An Automatic JPEG Ghost Detection Approach for Digital Image Forensics

Sepideh Azarian-Pour
Electrical Engineering Department
Sharif University of Technology
Tehran, 14588-89694, Iran
Email: sepide.azarian@gmail.com

Massoud Babaie-Zadeh
Electrical Engineering Department
Sharif University of Technology
Tehran, 14588-89694, Iran
Email: mbzadeh@sharif.edu

Amir Reza Sadri
Electrical Engineering Department
Isfahan University of Technology
Isfahan, 84156-83111, Iran
Email: ar.sadri@ec.iut.ac.ir

***Abstract*—In this paper we propose a new automatic method for discriminating original and tampered images based on "JPEG ghost detection" method, which is a subset of format-based image forensics approaches. The inconsistency of quality factors indicates that the photo is a composite one created from at least two different cameras and therefore it is a manipulated photo. Our classification algorithm first extracts the ghost border. Then the image is classified as original or tampered groups by thresholding a distance in feature space.**

I. Introduction

Easy access to digital cameras and photo editing software has resulted in creating counterfeiting images and changing the content of original ones. So the reliability and authenticity of digital media for a court of law is now questionable.

Any processing on images to transform a photograph into a desired image such as adding/ removing people or objects from the image scene, adjusting the brightness and contrast, scaling, rotating some parts of the image is called "image manipulation" or image forgery [1]. Digital Image Forensics (DIF) is an emerging subject which studies tools and methods for distinction of authentic images from digitally manipulated ones [2].

DIF methods are generally divided into two main categories: 1) passive (or blind) methods that use only the image under evaluation (the *"dubious image"*), and 2) active methods that use additional information, for example, the original (unmanipulated) image, or the embedded message in steganography applications [3]. Considering limitations of active approaches and widespread use of blind algorithms, the present paper focuses on passive methods.

In general, passive DIF techniques can be categorized into six different families [4]: 1) format-based methods which analyze inconsistencies in blocking, quality factor or quantization error in some lossy compression formats [5], [6]; 2) statistics-based methods that extract statistical features from the distribution function in each color channel [7]; 3) pixel-based methods which discover sampling history (by studying the adjacent pixel correlations) and reveal cloning, duplicating, resampling artifacts and copy-move regions [8]; 4) camera-based methods in which camera defects and imperfections are exploited for modeling the camera characteristics [9]; 5) geometric-based methods that make measurements according to perspective modeling and lens options [10]; 6) physics-based techniques that estimate 3-D lighting environment using the brightness gradient [11].

One of the format-based techniques is JPEG ghost detection [5] which can detect local forgery instead of global authentication [12]. This method estimates the quality factor of JPEG compression in every region by studying the compression artifacts (such as double-quantization effect in JPEG compression format which reveals post-processing of the image in a computer software and resaving it in JPEG format).

The JPEG ghost detection method has the advantage that it works for tampering detection of low-quality images [5]. However, this method suffers from a number of disadvantages. One of the fundamental shortcomings of this method is its lack of automation (method of [5] needs a manual search for the ghost). Another disadvantage is a problem with non-aligned Discrete Cosine Transform (DCT) blocks, that are involved in this method, as will be discussed in Section II.

In this paper, we are going to modify the JPEG ghost detection method by adding some post-processing and iterations to it. As a result, both the above mentioned limitations of the method of [5] is removed, that is, firstly the method will be automatic, and secondly, it will not have the problem of non-aligned DCT blocks.

The paper is organized as follows. In section II, we review ordinary JPEG ghost detection method of [5] and explain its disadvantages. The proposed method is explained in section III. Finally, section IV is devoted to experimental results.

II. A Review on JPEG Ghost Detection Method

The JPEG file format [13] has become the popular format of almost all compact cameras [1], [13]. As we have summarized in Fig. 1, JPEG compression contains some details on the choice of quantization tables and Huffman code-words. Different computer software and different camera models use different quantization tables and Huffman code-words. Even in one camera or software, the choice of different compression qualities or resolutions results in various values for these parameters. In this way, the set of these parameters is called fingerprint or signature [6], [14]. In a JPEG compressor, Q_Y, Q_{C_b}, Q_{C_r} tables of Fig. 1 result in a quality factor between $1, 2, \ldots, 100$ percent. Small elements of quantization tables make higher quality factors.

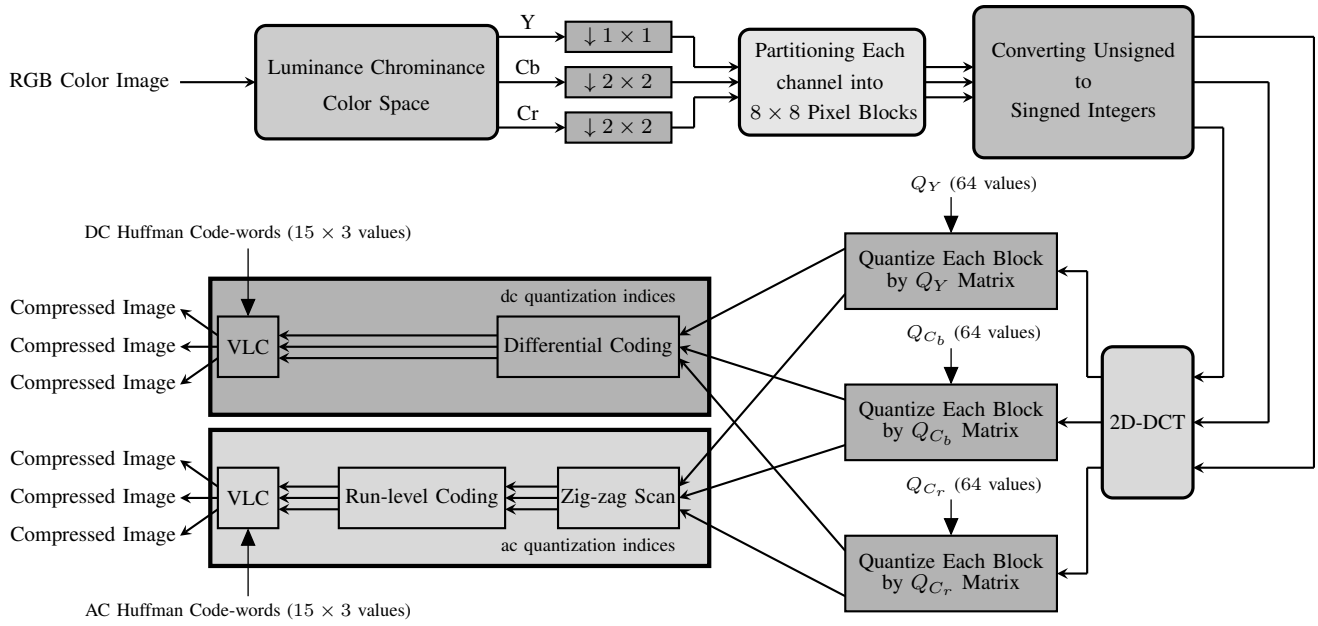


Figure 1: Standard JPEG Compression Scheme (this diagram is our summarization of the JPEG compression standard explained in [13]).

In a composite forged photo, often two JPEG images with different quantization tables were spliced together using a photo editing software. Then, the final image is resaved in JPEG file format by another quantization table which uses usually a higher quality quantization table. As a result, some parts of this image are double-quantized which are known as the tampered regions and other parts of the image are quantized by a higher quality factor which are original parts of the image. Such a tampered image is briefly called a double-quantized image. A simple diagram has been shown in Fig. 2, to illustrate this forgery scenario.

For detecting double-quantized parts of an image, [5] proposes the following approach. First, note that in the quantization block of a JPEG encoder, for single-quantized regions, each DCT coefficient, c, is quantized by a quantization step-size, s_1, from the 8×8 quantization table to yield c_{sq}:

$$c_{sq} = s_1 \lfloor \frac{c}{s_1} \rfloor. \tag{1}$$

Now, consider the double-quantized coefficient, c_{dq} which is quantized by step s_0 followed by quantization by step s_1, according to the two following equations:

$$c_0 = s_0 \lfloor \frac{c}{s_0} \rfloor \tag{2}$$

$$c_{dq} = s_1 \lfloor \frac{c_0}{s_1} \rfloor. \tag{3}$$

Compression history of c_{dq} can be determined by a subsequently quantization by a step-size s_2 to yield c_2:

$$c_2 = s_2 \lfloor \frac{c_{dq}}{s_2} \rfloor. \tag{4}$$

In [5] it is shown that assuming $s_1 < s_0$, then the energy function, $|c_{dq} - c_2|^2$ versus s_2 has a global minimum (zero) at $s_2 = s_1$ and a local minimum at $s_2 = s_0$.

Now, consider an image I which is compressed in JPEG file format at quality factor q_0 followed by another compression at quality factor q_1 ($q_1 > q_0$). By comparing the dubious image, I, and its JPEG-recompressed counterpart at quality factor q_2 in three color channels and calculating the sum of difference squares, an image d is obtained which is called difference energy image [5]

$$d(x, y, q_2) = \frac{1}{3} \sum_{c \in \{\mathrm{R,G,B}\}} (I(x, y, c) - I_{q_2}(x, y, c))^2 \tag{5}$$

where $I(x, y, c)$, in which $c = \mathrm{R}, \mathrm{G}, \mathrm{B}$, denotes each color channel of the image I and $I_{q_2}(x, y, c)$ denotes resaved version of $I(x, y, c)$ in quality factor of q_2. Moreover, for compensating the texture effect [5] in high frequency details or plain objects, the difference image is smoothed as follows:

$$\delta(x, y, q_2) = \frac{1}{3w^2} \sum_{c \in \{R,G,B\}} \sum_{i=0}^{w-1} \sum_{j=0}^{w-1} (I(x+i, y+j, c) - Iq_2(x+i, y+j, c))^2 \tag{6}$$

where the window size, w, is typically 16 [5]. Then, $\delta(x, y, q_2)$ is normalized into the interval $[0, 1]$ as in (7)

$$d(x, y, q_2) = \frac{\delta(x, y, q_2) - \min_q[\delta(x, y, q_2)]}{\max_q[\delta(x, y, q_2)] - \min_q[\delta(x, y, q_2)]}. \tag{7}$$

Now d is a grayscale image which depends on q_2. In case $q_2 = q_0$ tampered regions become dark and discriminable, as shown in Fig. 3b. Note that the local minimum of the energy plot in Fig. 3d occurs in $q_2 = q_0$. For $q_2 = q_1$ the whole image is dark (Fig. 3c) and make global minimum as shown in Fig. 3d.

One of the main limitation of the above-mentioned approach is the constraint $q_1 > q_0$ (this limitation will remain in

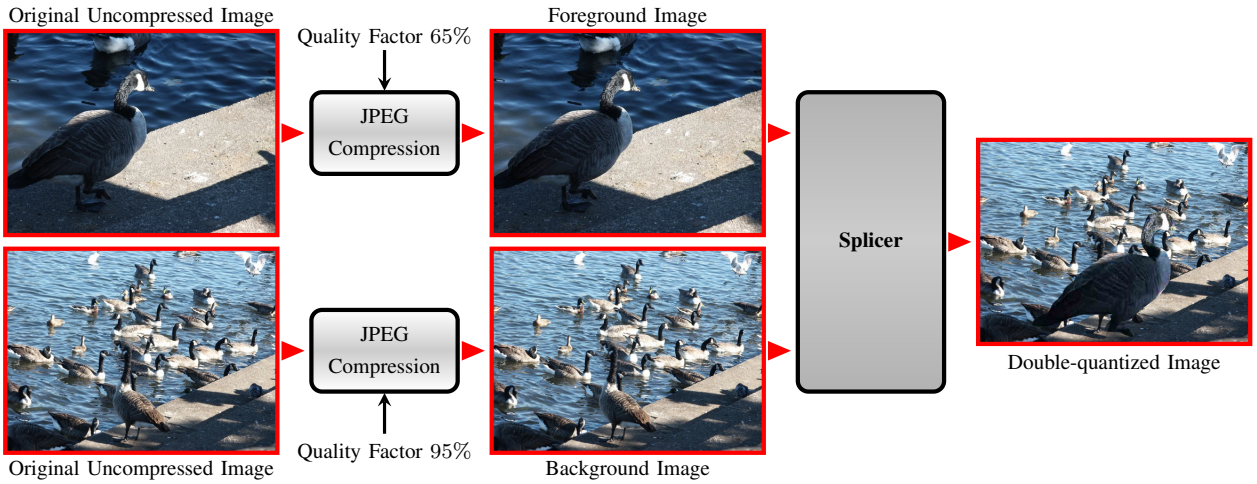


Figure 2: Hypothetical scenario for double-quantized image, composed of two different quality images.

our method, too). Otherwise, the ghost does not appear [5]. Additionally, in order to do an exact ghost detection, it is necessary for DCT grids of the JPEG of both original and tampered parts of the images to be aligned. But an image forger often needs to shift the objects of the image horizontally and/or vertically. So, this approach fails in this case. In other words, there is only one case among all $8 \times 8 = 64$ cases in which DCT grids are aligned and this method only works for this situation. The other main flaw of [5] is that there is no discussion in it on how the ghost has identified; because tampered images of [5] are artificial and the tampered region is always the central square of size 200×200 pixels in it.

III. Our Proposed Method

If the DCT grids of original and tampered images are not aligned, then with the method of [5] one has to manually run the algorithm for all possible shifts (8 pixels horizontal and 8 pixels vertical) and for all quality factors. For example, for 100 different quality factors, this results in 6400 manual runs of the algorithm. To solve this problem, we will use a segmentation algorithm to extract ghost borders, then we will propose a distance criterion to measure how much the JPEG ghost is different from the rest of the image. So, using this distance measure, we automatically calculate all the distance measures of all the above 6400 different cases, and take the decision based on the maximum value.

The steps of our method are explained in the following three subsections.

A. Step 1: Double Compression

In this step, firstly an $(m+d_x)\times(n+d_y)$ image $I'(x,y)$ is created by zero padding on the dubious $m \times n$ image $I(x,y)$, where $d_x = 0, 1, \ldots, 7$ and $d_y = 0, 1, \ldots, 7$ are the horizontal and vertical shifts in it. Now the ordinary JPEG ghost method is applied on the image $I'(x,y)$ similar to (6) as follows

$$\delta_{(q_2,dx,dy)}(x,y) \triangleq \frac{1}{3w^2} \sum_{c\in\{R,G,B\}} \sum_{i=0}^{w-1}\sum_{j=0}^{w-1} (I'(x+i,y+j,c) - I'_{q_2}(x+i,y+j,c))^2 \quad (8)$$

where $q_2 = 1, 2, \ldots, 100$ is the quality factor of the double compression block. In our simulations, we have used $w = 16$, as in [5]. Now similar to (7), difference image is normalized:

$$d_{(q_2,dx,dy)}(x,y) \triangleq (\delta_{(q_2,dx,dy)}(x,y) - \min_q[\delta_{(q_2,dx,dy)}(x,y))/ (\max_q[\delta_{(q_2,dx,dy)}(x,y)] - \min_q[\delta_{(q_2,dx,dy)}(x,y)]). \quad (9)$$

B. Step 2: Ghost Segmentation

In order to calculate the distance between the original and the tampered regions, it is necessary to identify JPEG ghost area. Thus, the SE-MinCut segmentation method of [15] is employed to obtain two segments (class-0 contains ghost area and class-1 for the rest of the image). Our reason for selecting this approach is its robustness against fractal noise, because JPEG ghost images are similar to images which are corrupted by fractal noise [15]. The output of this step is a binary indexed image, $Y(x,y)$.

C. Step 3: Classification

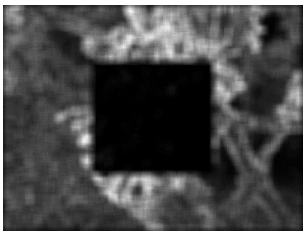
After performing the above segmentation, each pixel of $I(x,y)$ is labeled to belong to class-0 (ghost area) or class-1 (the rest of the image). Then, we need to define a criterion to decide whether or not the whole image $I(x,y)$ is a tampered image. To do so, we use the following one dimensional Bhattacharyya distance [16] between the classes 0 and 1:

$$B = \frac{1}{2}\ln\frac{\sigma_0^2+\sigma_1^2}{2\sigma_0\sigma_1} + \frac{(\mu_0-\mu_1)^2}{4(\sigma_0{}^2+\sigma_1{}^2)} \quad (10)$$

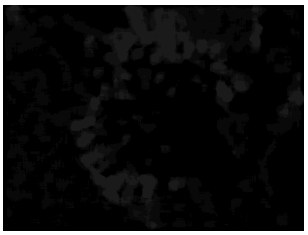
where μ_0, μ_1, $\sigma_0{}^2$, $\sigma_1{}^2$ are the mean and variances of the elements of class-0 and class-1, respectively.

After applying all these three steps 6400 times on the dubious image, one set of the parameters (q_2, d_x, d_y) maximizes the above distance criterion. These parameters are called ($q_{2,m}$, $d_{x,m}$, $d_{y,m}$) and the resulting distance is called $D_{\max}$. Now the classifier checks $D_{\max} > \text{Th}$, where Th is a threshold. Finally, if $D_{\max} > \text{Th}$ holds, then it is asserted that:

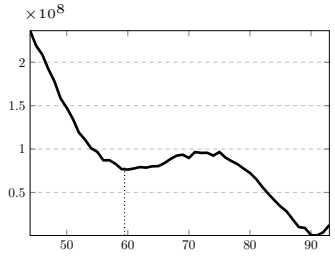
(a) A forged photo which is created by resaving central 200×200 region at quality factor of 60%. The quality factor of the original version of this image is 90% [17].



(b) The difference between the image and its JPEG-recompressed counterpart at quality factor of 60%, which contains JPEG ghost.



(c) The difference between the image and its JPEG-recompressed counterpart at quality factor of 90%, which has minimum energy among all difference images.



(d) The energy of the difference image versus q_2. The local minimum occurs at 60% (see Fig. 3b) and the global minimum occurs at 90% (see Fig. 3c)

Figure 3: Difference images at each quality factor reveals inconsistent regions.

1) The image $I(x, y)$ is a composite forged photo with different quality factors.
2) Class-0, which is compressed at lower quality than class-1 is the tampered region or double-compressed region. This is the final reported segmentation of the algorithm.
3) The quality factor of the tampered region is $q_{2,m}$.
4) $d_{x,m}$ and $d_{y,m}$ indicate the DCT grid misalignment.

The structure of our method is shown in Fig. 4 in the form of flowcharts.

IV. Evaluation of Results

For our simulations, we have used Uncompressed Color Image Database (UCID) [17]. This database contains 1338 TIFF images of size 512×384 pixels, and includes scenes of nature, people, objects, wildlife, cities, monuments, etc. According to the image forgery process, JPEG images in different qualities are needed to be spliced together. For this purpose, 1300 images have been saved as JPEG with 100 different quality factors. Each "quality group" includes 13 random images.

For creating tampered images, the background is chosen from q_1 group and the foreground is chosen from the lower quality q_0 group which is cropped with a random mask and inserted as the tampered region in the background. Thus 4950 groups, including 13 tampered images are obtained. The crop mask which is employed here is used as Ground Truth (GT) later, in training and evaluation.

All experiments were carried out in MATLAB R2014a using Intel® Core™ i7-2670QM (2.20GHz) processor and 4GB RAM.

A. Segmentation Results

In each case by comparing the final segmentation result and GT, the values of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) were determined and the values of the accuracy [16] ($\frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}$) and the precision [16] ($\frac{\text{TP}}{\text{TP+FP}}$) were also obtained. The mean value of accuracy and precision were respectively 97.73% and 91.01%.

B. Training Classifier

Since the classification step is a simple thresholding on the distance measure, the training step is determining the value of Th. For this purpose, we applied 1000 original and 1000 tampered photos on our algorithm to obtain final segmentation results. Then, we set the Th in a way which minimizes the classification error rate on our database (FP and FN are equal) which resulted to $\text{Th} = 0.19$.

C. Sensitivity and Specificity of Final Classification

Having determined TN and FP for classifying authentic images in the previous subsection, the specificity [16] of our algorithm ($\frac{\text{TN}}{\text{TN+FP}}$) can be calculated. Figure 5a shows this specificity versus the quality factor of the original images. Furthermore, the sensitivity [16] of our algorithm for classifying tampered images ($\frac{\text{TP}}{\text{TP+FN}}$) depends on both q_1 and q_0. Figure 5b depicts the average value of sensitivity versus $\Delta q \triangleq q_1 - q_0$ (for each Δq, averaging is done over the values of sensitivity for all values of q_1 and q_0 that give rise to that Δq). It is seen that for $\Delta q > 22$, the averaged sensitivity is more than 95.15%. Note that the sensitivity does not depend only on Δq; it depends on q_1 and q_0, too. Especially, our experience with the algorithm showed us that for very low and very high values of q_1, only a small quality difference creates a clear JPEG ghost. For example, $q_1 = 98, q_0 = 97$ results in $D_{\max} = 0.85$ which results in a correct detection (note that our threshold was $\text{Th} = 0.19$). As another example, $q_1 = 8, q_0 = 3$ results in $D_{\max} = 0.61$ and again a correct classification. Figure 6 shows the values of Δq versus q_1 that result in sensitivity equal to 90%.

V. Conclusion

Our paper proposed a new technique for automatic image forensics, based on JPEG ghost detection. JPEG ghost is a symptomatic of image manipulation in which the quality factors of two components are inconsistent. This inconsistency is revealed by a simple difference between the image and its JPEG-recompressed counterpart. Detecting ghost by manual search can be tedious and error-prone. In our method, after applying an ordinary JPEG ghost detection method, the ghost borders are extracted by the SE-MinCut segmentation

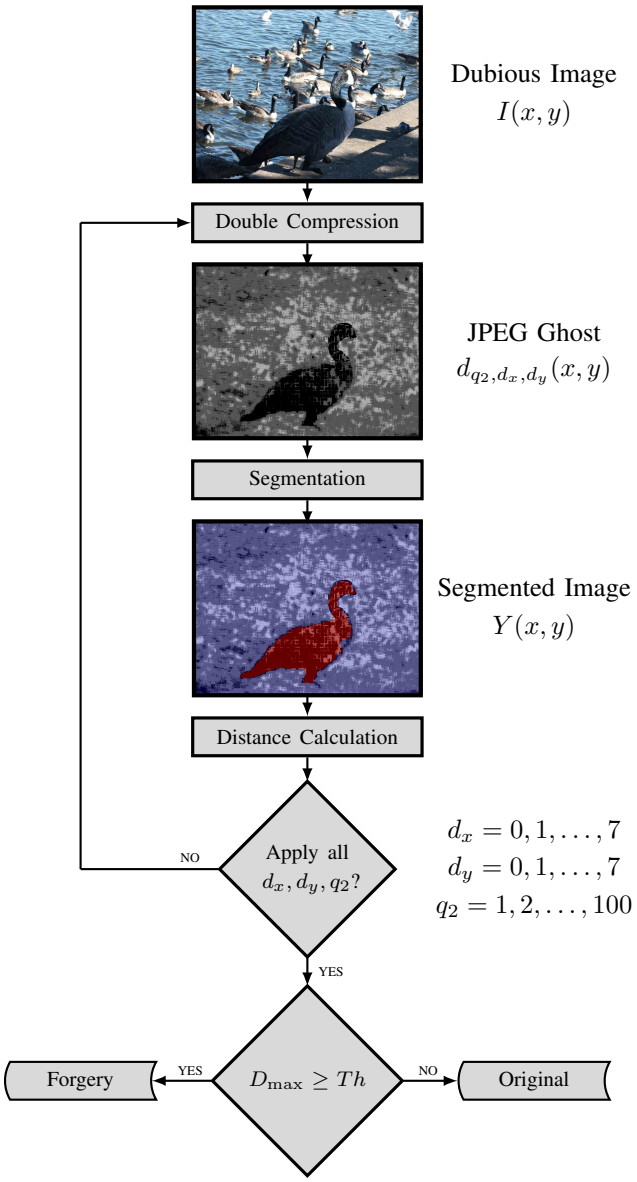
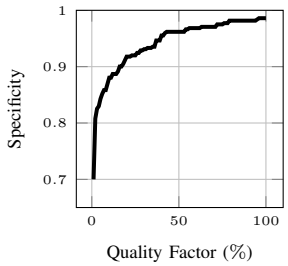
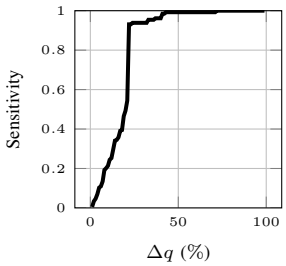


Figure 4: The flowchart of our proposed method on a sample dubious image. Black distinct region in the JPEG ghost image indicates lower quality area. In the segmentation step, class-0 is shown by red color and class-1 by blue. $D_{\max}$ for this sample image is 0.78 which is greater than the threshold. So, it is categorized in the forged group.

algorithm and then the classifier compares the Bhattacharyya distance of the two classes with a specific threshold. Similar to [5], our algorithm has the limitation that it is assuming that a JPEG image is inserted into a higher quality JPEG image.



(a) The specificity of correctly classifying authentic images versus their quality factors.



(b) The sensitivity of correctly classifying tampered images versus the difference of quality factors (Δq).

Figure 5: Classification results.

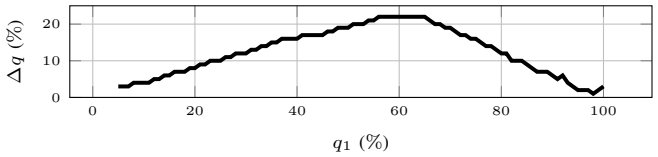


Figure 6: Minimum necessary Δq for satisfying sensitivity greater than 90%, versus background quality factor.

References

[1] H. Farid, "Digital image forensics," *Scientific American*, vol. 298, no. 6, pp. 66–71, May 2008.

[2] A. Redi, W. Taktak, and J. Dugelay, "Digital image forensics: a booklet for beginners," *Multimedia Tools and Applications*, vol. 51, no. 1, pp. 133–162, May 2011.

[3] S. Katzenbeisser and P. Fabien, *Information hiding techniques for steganography and digital watermarking*. Artech house, 2000.

[4] H. Farid, "Image forgery detection," *IEEE Signal Processing Magazine*, vol. 26, no. 2, pp. 16–25, Mar. 2009.

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

[6] E. Kee, M. K. Johnson, and H. Farid, "Digital Image Authentication from JPEG Headers," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 1066–1075, Feb. 2011.

[7] H. Farid and S. Lyu, "Higher-order wavelet statistics and their application to digital forensics," *in Proceedings Conference Computer Vision and Pattern Recognition Workshop*, p. 94. 2003.

[8] A. Fridrich, B. Soukal, and A. Lukas, "Detection of copy-move forgery in digital images," *in Proceedings on Digital Forensic Research Workshop*, 2003.

[9] J. Lukas, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, June 2006.

[10] F. Obrien and H. Farid, "Exposing photo manipulation with inconsistent reflections," *ACM Transactions on Graphics*, vol. 31, no. 1, 2012.

[11] K. Johnson and H. Farid, "Exposing digital forgeries by detecting inconsistencies in lighting," *in Proceedings of the 7th workshop on Multimedia and security*, pp. 1–10, 2005.

[12] F. Zach, C. Riess, and E. Angelopoulou, "Automated image forgery detection through classification of JPEG ghosts," *Pattern Recognition*, vol. 7476, pp.185–194, 2012.

[13] K. Sayood, *Introduction to Data Compression*. Elsevier, Newnes, MA, 2012, pp. 410–423.

[14] M. Goljan, J. Fridrich, and T. Filler, "Large scale test of sensor fingerprint camera identification," *SPIE Conference on Media Forensics and Security*, 2009.

[15] J. Estrada and A. Jepson, "Benchmarking image segmentation algorithms," *International Journal of Computer Vision*, vol. 85, no. 2, pp. 167–181, May 2009.

[16] S. Theodoridis and K. koutroumbas, *Pattern recognition*. Chapman & Hal, UK: London, 2008, pp. 261–285.

[17] G. Schaefer and M. Stich, "UCID an uncompressed colour image database," *School Computer Science and Mathematics, Nottingham Trent University*, Nottingham, U.K., Technical Report, 2003.