

THE COST OF JPEG COMPRESSION ANTI-FORENSICS

G. Valenzise, M. Tagliasacchi, S. Tubaro

Dipartimento di Elettronica e Informazione
Politecnico di Milano, Italy

ABSTRACT

The statistical footprint left by JPEG compression can be a valuable source of information for the forensic analyst. Recently, it has been shown that a suitable anti-forensic method can be used to destroy these traces, by properly adding a noise-like signal to the quantized DCT coefficients. In this paper we analyze the cost of this technique in terms of introduced distortion and loss of image quality. We characterize the dependency of the distortion on the image statistics in the DCT domain and on the quantization step used in JPEG compression. We also evaluate the loss of quality as measured by means of a perceptual metric, showing that a perceptually-optimized version of the anti-forensic method fails to completely conceal the forgery. Our conclusion is that removing the traces of the JPEG compression history could be much more challenging than it might appear, as anti-forensic methods are bound to leave characteristic traces.

Index Terms—digital image forensics; anti-forensics; JPEG compression

1. INTRODUCTION

The increased possibility to tamper with digital content has lately cast doubts on the traditional trust in images as representation of the reality. To re-establish part of this credibility, a number of digital image forensic techniques have been proposed to detect possible image forgeries. These methods do not rely on extrinsic information such as metadata or watermarks embedded into the image, but analyze the image content in order to find traces and footprints left by specific acquisition, coding or editing operations. This approach has been employed, e.g., for identifying which camera snapped a picture [1, 2]; for reconstructing an image’s history through traces left by coding [3, 4], re-sampling [5], cropping [6] or contrast enhancement [7]; and for offering evidence that an image’s content has been altered [8].

Several image forensic techniques leverage the statistical footprints left by JPEG compression. When an image is compressed using JPEG, the histogram of the quantized discrete cosine transform (DCT) coefficients exhibits a characteristic comb-like shape. This fact has been employed to find the original quantization matrix used to compress the image from the decoded pixels only [3], to identify double JPEG compression [4] and copy-move forgeries [9]. Recently, Stamm et al. [10] have shown that adding noise with a certain distribution to the quantized DCT coefficients is sufficient to remove the statistical traces left by JPEG compression and reconstruct the original coefficient distribution. However, the dithering signal added to destroy the JPEG compression footprints leaves traces in the forged image. Indeed, this anti-forensic tool effectively restores the original distribution of DCT coefficients, but it cannot recover the underlying image content lost during quantization. Therefore, it results in an overall degradation of the doctored image quality.

The objective of this paper is to analyze the cost of anti-forensic methods used to remove statistical traces of JPEG compression. The cost is quantified in terms of introduced distortion and loss of image quality. Specifically, we describe two contributions. First, we analyze the dependency of the mean square error (MSE) distortion introduced by anti-forensic dithering in terms of both the quantization step size and the distribution of the original DCT coefficients. This observation enables the characterization of the footprint left by the anti-forensic method in the DCT domain. Second, we show that the anti-forensic dithering degrades seriously the image quality, measured through an objective metric which is known to be well correlated to perceived quality. To support this observation, we consider a variation of the algorithm in [7], which makes use of a content-dependent perceptual model to insert the dithering signal mostly in regions of the image where it is less likely to be observed. We show that, even in this circumstance, the forgery is not adequately concealed. These results indicate that, especially at low JPEG quality factors, destroying the traces of a previous JPEG quantization is feasible, but it is accompanied by a serious degradation of the doctored image quality.

The rest of the paper is organized as follows. Section 2 summarizes the anti-forensic tool described in [7]. Section 3 analyzes the mean square error distortion introduced by this tool in order to characterize the dependency on the quantization step size and on the image content. Section 4 shows the image quality degradation introduced by the method in [7] and proposes a perceptually-driven insertion of the dithering signal. Finally, Section 5 gives some concluding remarks.

2. ANTI-FORENSIC DITHER

In the JPEG compression standard, a greyscale image is first divided into B non-overlapping pixel blocks of size 8×8 . Then, the DCT of each block is computed. Let X_i^b , $1 \leq b \leq B$, $1 \leq i \leq 64$, denote the i -th coefficient of the b -th block according to some scanning order (e.g. zig-zag). Let $\mathbf{X}_i = [X_i^1, \dots, X_i^B]^T$ denote the set of DCT coefficients of the i -th subband. Each DCT coefficient X_i^b , $1 \leq i \leq 64$, is quantized with a quantization step size q_i as given by the JPEG quantization table. Note that the JPEG quantization table is not specified by the standard. The quantization levels Y_i^b are obtained from the original coefficients X_i^b as $Y_i^b = \text{round}(X_i^b/q_i)$. The quantization levels are entropy coded and written in the JPEG bitstream. When the bitstream is decoded, the DCT values are reconstructed from the quantization levels as $\hat{X}_i^b = q_i Y_i^b$. Then, the inverse DCT is applied to each block, and the result is rounded and truncated in order to take integer values on $[0, 255]$. Since the quantized coefficients \hat{X}_i^b can only assume values that are integer multiples of the quantization step size q_i , the histogram of quantized coefficients of the i -th DCT subband, i.e. $\hat{\mathbf{X}}_i = [\hat{X}_i^1, \dots, \hat{X}_i^B]$, is

composed of a train of spikes at integer multiples of q_i . The process of rounding and truncating the decompressed pixel values perturbs the comb-like distribution of $\hat{\mathbf{X}}_i$, as it can be recovered at the decoder; however, the DCT coefficient values typically remain tightly clustered around integer multiples of q_i , thus revealing that: a) a quantization process has occurred; and b) which was the original quantization step.

The work in [10] proposes to conceal the traces of JPEG compression by filling the gaps in the comb-like distribution of $\hat{\mathbf{X}}_i$ by adding a dithering, noise-like, signal \mathbf{N}_i in such a way that the distribution of the dithered coefficients $\mathbf{Z}_i = \hat{\mathbf{X}}_i + \mathbf{N}_i$ approximates the original distribution of \mathbf{X}_i . The distribution of \mathbf{N}_i depends on the DCT subband index i , since it is related to the distribution of $\hat{\mathbf{X}}_i$. The original AC coefficients ($2 \leq i \leq 64$) are typically assumed to be distributed according to the Laplacian distribution:

$$p_{\mathbf{X}_i}(x) = \frac{\lambda_i}{2} e^{-\lambda_i|x|}, \quad (1)$$

where the decay parameter λ_i typically takes values between 10^{-3} and 1 for natural imagery. In practice, only the JPEG-compressed version of the image is available, and the original AC coefficients \mathbf{X}_i are unknown. Therefore, the parameter λ_i in (1) must be estimated from the quantized coefficients $\hat{\mathbf{X}}_i$, e.g. using the maximum-likelihood method in [11], which will result in an estimated parameter $\hat{\lambda}_i$. According to [10], in order to remove the statistical traces of quantization in $\hat{\mathbf{X}}_i$, the dithering signal \mathbf{N}_i needs to be designed in such a way that its distribution depends on whether the corresponding quantized coefficients $\hat{\mathbf{X}}_i$ are equal to zero. That is, for DCT coefficients quantized to zero:

$$p_{\mathbf{N}_i}(n|\hat{X}_i = 0) = \begin{cases} \frac{1}{c_0} e^{-\hat{\lambda}_i|n|} & \text{if } -\frac{q_i}{2} \leq n < \frac{q_i}{2} \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

where $c_0 = \frac{2}{\hat{\lambda}_i} (1 - e^{-\hat{\lambda}_i q_i/2})$. Conversely, for the other coefficients

$$p_{\mathbf{N}_i}(n|\hat{X}_i = x) = \begin{cases} \frac{1}{c_1} e^{-\text{sgn}(x)\hat{\lambda}_i(n+q_i/2)} & \text{if } -\frac{q_i}{2} \leq n < \frac{q_i}{2} \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

where $c_1 = \frac{1}{\hat{\lambda}_i} (1 - e^{-\hat{\lambda}_i q_i})$. Note that the value of the DCT coefficient \hat{X}_i enters in the definition of the p.d.f. in (3) only through its sign.

For some DCT subbands, all the quantized coefficients may be quantized to zero, and $\hat{\lambda}_i$ cannot be determined. In those cases, the anti-forensic method leaves the reconstructed coefficients unmodified, i.e. $\mathbf{Z}_i = \hat{\mathbf{X}}_i$.

Since there is no general model for representing the distribution of DC coefficients, the anti-forensic dithering signal for the DC coefficient ($i = 1$) is sampled from the uniform distribution

$$p_{\mathbf{N}_1}(n) = \begin{cases} \frac{1}{q_i} & \text{if } -\frac{q_i}{2} \leq n < \frac{q_i}{2} \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

3. ANALYSIS OF THE DISTORTION INTRODUCED BY ANTI-FORENSIC DITHER

The measured MSE distortion ε_i introduced by the dithering signal in the i -th subband is given by

$$\hat{\varepsilon}_i = \frac{1}{64B} \sum_{b=1}^B \sum_{i=1}^{64} |Z_i^b - \hat{X}_i^b|^2 = \frac{1}{64B} \sum_{b=1}^B \sum_{i=1}^{64} |N_i^b|^2 \quad (5)$$

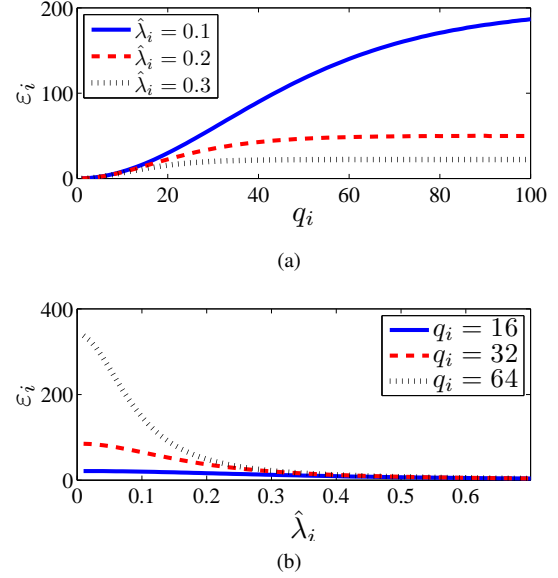


Fig. 1. MSE distortion ε_i introduced by antiforensic dithering for: a) different values of $\hat{\lambda}_i$; b) different values of q_i .

An analytical expression for ε_i can be obtained directly from the distribution of the dithering signal \mathbf{N}_i , as expressed in (2)-(4):

$$\varepsilon_i = \sum_{k=-\infty}^{+\infty} \Pr(\hat{X}_i = kq_i) \int_{-q_i/2}^{+q_i/2} x^2 p_{\mathbf{N}_i}(x|\hat{X}_i = kq_i) dx, \quad (6)$$

where $\Pr(\hat{X}_i = kq_i)$ represents the probability mass function of the quantized DCT coefficients. For AC coefficients, equation (6) can be rewritten according to the definitions given in (2)-(3). That is,

$$\varepsilon_i = m_i^0 \varepsilon_i^0 + (1 - m_i^0) \varepsilon_i^1, \quad \text{for } 1 < i \leq 64 \quad (7)$$

where

$$\varepsilon_i^0 = \int_{-q_i/2}^{+q_i/2} x^2 p_{\mathbf{N}_i}(x|\hat{X}_i = 0) dx, \quad (8)$$

$$\varepsilon_i^1 = \int_{-q_i/2}^{+q_i/2} x^2 p_{\mathbf{N}_i}(x|\hat{X}_i = kq_i) dx, \quad (9)$$

and $m_i^0 = 1 - e^{-\hat{\lambda}_i q_i/2}$ is the fraction of coefficients quantized to zero.

For DC coefficients, the mean square error ε_1 is equal to that of a uniform scalar quantizer, i.e. $\varepsilon_1 = q_1^2/12$. Instead, for AC coefficients, an expression can be found in closed form by solving the integrals in (8) and (9), as a function of the quantization step size and the parameter of the Laplacian distribution, i.e. $\varepsilon_i(q_i, \hat{\lambda}_i)$. Figure 1(a) shows the error ε_i as a function of q_i , for different values of $\hat{\lambda}_i$. A larger value of q_i implies a wider spacing between the spikes in the comb-like distribution of $\hat{\mathbf{X}}_i$. Thus, to restore the original coefficient distribution, a greater amount of noise must be added, and the resulting distortion in the forged image is bound to increase. The growth of mean square error ε_i depends on the value of $\hat{\lambda}_i$. In the Laplacian model (1), a larger $\hat{\lambda}_i$ implies a faster decay, and thus the quantized coefficients $\hat{\mathbf{X}}_i$ will be more clustered around zero (which, in turn, implies that their variance is smaller). When coefficients with a small variance are quantized with a large quantization step, all the

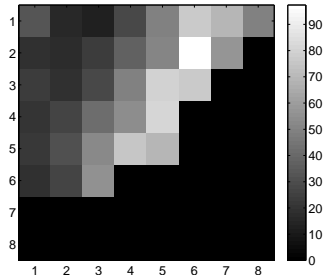


Fig. 2. MSE $\hat{\varepsilon}_i$ in the DCT subbands for the *Lenna* image. The image was originally compressed with a quality factor $Q = 30$ and then the anti-forensic algorithm of [10] has been applied.

coefficients will fall in the zero bin of the quantizer, which is equivalent to setting $m_0 = 1$ in (7). This is more difficult to observe if $\hat{\lambda}_i$ is small. In general, for fine quantization (left part of Figure 1(a)), the MSE distortion grows quadratically, since the underlying data p.d.f. is smooth, and the quantization step is small. We can make similar observations based on Figure 1(b), which illustrates the dependence of ε_i on $\hat{\lambda}_i$. If the quantization step is less than or equal to 16, the distortion introduced by the anti-forensic method is relatively small.

The quantization matrix used by JPEG is typically designed in such a way that frequency masking is taken into account, i.e. the quantization step size q_i is larger at higher frequencies. On the other hand, high frequency components have lower energy (higher $\hat{\lambda}_i$) due to the piecewise smoothness of natural images. According to the previous discussion, this is expected to produce a maximum of the error at intermediate frequencies. This is demonstrated in Figure 2, which shows $\hat{\varepsilon}_i$ averaged over all the 8×8 blocks of the *Lenna* image. This picture was obtained by previously compressing the image with a JPEG quality factor $Q = 30$ and the standard IJG quantization matrix. Notice that high frequency values, for which the error is close to zero, are likely to have all coefficients quantized to zero; as mentioned above, the anti-forensic algorithm does not introduce any dither to these values. This observation holds also for other images, and suggests the fact that the introduced dithering signal has, in general, a characteristic, content-dependent, spectrum.

Finally, the considerations emerged from Figure 1 may give hints in predicting the cost of anti-forensics depending on the smoothness characteristics of an image. Recall that smoother images have generally larger $\hat{\lambda}_i$ at high DCT frequencies. Thus, the average cost of the anti-forensic method is going to be smaller for smooth images. This is illustrated in Figure 3, where we show ε_i for two images characterized by different content. The *Mandril* image is characterized by high-frequency, noise-like textures, and this reflects in higher cost of anti-forensically doctoring it.

4. LOSS OF IMAGE QUALITY DUE TO ANTI-FORENSICS

The direct consequence of the distortion introduced by the dithering signal is a loss of perceived image quality, with respect to both the original (uncompressed) and to the JPEG-compressed image. The ability of MSE and of the related Peak Signal-to-Noise Ratio (PSNR) metric to predict image quality as judged by human observers has been widely questioned in the past literature [12]. Therefore, we adopt here the well-known Structural SIMilarity (SSIM) metric described in [13] as an objective measurement of the perceived image

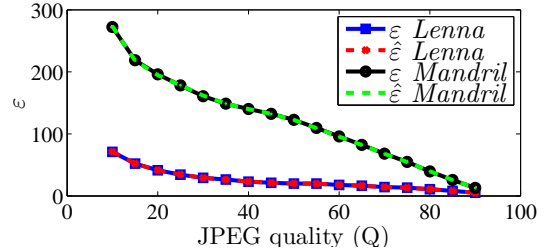


Fig. 3. The MSE $\hat{\varepsilon}_i$ for two images with different smoothness characteristics. The cost of removing JPEG statistical footprints from images with a plenty of high-frequency detail is in general higher than for smooth images.

quality. The correlation between the scores produced by this metric and users' quality scores has proved to be reliable over a broad range of testing conditions [14], including the types of distortion addressed in this work. Since the original method in [10] was not conceived to maximize image quality, we modify it by embedding the dithering signal in a perceptual-aware fashion. Then, we compare the results of this method with the baseline anti-forensic method.

4.1. Perceptual anti-forensic dither

The method in [10] is not optimized for minimizing the quality loss in the doctored image. In fact, the dithering signal sampled from the p.d.f.'s (2)-(4) is added indiscriminately to quantized coefficients. However, the ability of the human eye to perceive distortions depends strongly on the local characteristics of an image region, and can be described in terms of "just-noticeable distortion" (JND) [15]. A JND is the maximum distortion which cannot be perceived by the human eye. We employ a state-of-the-art JND model [16] which works directly in the DCT domain, giving for each DCT coefficient \hat{X}_i^b , a JND threshold τ_i^b . The values τ_i^b are then used to drive the insertion of dither into $\hat{\mathbf{X}}$, with the intuition that the dithering signal is expected to be unnoticeable if $Z_i^b \in [\hat{X}_i^b - \tau_i^b, \hat{X}_i^b + \tau_i^b]$ or, equivalently, $N_i^b \in [-\tau_i^b, +\tau_i^b]$.

The adaptive insertion of the dithering signal can be modeled as a minimum-cost bipartite graph matching problem. Let \mathcal{X} be the set of quantized DCT coefficients \hat{X}_i^b to which some dither N_i^b has to be added. Let \mathcal{N} denote the set of dithering values sampled from (2)-(4). The sets \mathcal{X} and \mathcal{N} have the same cardinality M , and can be represented by the two node partitions in a bipartite graph. Note that M is equal to the number of nonzero (zero) coefficients in the i -th DCT subband, hence $M \leq B$. The goal is to assign each dithering sample in \mathcal{N} to a quantized coefficient $\hat{X}_i^b \in \mathcal{X}$, in such a way that the total cost of the assignment

$$C = \sum_{k=1}^M \frac{|N_i^b|}{\tau_i^b} \quad (10)$$

is minimized. The cost C here is measured in JND units as proposed in [15], and is thus directly related to perceptual loss of quality.

The solution of the minimum-cost matching problem can be carried out in $O(M^3)$ time, where M is in the order of the number of 8×8 blocks in an image. Even with images at modest spatial resolution, $M \simeq 10^4$, thus making the exact solution of the problem impractical. Therefore, in our experiments we implemented a greedy algorithm which provides an approximate answer in $O(M^2)$ time.

The algorithm works as follows. First, the elements of the set \mathcal{N} are sorted in descending order of absolute value. Then, starting from the first element of \mathcal{N} , all edges departing from that node to each of the elements of \mathcal{X} are scanned, and the one with minimum cost is selected as a matching pair. Matching elements are removed from both sets, and the process is applied to the next element in the sorted list, until all nodes have been matched. In our simulations, we found that for all the tested images the solution of the greedy algorithm is typically within 5% of the optimal solution.

4.2. Analysis of perceptual quality loss due to anti-forensic dithering

In this section we evaluate the cost in term of perceptual quality loss for both the baseline anti-forensic method and the perceptually-modified version described above. The comparison here is with respect to the *original, uncompressed image*. This perspective has two advantages: first, it allows to compare the quality of the doctored image with the quality of the JPEG-compressed image, in terms of the loss with respect to the ground-truth original image; second, it is a more realistic indicator of the quality perceived by the observer, as it can be assumed that the quality of the original image is higher than its JPEG version.

Figure 4 shows the SSIM score for the same two images of Figure 3. SSIM values close to 1 indicate a higher quality. Notice that both the baseline and the perceptual anti-forensic algorithms guarantee the same degree of concealment of the JPEG footprint, as the noise is sampled from the very same distribution. We can observe that the perceptual anti-forensic algorithm is able to insert the dither achieving a better score than the method in [10]. The increase is especially apparent for the *Mandril* image. Indeed, the JND thresholds τ_i^b are larger in regions with high-contrast textures, and this increases the possibilities of concealing the dithering signal. Nevertheless, for both images the image quality after the anti-forensic forgery is far from being similar to that of the initial JPEG or to the original image.

5. CONCLUSIONS

Concealing the traces of a JPEG compression is possible, but it incurs in a serious cost in terms of the quality of the doctored image. In this paper, we have characterized such a loss both analytically and experimentally. The distortion left by the analyzed anti-forensic method could actually be tell-tale of anti-forensic manipulations, and might be used in the future to design methods that counteract JPEG anti-forensic attacks.

6. REFERENCES

- [1] M. Chen, J. Fridrich, M. Goljan, and J. Lukas, "Determining Image Origin and Integrity Using Sensor Noise," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 1, pp. 74–90, 2008.
- [2] H. Farid, "Digital image ballistics from JPEG quantization," *Dept. Comput. Sci., Dartmouth College, Tech. Rep. TR2006-583*, 2006.
- [3] Z. Fan and R. L. de Queiroz, "Identification of bitmap compression history: JPEG detection and quantizer estimation," *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 230–235, February 2003.
- [4] J. Lukáš and J. Fridrich, "Estimation of primary quantization matrix in double compressed JPEG images," *Proc. of Digital Forensic Research Workshop*, 2003.
- [5] A.C. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of resampling," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 758 – 767, February 2005.

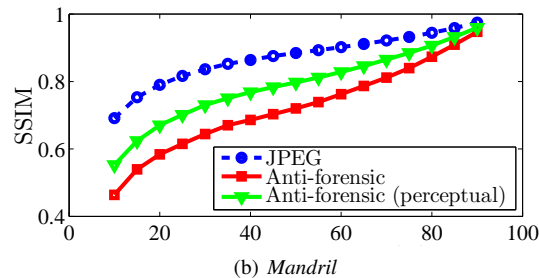
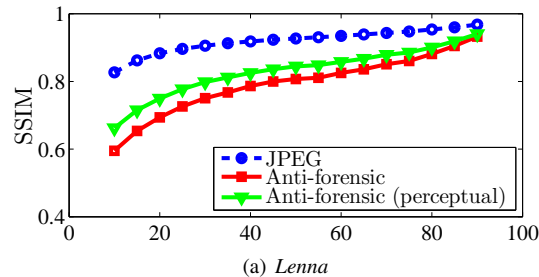


Fig. 4. SSIM distortion with respect to the *original* uncompressed image computed with the method in [10], and with the JND perceptual-based dithering. The quality of the JPEG image used as starting point is shown for reference.

- [6] W. Luo, Z. Qu, J. Huang, and G. Qiu, "A novel method for detecting cropped and recompressed image block," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, April 2007, vol. 2, pp. 217–220.
- [7] M. Stamm and K.J.R. Liu, "Blind forensics of contrast enhancement in digital images," in *Proceedings of the International Conference on Image Processing*, San Diego, CA, USA, October 2008, vol. 1, pp. 3112–3115.
- [8] S. Bayram, H.T. Sencar, and N. Memon, "A Survey of Copy-Move Forgery Detection Techniques," in *Proc. IEEE Western New York Image Processing Workshop*, Rochester, NY, USA, October 2008.
- [9] J. Fridrich, D. Soukal, and J. Lukáš, "Detection of Copy-Move Forgery in Digital Images," in *Proc. of Digital Forensic Research Workshop*, Cleveland, USA, August 2003.
- [10] M.C. Stamm, S.K. Tjoa, W.S. Lin, and K.J.R. Liu, "Anti-forensics of JPEG compression," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Dallas, TX, USA, April 2010.
- [11] J.R. Price and M. Rabbani, "Biased reconstruction for jpeg decoding," *IEEE Signal Process. Lett.*, vol. 6, no. 12, pp. 297–299, December 1999.
- [12] B. Girod, "What's wrong with mean-squared error?," *MIT Press Cambridge, MA, USA*, 1993.
- [13] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, April 2004.
- [14] H.R. Sheikh, M.F. Sabir, and A.C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, November 2006.
- [15] A.B. Watson, "DCT quantization matrices visually optimized for individual images," in *Proc. SPIE*. Citeseer, 1993, vol. 1913, pp. 202–216.
- [16] Z. Wei and K.N. Ngan, "Spatio-temporal just noticeable distortion profile for grey scale image/video in DCT domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 3, pp. 337–346, March 2009.