ac.in

ABSTRACT

The Double JPEG problem in image forensics has been gaining importance since it involves two compression cycles and there is a possibility of tampering having taken place after the first cycle thereby calling for accurate methods to detect and localize the introduced tamper. First quantization matrix estimation which basically retrieves the missing quantization table of the first cycle is one of the ways of image authentication for Double JPEG images. This paper presents a robust method for first quantization matrix estimation in case of double compressed JPEG images by improving the selection strategy which chooses the quantization estimate from the filtered DCT histograms. The selection strategy is made robust by increasing the available statistics utilizing the DCT coefficients from the double compressed image under investigation coupled with performing relative comparison between the obtained histograms followed by a novel priority assignment and selection step, which accurately estimates the first quantization value. Experimental testing and comparative analysis with two state-of-art methods show the robustness of the proposed method for accurate first quantization estimation. The proposed method finds its application in image forensics as well as in steganalysis.

CCS Concepts

•**Computing methodologies** → **Image processing;** •**Applied computing** → *Evidence collection, storage and analysis;*

Keywords

Image forensics; double JPEG compression; tampering; histogram analysis; prioritized lists; split noise.

ICVGIP, December 18-22, 2016, Guwahati, India

© 2016 ACM. ISBN 978-1-4503-4753-2/16/12. . . $15.00

DOI: http://dx.doi.org/10.1145/3009977.3010067

1. INTRODUCTION

Image forensics focuses on analyzing image data for possible tampering which could have taken place due to splicing, cloning, retouching, and/or other digital manipulations [5, 9]. Image forensic methods can be broadly divided into two categories: active and passive. Inserting watermark is an active forensic approach where data is embedded during the time when the image is formed so that tampering detection at later instant of time can be possible. Such class of methods which insert an imperceptible watermark or signature falls under the category of active forensics. On the contrary, passive forensic doesn't utilize any additional information such as watermark and/or a signature and hence regarded as passive since the analysis is blind. Passive techniques focus their analysis on the statistical behaviour of image data, its origin, compression pipeline, its storage and what operation could have possibly been carried out on these images i.e. trying to reconstruct its history and checking if any changes have been brought about deliberately. JPEG is one of the very popular image compression standards which is used by most cameras as well as image editing softwares like Adobe, GIMP etc. JPEG based forensic methods basically can be divided into two categories; firstly, techniques which analyse the metadata and/or structure [8, 10] while second group that analyse the compression pipeline (i.e. transform, quantize, encode) and reconstruct its history [4, 2, 12, 3]. In this context double JPEG compression has been studied [11, 7]. Double compression occurs when a previously saved JPEG image is decompressed and then resaved. This might have taken place because of the replacement of certain portion of previously saved image by another portion from altogether different image. In such cases two quantization matrices will be participating, the first quantization matrix before the forgery occurs and second quantization matrix which results after the portion is pasted and resaved. Since the first quantization matrix is unavailable in such cases, one must determine it accurately which is the focus of the current work. The estimation of the first quantization matrix is also useful in steganalysis [6, 13].

Researchers have previously focused their attention on this problem and proposed various schemes for estimation of first quantization matrix. Few among them are highlighted here. For a detailed review, readers are kindly referred to [15]. Bianchi and Piva [1] investigated a novel approach to detect and localize tampering under double JPEG compres-

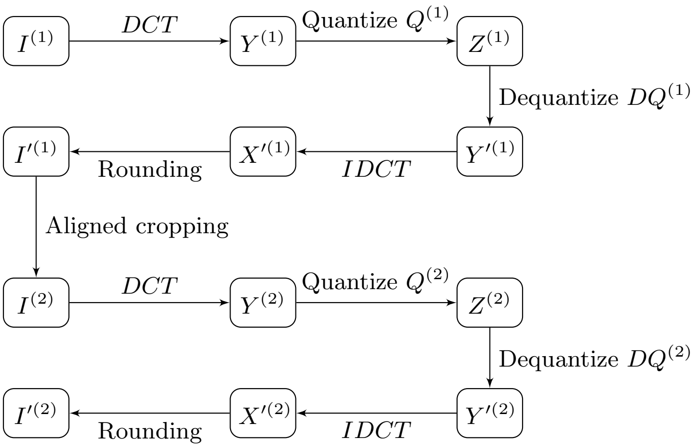


Figure 1: Pipeline showing the $first^{(1)}$ JPEG operations and $second^{(2)}$ JPEG operations for the Double JPEG case

sion by using a statistical model on the DCT coefficients and deriving a likelihood map which showed the probability of each DCT block being double compressed. Lucas et al. [11] proposed three methods for first quantization matrix estimation which were based on histogram normalization. Farid [4] proposed a tamper localization technique for low resolution JPEG images which have undergone double compression by detecting spatially localized minima in the difference between the image and its JPEG compressed counterpart. They carried out a third quantization step and computed the error between the DCT coefficients by varying the quantization step of the third quantization which was able to localize two minima corresponding to first quantization and second quantization. This method was improved by Galvan et al. [7] to detect and localize large tampered regions. Their approach introduced an improved error computing term by accounting for split as well as residual noise utilizing the histogram filtering strategy with the aim to counter the round-off and truncation error impact on the coefficient values. However, it has been observed that although they achieve better estimate of first quantization values in comparison to Bianchi and Piva [1] as well as Farid [4] yet their method gives estimates which are often very close to the actual first quantization values. In forensic applications where the integrity of image is being tested, such approximate estimates will hamper the investigation process. Motivated by this, we investigate a robust estimation method for first quantization matrix based on utilizing relative priority assignment as a strategy for accurate first quantization matrix estimation. Rest of the paper has been organized as follows: Section 2 gives a brief overview of the JPEG compression pipeline and the noises introduced by the compression process. This is followed by motivation and problem formulation in section 3. Section 4 describes the proposed method which is followed by section 5 where the experimental validation, performance evaluation and robustness analysis has been discussed to prove the superiority of the proposed method in comparison to state-of-art methods. Finally, in section 6 the conclusions have been drawn along with scope of future work.

2. JPEG COMPRESSION OVERVIEW

JPEG compression pipeline consists of DCT, quantization by fixed tables followed by Huffman encoding. The decompression does the inverse of each of these steps i.e. decode, dequantize and IDCT. The image is partitioned into (8×8) non-overlapping pixel blocks B(x,y); $0 \leq x, y \leq 7$. The block wise DCT is calculated using

$$D(u,v) = \sum_{x=0}^{7}\sum_{y=0}^{7} B(x,y)g(x,y,u,v)$$

$$g(x,y,u,v) = \alpha(u)\alpha(v)cos\Big(\frac{(2x+1)\pi u}{16}\Big)\Big(cos\frac{(2y+1)\pi v}{16}\Big)$$

$$\alpha(u) = \alpha(v) = \begin{cases}\sqrt{\frac{1}{N}}, & u,v=0\\ \sqrt{\frac{2}{N}}, & u,v=1,2,..,7\end{cases} \tag{1}$$

The matrix D_{uv}, $(u, v\epsilon\{0,1,2,...7\})$ is then quantized using quantization matrix Q_{uv}.

$$D^q_{uv} = round\Big(\frac{D_{u,v}}{Q_{u,v}}\Big)$$

Coefficients of D^q_{uv} are ordered using zig-zag scan and encoded using the Huffman encoder. Decompression is carried out by multiplying D^q_{uv} by Q_{uv} (i.e. $D_{uv} = Q_{uv} \times D^q_{uv}$) and taking IDCT followed by rounding the result and truncating to integer values in the range [0, 255]. Figure 1 shows the JPEG pipeline for the Double-JPEG (DJPEG) scenario with superscript (1) denoting first compression and (2) denoting second compression. Rounding and truncation error are introduced by the DCT step while quantization error is introduced by the quantization operation. Entropy coding/decoding is neglected since it is assumed to be perfectly reversible. The two successive compressions occurring in DJPEG can be modeled as below:

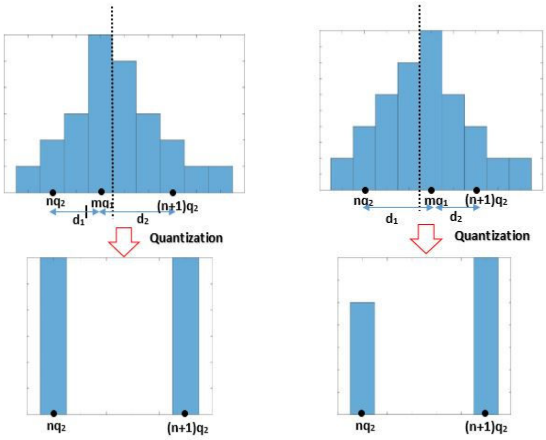
$$I'^{(1)} = IDCT^{(1)}(DQ^{(1)}(Q^{(1)}(DCT^{(1)}(I^{(1)})))) \tag{2}$$

$$I'^{(2)} = IDCT^{(2)}(DQ^{(2)}(Q^{(2)}(DCT^{(2)}(I^{(2)})))) \tag{3}$$

where $I^{(1)}$ is the uncompressed image acting as the input to the pipeline, $I^{(2)}$ is the image formed after tampering with quantization step being represented as Q and dequantization step as DQ.

3. MOTIVATION AND PROBLEM FORMULATION

The rounding and truncation operation present in the JPEG compression pipeline is a source of error and has been a major challenge for forensic experts. This error gets compounded in case of DJPEG images since the second DCT step results in spreading of the coefficients around the multiples of first quantization step. Due to this spread when quantization is performed, the coefficients may move to a wrong bin resulting in split noise and residual noise [7]. Galvan et al. [7] considered the case where bin of the first quantization (q_1) was equidistant from two neighbouring bins of



(a) (b)

Figure 2: Additional instances of split noise (a) $d_1 < d_2$ **(b)** $d_1 > d_2$

the second quantization (q_2) values i.e.

$$mq_1 = \left(\frac{nq_2 + (n+1)q_2}{2}\right), n, m \in \mathbb{N}^+ \tag{4}$$

where $\mathbb{N}^+$ is a set of natural numbers. Galvan et al.[7] addressed this case of split noise by filtering the DCT histograms $\langle H_{filtq1i} \rangle$ followed by the utilization of error function given by

$$\begin{aligned} &f(c, q_1, q_2, q_3) \\ &= \left| \left[\left[\left[\left[\frac{c}{q_1} \right] \frac{q_1}{q_2} \right] \frac{q_2}{q_3} \right] \frac{q_3}{q_2} \right] q_2 - \left[\left[\frac{c}{q_1} \right] \frac{q_1}{q_2} \right] q_2 \right| \end{aligned} \tag{5}$$

where, $[.]$ denotes the rounding function and q_3 denotes the probable first quantization step size. The error function values i.e. $\langle f_{q1i} \rangle$ (Eq. 5) were estimated corresponding to the probable first quantization estimates referred as $\langle q_{1i} \rangle$. Assuming $q_1 > q_2$, a subset $\langle f_{q1s} \rangle$ was formed which contained certain number of minimum values from $\langle f_{q1i} \rangle$ based on a threshold along with its corresponding $\langle q_{1i} \rangle$ entries referred as $\langle q_{1s} \rangle$. The estimation of the first quantization value ($\hat{q_1}$) was then performed utilizing the selection strategy

$$\hat{q_1} = \min_{q_{1s}} \sum_{i=1}^{N} min(max_{diff}, |H_{real}(i) - H_{q1s}(i)|) \tag{6}$$

where, $\langle q_{1s} \rangle$ was the set of probable first quantization values, max_{diff} was a threshold, H_{real} was the histogram of double compressed JPEG image under consideration, $H_{q_{1s}}$ was the synthetic histogram obtained when double quantization process is simulated on the double compressed image under investigation by cropping the image by 4×4 pixels followed by first quantization utilizing $\langle q_{1s} \rangle$ and second quantization utilizing data available from the JPEG header. However, there are many other scenarios apart from the equidistant case ($d_1 = d_2$) considered by Galvan et al. [7] like the one depicted in Figure 2 where Eq. (4) is strictly not satisfied i.e. for $d_1 < d_2$ (Figure 2(a)), the location of mq_1 is displaced towards nq_2 while for $d_1 > d_2$ (Figure 2(b)) the location of mq_1 is displaced towards $(n+1)q_2$ which are also instances of split noise and residual noise that have not been accounted by Galvan's work. In addition, Galvan's work [7] also did not account for the split and residual noise introduced while performing double JPEG compression to form histogram H_{q1s} (Eq. 6). These shortcomings led to incorrect estimation of first quantization value by their method. Motivated from this shortcoming, a robust method which utilizes the DCT coefficients from the image under investigation for increasing the available statistics coupled with carrying out a relative comparison between obtained histograms followed by a novel priority assignment and selection strategy is proposed in this paper which achieves accurate estimation of first quantization values as described below.

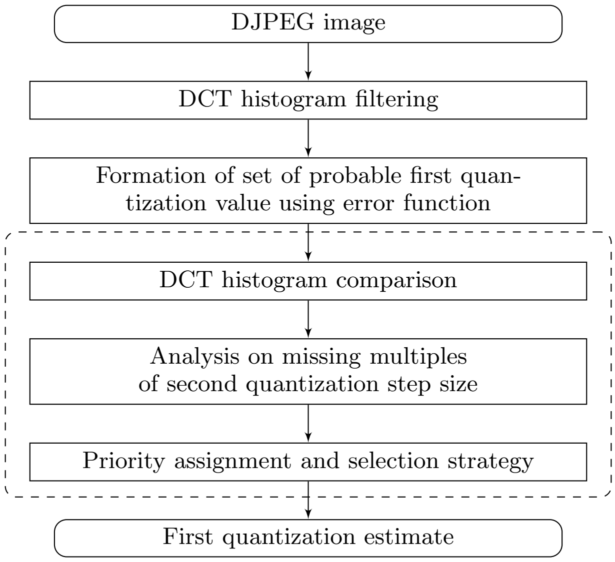


Figure 3: Overview of the proposed method

4. PROPOSED METHOD

Figure 3 shows the overview of the proposed method with the dotted area being our novel contribution for accurate estimation of first quantization matrix estimates. The DCT histograms are filtered to counter the effect of split noise and residual noise, in-line with Galvan's work [7] followed by selection of probable values of first quantization estimate using fixed threshold where q_1 values corresponding to first 6 minima of f_{q1i} form the set $\langle q_{1s} \rangle$. However, as pointed out earlier, the choice of first quantization estimate by Galvan is based on choosing the absolute minimum (Eq. 6) which doesn't arrive at the correct first quantization estimate. In-order to overcome this shortcoming, instead of performing the DCT histogram selection step as given by Galvan, the following proposed selection strategy is adopted:

1. DCT histogram comparison is performed utilizing the histogram of actual DJPEG image and the synthesized DJPEG image to obtain a list of values denoted as $\langle h_{q1s} \rangle$ by using

$$h^k_{q1s} = \sum_{i=1}^{length(H_{real})} |H_{real}(i) - H^k_{q1s}(i)| \tag{7}$$

Eq. (7) results in a set of values $\langle h^1_{q1}, h^2_{q1} \ldots h^6_{q1} \rangle \in \langle h_{q1s} \rangle$ for corresponding $\langle q_{1s} \rangle$.

2. Utilizing the DCT coefficients (c_{DQ}) from the double compressed JPEG image under investigation, the missing multiples of q_2 in the range: [minimum (c_{DQ})] to [maximum (c_{DQ})] are located i.e. $\langle k^1, k^2, \dots k^n \rangle$. This is followed by application of error function f (Eq. 5) to form a list $\langle f_{mk^n q1s} \rangle$ where $'m'$ represents missing multiples. The elements of this list are added using

$$f_{mq1s} = \sum_{\forall k^n} f_{mk^n q1s} \tag{8}$$

Algorithm 1 First Quantization estimation using priority lists

Input: Lists $\langle f_{q1s} \rangle$, $\langle h_{q1s} \rangle$, $\langle f_{mq1s} \rangle$
Output: First quantization value ($\hat{q_1}$)
1. $Priority\_f_{q1s} = Extract\_Min(f_{q1s})$
2. $Priority\_h_{q1s} = Extract\_Min(h_{q1s})$
3. $Priority\_f_{mq1s} = Extract\_Max(f_{mq1s})$
4. $F_{q1s} = Priority\_f_{q1s} + Priority\_h_{q1s} + Priority\_f_{mq1s}$
5. **if** $\langle F_{q1s} \rangle$ has a unique minimum value **then**
6. $\quad \hat{q_1} = \min_{\boldsymbol{q_{1s}}} \langle F_{q1s} \rangle$
7. **else**
8. $\quad \langle q_1 ms \rangle = Extract\_Multiple(F_{q1s}, q_{1s})$
9. $\quad$ **if** $\left(\dfrac{q^j_{1ms}}{q^i_{1ms}} \right) == r$, $(r \in \mathbb{N}^+)$ **then**
10. $\quad\quad$ delete q^i_{1ms} from $\langle q_{1ms} \rangle$
11. $\quad$ **end if**
12. $\quad \hat{q_1} = \min_{\boldsymbol{q_{1ms}}} \langle Priority\_h_{q1s} \rangle$
13. **end if**

Algorithm 2 $Extract\_Min(X)$

Input: List X
Output: Priority list of X

for $i = 1$ to length(X) **do**
$\quad (value, location) = minimum(X)$
$\quad Priority\_X(location) = i$
$\quad$ delete(X(location))
end for

Algorithm 3 $Extract\_Max(X)$

Input: List X
Output: Priority list of X

for $i = 1$ to length(X) **do**
$\quad (value, location) = maximum(X)$
$\quad Priority\_X(location) = i$
$\quad$ delete(X(location))
end for

Algorithm 4 $Extract\_Multiple(F_{q1s}, q_{1s})$

Input: List $\langle F_{q1s} \rangle$, $\langle q_{1s} \rangle$
Output: List $\langle q_{1ms} \rangle$

$(value, location) = minimum(F_{q1s})$
$j = 1$
for i = location to $size(F_{q1s})$ **do**
$\quad$ **if** $(F_{q1s}(i) == value)$ **then**
$\quad\quad q_{1ms}(j) = q_{1s}(i)$
$\quad\quad j++$
$\quad$ **end if**
end for

3. The lists $\langle h_{q1s} \rangle$, $\langle f_{q1s} \rangle$ and $\langle f_{mq1s} \rangle$ are further utilized for carrying out the first quantization estimate as described below (Algorithm 1):

   The elements of the list $\langle f_{q1s} \rangle$ and $\langle h_{q1s} \rangle$ are assigned priority based on relative ranking by considering the minimum criterion (line 1-2) as the minimum of the list $\langle f_{q1s} \rangle$ and $\langle h_{q1s} \rangle$ in the ideal case would indicate to first quantization value. Next, the elements of the list $\langle f_{mq1s} \rangle$ are assigned priority based on relative ranking by considering the maximum criterion (line 3) as backtracking over missing multiple by using correct estimate is expected to give maximum error. The relative priority assignment strategy to the lists using $Extract\_Min()$ and $Extract\_Max()$ procedure aids in countering the split noise effects depicted in Figure 2 as well as effect of residual and split noise while forming H_{q1s}. The three lists which have been assigned priority are added to calculate the final list $\langle F_{q1s} \rangle$ which will be further utilized for localizing the first quantization estimate. The summation of the priority lists (line 4) is taken under the premise that though the three lists may individually be affected with noise yet while considering their summation these noises are likely to be neutralized. If the list $\langle F_{q1s} \rangle$ has unique minimum value then that value is selected and the corresponding $\langle q_{1s} \rangle$ is chosen as first quantization estimate. However, the list $\langle F_{q1s} \rangle$ may contain multiple values having equal minimum priority. In such a case, the entries corresponding to the equal minimum priority are extracted from $\langle q_{1s} \rangle$ and stored in a list $\langle q_{1ms} \rangle$ using procedure $Extract\_Multiple()$ (line 8). Further, the list $\langle q_{1ms} \rangle$ is examined and if integer multiples are found then the lower integer value is removed because the lower value is likely to follow the trend of its multiple (line 9-11). Finally, the minimum of $\langle Priority\_h_{q1s} \rangle$ for corresponding value of $\langle q_{1ms} \rangle$ indicates towards the true value of first quantization estimate ($\hat{q_1}$) (line 12). Table 1 lists the notations used in this paper for clarity and quick reference. The proposed selection strategy achieves accurate first quantization estimates as demonstrated in detail in the next section.

5. RESULTS AND DISCUSSIONS

MATLAB 8.1 is used for experimentation. The proposed method is tested considering double compressed JPEG images, obtained from UCID v2 [14] dataset which contains

Table 1: Table of Notations

NOTATION	DEFINITION
$\hat{q_1}$	First quantization estimate
q_2	Second quantization value
H_{real}	Histogram of double compressed JPEG image under consideration
H_{q1s}	Synthetic histogram formed in DCT histogram comparison step
q_{1s}	List of probable first quantization estimates
f_{q1s}	List containing error function values for existing multiples of q_2
h_{q1s}	List containing values obtained in DCT histogram comparison stage
f_{mq1s}	List containing error function values for missing multiples of q_2
F_{q1s}	List utilized for localizing the first quantization estimate
q_{1ms}	Set containing probable first quantization estimates for multiple minimum priority scenario

Table 2: ($QF_1 = 50, QF_2 = 80$), $q_2 = 5$ of (1, 0) coefficient for ucid01338.tif

q_{1s}	6	8	9	10	11	12
f_{q1s}	155	2825	3100	4245	5130	160
h_{q1s}	339	1559	1469	1839	2163	351
f_{mq1s}	70	160	190	205	225	395
$Priority\_f_{q1s}$	1	3	4	5	6	2
$Priority\_h_{q1s}$	1	4	3	5	6	2
$Priority\_f_{mq1s}$	6	5	4	3	2	1
F_{q1s}	8	12	11	13	14	5

1338 uncompressed TIFF images with different scene content and settings. From this a compressed dataset has then been built with quality factors (QF_1, QF_2) in the range 50 to 100 in steps of 10 under the assumption $q_1 > q_2$. Based on this variation in quality factor, each uncompressed image will form 15 double compressed images, thereby forming a dataset which comprises of 20,070 (1338 × 15) double compressed JPEG images. This research work is carried out in a reproducible manner and the MATLAB code needed to reproduce the presented results is available at https://sites.google.com/site/manishokade/publications.

5.1 Validation

The proposed method is validated considering the DC coefficient and two of the AC coefficients in order to demonstrate the accurate estimation capability of the first quantization value. Table 2 shows the estimation steps for AC coefficient (1, 0) i.e. 3^{rd} DCT coefficient when 'ucid01338.tif' is the image under consideration. Galvan's method [7] considers the absolute minimum of $\langle h_{q1s} \rangle$ which in this case is # 339 and the corresponding $\langle q_{1s} \rangle$ entry # 6 is chosen as the first quantization estimate which is incorrect as verified from standard JPEG quantization table. The proposed method

Table 3: ($QF_1 = 80, QF_2 = 100$), $q_2 = 1$ of (0,0) coefficient for ucid01338.tif

q_{1s}	2	3	4	5	6	8
f_{q1s}	254	579	2626	3628	904	5340
h_{q1s}	651	851	2913	3967	897	4005
f_{mq1s}	687	958	1267	1514	2361	2509
$Priority\_f_{q1s}$	1	2	4	5	3	6
$Priority\_h_{q1s}$	1	2	4	5	3	6
$Priority\_f_{mq1s}$	6	5	4	3	2	1
F_{q1s}	8	9	12	13	8	13

Table 4: ($QF_1 = 70, QF_2 = 80$), $q_2 = 8$ of (4,0) coefficient for ucid00003.tif

q_{1s}	9	10	11	12	13	14
f_{q1s}	0	8	16	8	8848	7560
h_{q1s}	309	631	149	955	2151	1929
f_{mq1s}	48	80	104	136	72	112
$Priority\_f_{q1s}$	1	2	3	2	5	4
$Priority\_h_{q1s}$	2	3	1	4	6	5
$Priority\_f_{mq1s}$	6	4	3	1	5	4
F_{q1s}	9	9	7	7	16	11

in which the three prioritized lists i.e. $\langle Priority\_h_{q1s} \rangle$, $\langle Priority\_f_{q1s} \rangle$, $\langle Priority\_f_{mq1s} \rangle$ have been used followed by a final priority assignment $\langle F_{q1s} \rangle$ indicates a minimum value of # 5, which when mapped to $\langle q_{1s} \rangle$ gives the estimate of first quantization value as # 12. This is the correct estimate as verified from the standard JPEG quantization table (QF_1 = 50, QF_2 = 80) for (1, 0) coefficient. Table 3 shows the estimation steps for DC coefficient (0, 0) i.e. 1^{st} DCT coefficient with 'ucid01338.tif' taken under consideration. Galvan's method [7] chooses the absolute minimum from $\langle h_{q1s} \rangle$ i.e. # 651 and picks its corresponding entry from $\langle q_{1s} \rangle$ i.e. # 2 as the first quantization estimate which is erroneous as verified from the standard JPEG quantization table. On the other hand, the proposed method utilizes the minimum value of $\langle F_{q1s} \rangle$ to localize first quantization estimate ($\hat{q_1}$). However, as observed from Table 3, $\langle F_{q1s} \rangle$ contains multiple values having equal minimum priority (i.e. #8). The q_{1s} values corresponding to these multiples i.e. $\langle 2, 6 \rangle$ are extracted and stored in a list $\langle q_{1ms} \rangle$ (line 8, Algorithm 1). Now the list $\langle q_{1ms} \rangle$ is tested for integer multiples and the value corresponding to the lower entry is removed i.e. # 2 is removed from $\langle q_{1ms} \rangle$ since its an integer multiple of #6. Finally, $\langle q_{1ms} \rangle$ contains # 6 with corresponding $\langle Priority\_h_{q1s} \rangle$ entry being # 3. Since, the $\langle q_{1ms} \rangle$ value corresponding to minimum of selected entries of $\langle Priority\_h_{q1s} \rangle$ gives the first quantization estimate, #6 is the estimated first quantization value by the proposed method. This is the correct estimate as verified from standard JPEG quantization table (QF_1=80, QF_2=100). Table 4 shows the estimation steps for AC coefficient (4,0) i.e. the 11^{th} DCT coefficient with image under consideration being 'ucid00003.tif'. The proposed method utilizes the minimum value of $\langle F_{q1s} \rangle$ to localize first quantization estimate ($\hat{q_1}$). However, as observed from Table 4, $\langle F_{q1s} \rangle$ contains multiple values having equal minimum priority (i.e. #7). The q_{1s} values corresponding

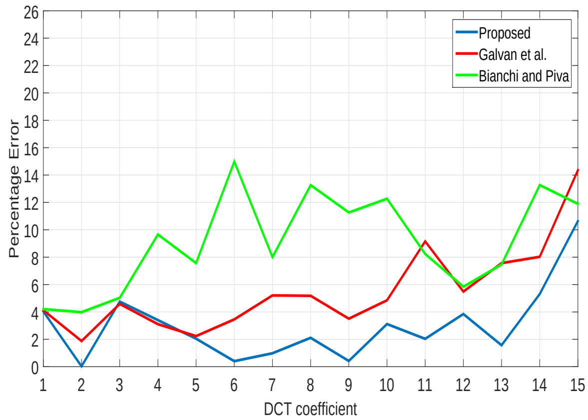


Figure 4: Percentage error v/s DCT coefficient position (zig zag scan) for first quantization ($\hat{q}_1$) estimation.

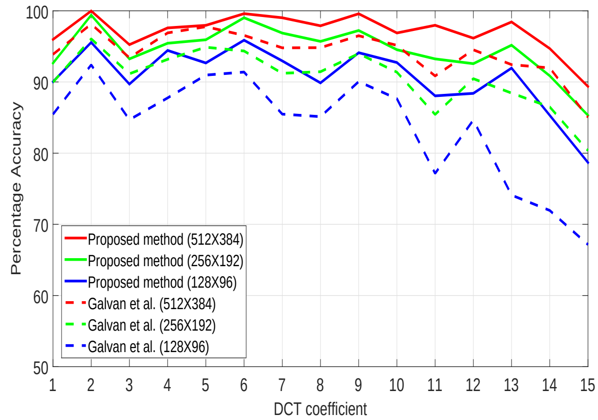


Figure 5: Percentage accuracy v/s DCT coefficient position (zig zag scan) at varying patch sizes for first quantization ($\hat{q}_1$) estimation.

to these multiples i.e. $\langle 11, 12\rangle$ are extracted and stored in a list $\langle q_{1ms}\rangle$ (line 8, Algorithm 1). Now the list $\langle q_{1ms}\rangle$ is tested for integer multiples. Since, multiplicity criteria is not being satisfied by these elements, $\langle q_{1ms}\rangle$ list along with corresponding $\langle Priority\_h_{q1s}\rangle$ values i.e. $\langle 1, 4\rangle$ is considered. Now the minimum of $\langle Priority\_h_{q1s}\rangle$ is considered which evaluates to # 1 and its corresponding $\langle q_{1ms}\rangle$ entry i.e. # 11 is the estimated first quantization value by the proposed method. This is the correct estimate as verified from standard JPEG quantization table (QF_1=70, QF_2=80). It is to be noted that Galvan's method [7] which considers the absolute minimum of $\langle h_{q1s}\rangle$ i.e. # 149 also converges to # 11 which is the correct estimate.

5.2 Performance Evaluation

This section is devoted towards analyzing the performance of the proposed method with respect to the DCT coefficient position. We experimented on the Double JPEG dataset comprising of 20070 double compressed JPEG images. The proposed method was verified for estimation of first 15 DCT coefficients considered in zig-zag order, in-line with the JPEG ordering strategy [i.e. (0,0)→1, (0,1)→2, (1,0)→3, (2,0)→4 . . . (0,4)→15]. First 15 coefficients were considered instead of all 64 coefficients because of the fact that majority of the higher order terms are zero, thereby losing there significance. Table 5 shows the percentage error obtained for first quantization matrix estimation for first 15 DCT coefficients at varying Quality Factors (QF_1, QF_2). As observed for DC coefficient i.e. (0,0), the percentage errors for first quantization estimation is relatively high for higher values of QF_2 due to blocking artifact impacting the DC value as well as the DCT histogram comparison step i.e. Eq. (7) accumulating the rounding, truncation and quantization errors thereby estimating a value close to the correct first quantization value. For higher AC frequencies, the percentage error for first quantization value estimation is high when values of QF_1 and QF_2 are low as blocking artifacts significantly affect DCT coefficients with low quality factor. Further, the number of histogram bins available reduces with the decrease in quality factor, thereby reducing the available statistics and increasing the percentage error.

Comparative analysis of the proposed method with Galvan et al.'s [7] and Bianchi and Piva's method [1] is carried out next. Figure 4 shows a plot of percentage error v/s the DCT coefficient position for the proposed method, Galvan et al.'s method [7] and Bianchi and Piva's method [1]. It is observed that the DC coefficient has high error percentage in comparison to the next 10 AC coefficients. This occurs mainly due to blocking artifacts significantly influencing the DC coefficient since the DC value has maximum weightage thereby bearing the maximum block artifact error. Observation from Figure 4 on the behaviour of the AC coefficients indicate low percentage error for initial coefficients followed by increase in error values for the later coefficients. However, it is worth mentioning that this trend is not followed by coefficient numbers # 3 and # 6 due to them being related to horizontal edge position and vertical edge position, respectively thereby being more sensitive to the cropping errors as well as blocking artifacts. Similar observations have been reported by Galvan et al. [7] as observed from Figure 4. However, their method has high error percentage values in comparison to the proposed method due to few instances of split noise like the scenario depicted in Figure 2 not being accounted coupled with presence of rounding off errors present in the DCT histogram comparison step. On the other hand, since the proposed method accounts for all the instances of split noise by utilizing the three prioritized lists i.e. $\langle Priority\_f_{q1s}\rangle$, $\langle Priority\_h_{q1s}\rangle$ and $\langle Priority\_f_{mq1s}\rangle$, we are able to arrive at an accurate estimate for the first quantization value in the DJPEG scenario. The highest error percentage is found with Bianchi and Piva's method [1] since they utilize Expectation Maximization (EM) which suffers from parameter initialization and convergence issues along with outlier handling in a globalized way, thereby leading to inaccurate first quantization estimates.

5.3 Robustness Analysis

This section presents the sensitivity analysis for the proposed method against varying patch size of the tampered JPEG images. Figure 5 shows the plot of percentage accuracy versus position of DCT coefficient for varying sizes of the pasted patch. The continuous line is used to denote

Table 5: Study of percentage error for first quantization matrix estimation for first 15 DCT coefficients in zig-zag scan order at varying Quality Factor(QF_1, QF_2) values.

(0,0)		QF_2				
		60	70	80	90	100
QF_1	50	1.57	0.30	0.00	0.00	8.52
	60		4.71	2.99	5.28	6.54
	70			4.18	6.72	12.53
	80				10.86	11.48
	90					0.00

(0,1)		QF_2				
		60	70	80	90	100
QF_1	50	0.00	0.00	0.00	0.00	0.15
	60		0.00	0.00	0.00	0.00
	70			0.00	0.00	0.00
	80				0.00	0.22
	90					0.00

(1,0)		QF_2				
		60	70	80	90	100
QF_1	50	6.81	26.65	0.00	0.00	38.57
	60		7.09	0.00	0.00	0.00
	70			0.08	0.00	0.00
	80				0.00	0.00
	90					0.00

(2,0)		QF_2				
		60	70	80	90	100
QF_1	50	31.06	0.00	0.00	0.00	12.40
	60		1.35	0.00	0.00	0.00
	70			0.15	0.00	2.91
	80				0.00	3.58
	90					0.00

(1,1)		QF_2				
		60	70	80	90	100
QF_1	50	16.44	14.66	0.00	0.00	4.55
	60		4.55	0.00	0.00	0.00
	70			0.00	0.00	0.00
	80				0.00	0.00
	90					0.00

(0,2)		QF_2				
		60	70	80	90	100
QF_1	50	0.12	0.08	0.00	0.00	3.14
	60		0.02	0.00	0.00	2.09
	70			0.00	0.00	0.07
	80				0.00	0.52
	90					0.00

(0,3)		QF_2				
		60	70	80	90	100
QF_1	50	5.76	1.05	0.45	0.08	1.869
	60		1.50	1.05	0.00	0.00
	70			0.15	0.00	0.60
	80				2.09	0.15
	90					0.00

(1,2)		QF_2				
		60	70	80	90	100
QF_1	50	29.07	0.75	0.00	0.00	0.22
	60		0.00	0.38	0.00	0.00
	70			0.00	0.00	0.00
	80				0.00	0.00
	90					0.00

(2,1)		QF_2				
		60	70	80	90	100
QF_1	50	0.45	0.00	0.45	0.00	0.00
	60		5.00	0.37	0.00	0.00
	70			0.00	0.00	0.00
	80				0.00	0.00
	90					0.00

(3,0)		QF_2				
		60	70	80	90	100
QF_1	50	29.64	0.45	0.00	0.00	2.99
	60		0.32	0.52	0.08	0.00
	70			0.22	0.00	7.84
	80				0.00	3.21
	90					0.00

(4,0)		QF_2				
		60	70	80	90	100
QF_1	50	6.43	3.69	3.29	1.72	2.99
	60		5.90	0.67	0.00	1.35
	70			2.24	0.90	0.67
	80				1.34	0.07
	90					0.22

(3,1)		QF_2				
		60	70	80	90	100
QF_1	50	7.47	7.24	34.27	0.22	0.00
	60		7.47	0.522	0.00	0.00
	70			0.30	0.00	0.00
	80				0.08	0.00
	90					0.00

(2,2)		QF_2				
		60	70	80	90	100
$QF1$	50	14.50	1.20	0.45	0.00	1.94
	60		0.97	1.94	0.00	0.00
	70			0.30	0.00	0.00
	80				1.57	0.00
	90					0.00

(1,3)		QF_2				
		60	70	80	90	100
$QF1$	50	12.93	6.96	3.74	0.52	0.08
	60		44.92	8.82	0.08	0.00
	70			0.43	0.15	0.00
	80				0.00	0.08
	90					0.00

(0,4)		QF_2				
		60	70	80	90	100
QF_1	50	24.23	7.85	36.62	11.14	19.73
	60		22.35	24.51	4.49	0.15
	70			5.61	1.50	0.22
	80				0.08	0.07
	90					0.07

the proposed method while the dotted line denotes Galvan's method. Similar patch size is represented by using same color i.e. red represents patch size of 512 $\times$ 384 while green and blue represents patch size of 256 $\times$ 192 and 128 $\times$ 96 respectively. As observed from Figure 5, when the size of the pasted patch decreases, the percentage accuracy drops thereby indicating that smaller patches of tampering are difficult to detect. However, as observations from Figure 5 reveal, the proposed method is more robust for varying patch sizes in comparison to Galvan's method [7] as observed from higher values for percentage accuracy, which is a vital aspect in forensic applications since the integrity of image is being tested.

6. CONCLUSIONS

This paper investigated a robust method for first quantization matrix estimation in case of double compressed JPEG images by improving the DCT histogram selection step. Analysis carried out in this paper showed that existing methods were only accounting for the split noise and residual noise effects when bin of first quantization was equidistant from two neighbouring bins of second quantization thereby leading to incorrect first quantization estimations. In addition, they did not account for the split noise and residual noise introduced while forming synthetic histogram in DCT histogram comparison step. The DCT histogram selection step was improved by accounting for the different instances of split noise and residual noise via a novel relative priority assignment and selection strategy applied on error function values of both existing and missing multiples of second quantization step along with the histogram comparison values, thereby achieving robust first quantization estimates. Experimental validation and comparison with two of the state-of-the-art methods showed the superiority of the proposed method for forensic application. Our future work is focussed on estimation of first quantization matrix in DJPEG images where tampering in the form of resizing has taken place.

7. REFERENCES

[1] T. Bianchi and A. Piva. Image forgery localization via block-grained analysis of JPEG artifacts. *IEEE Transactions on Information Forensics and Security*, 7(3):1003–1017, June 2012.

[2] G. Cao, Y. Zhao, R. Ni, and X. Li. Contrast enhancement-based forensics in digital images. *IEEE Transactions on Information Forensics and Security*, 9(3):515–525, March 2014.

[3] Y. L. Chen and C. T. Hsu. Detecting recompression of JPEG images via periodicity analysis of compression artifacts for tampering detection. *IEEE Transactions on Information Forensics and Security*, 6(2):396–406, June 2011.

[4] H. Farid. Exposing digital forgeries from JPEG ghosts. *IEEE Transactions on Information Forensics and Security*, 4(1):154–160, March 2009.

[5] J. Fridrich. Digital image forensics. *Signal Processing Magazine, IEEE*, 26(2):26–37, March 2009.

[6] J. Fridrich, M. Goljan, and D. Hogea. Steganalysis of JPEG images: Breaking the F5 algorithm. In *5th International Workshop on Information Hiding*, IH '02, pages 310–323, 2003.

[7] F. Galvan, G. Puglisi, A. R. Bruna, and S. Battiato. First quantization matrix estimation from double compressed JPEG images. *IEEE Transactions on Information Forensics and Security*, 9(8):1299–1310, Aug 2014.

[8] E. Kee, M. K. Johnson, and H. Farid. Digital image authentication from JPEG headers. *IEEE Transactions on Information Forensics and Security*, 6(3):1066–1075, Sept 2011.

[9] T. V. Lanh, K. S. Chong, S. Emmanuel, and M. S. Kankanhalli. A survey on digital camera image forensic methods. In *2007 IEEE International Conference on Multimedia and Expo*, pages 16–19, July 2007.

[10] W. S. Lin, S. K. Tjoa, H. V. Zhao, and K. J. R. Liu. Digital image source coder forensics via intrinsic fingerprints. *IEEE Transactions on Information Forensics and Security*, 4(3):460–475, Sept 2009.

[11] J. Lukas and J. Fridrich. Estimation of primary quantization matrix in double compressed JPEG images. In *Digit. Forensic Res. Workshop (DFRWS)*, pages 5–8, 2003.

[12] W. Luo, J. Huang, and G. Qiu. JPEG error analysis and its applications to digital image forensics. *IEEE Transactions on Information Forensics and Security*, 5(3):480–491, Sept 2010.

[13] T. Pevny and J. Fridrich. Detection of double-compression in JPEG images for applications in steganography. *IEEE Transactions on Information Forensics and Security*, 3(2):247–258, June 2008.

[14] G. Schaefer and M. Stich. UCID: an uncompressed color image database. In *Proc. SPIE*, volume 5307, pages 472–480, 2003.

[15] M. C. Stamm, M. Wu, and K. J. R. Liu. Information forensics: An overview of the first decade. *IEEE Access*, 1:167–200, 2013.