A NOVEL QUANTIZATION-BASED WATERMARKING APPROACH INVARIANT TO GAIN ATTACK

Mohsen Zareian[*], *Hamid Reza Tohidypour*[1], *and Z. Jane Wang*[1]

[*]Department of Electrical Engineering, Amirkabir University of Technology, Iran, Tehran
[1]Department of Electrical and Computer Engineering, University of British Columbia, BC, Vancouver
smohsenz@aut.ac.ir, htohidyp@ece.ubc.ca, and zjanew@ece.ubc.ca

ABSTRACT

In this paper, a novel quantization-based information hiding approach which is invariant to gain attack is presented. For the data embedding, the host vector signal is first divided into two separate vectors. Then the ratio of the magnitude of the vectors is quantized according to the watermark data. The decoding scheme is performed blindly using the euclidean distance. The performance of the proposed method is analytically studied and assessed by simulations on artificial signals. The proposed method is applied to various test images as well. The experimental results confirm the superiority of the proposed technique against common attacks in comparison with the recently proposed methods.

Index Terms— quantization-based information hiding, ratio of the magnitude of vectors, gain attack

1. INTRODUCTION

In the past decade, the quantization-based watermarking has grabbed the attention of researchers due to its good rate-distortion-robustness trade-offs. Quantization index modulation (QIM) is one of the most important methods proposed by Chen and Wornel [1]. In the QIM method, the watermark data is embedded by quantizing the host signal features using a set of quantizers, each of which associated with a different message. Spread-transform dither modulation (ST-DM) [1] is a quantization based scheme in which the data embedding is performed by quantizing the projection of the host signal vector along a random direction. Many QIM based methods have been developed so far. In [2], Kalantari and Ahadi has proposed a logarithmic QIM (LQIM) which poses a better robustness and perceptual advantages in comparison with the traditional QIM due to performing a logarithmic transform on the host signal before quantization.

The main drawback of quantization-based methods is their extreme sensitivity to valumetric attacks since this attack that produces a mismatch between the encoder and the decoder. During the last few years, many improved techniques have been proposed to deal with this issue. Among recent methods to alleviate this problem, Rational Dither Modulation (RDM) [3] is a simpler and effective one. In RDM, a division function is used to overcome the gain attack. Another approach named Hyperbolic RDM [4] has been proposed to obtain invariance to both amplitude and power law attacks. RDM based methods can asymptotically achieve the performance of the QIM method. The high peak to average power ratio (PAPR) of RDM is a disadvantage that should be addressed. In [5], Ourque et al. has introduced a watermarking scheme called angle QIM (AQIM). AQIM embeds watermark in the angle vector which makes the algorithm robust to amplitude scaling attacks. However, it has been shown that AQIM has a low robustness against AWGN attack. In [6], Akhaee et al. has introduced a gain invariant watermarking method named Sample Projection (SP) in which the watermarking code is embedded by projecting the line segment obtained from samples of the host signal on some specific lines in the 2-D space according to message bits. It has been shown that this approach outperforms than the AQIM and Hyperbolic RDM methods. Recently, a ST-DM based method called Normalized Correlation based Dither Modulation (NC-DM) which is robust against scaling attacks has been proposed [7]. In this method, the watermark data is embedded by quantizing the normalized cross correlation between the host signal vector and a random vector. However, similar to the ST-DM scheme, the random vector must be sent to the decoder which decreases the security of the algorithm,since the malicious attacker can change either the watermark or the random vector.

In this paper, we introduce a novel gain invariant watermarking scheme based on quantization. Partitioning the host signal vector into two separate parts, the data is embedded by quantizing the ratio of the normalized magnitude of each part. In the decoding process, the watermark code is extracted using the euclidean distance. The embedding distortion is first derived. Then, the error probability is analytically calculated. By simulations on artificial signals, the analytical derivations are verified. Simulation results on image signals show the superior performance of the proposed method in comparison with recent watermarking methods.

2. PROPOSED METHOD

2.1. Watermark Embedding

Let $\mathbf{u}$ be the host signal consisting N variables $u_1, u_2, ..., u_N$. The N samples of $\mathbf{u}$ are divided into two subsequences $\mathbf{x}$ and $\mathbf{y}$ containing the even and odd indexed terms, respectively: $x_i = u_{2i}, y_i = u_{2i-1}$, $i = 1, ..., \frac{N}{2}$. In order to embed the watermark message $m \in \{0, 1\}$ in $\mathbf{u}$, the normalized magnitude of $\mathbf{x}$ and $\mathbf{y}$ are calculated:

$$s_x = \sqrt{\frac{2}{N}\sum_{i=1}^{\frac{N}{2}} u_{2i}^2}, \qquad s_y = \sqrt{\frac{2}{N}\sum_{i=1}^{\frac{N}{2}} u_{2i-1}^2} \tag{1}$$

where s_x and s_y are the normalized magnitudes of $\mathbf{x}$ and $\mathbf{y}$, respectively. Then, the QIM method is applied to the ratio of s_x and s_y, $z = \frac{s_x}{s_y}$, as follows:

$$z_q = Q_m(z) = \Delta round\left(\frac{z + m\Delta/2}{\Delta}\right) - m\frac{\Delta}{2} \tag{2}$$

Then, x_i and y_i are updated as follows:

$$x_i' = \sqrt{\frac{z_q}{z}} x_i, \qquad y_i' = \sqrt{\frac{z}{z_q}} y_i \tag{3}$$

Repositioning $\mathbf{x}'$ and $\mathbf{y}'$ in the even and odd positions of $\mathbf{u}$ we finally obtain the watermarked signal $\mathbf{u}'$.

2.2. Watermark Extraction

At the decoder side, the received signal $\mathbf{u}''$ is first split between two subsequences $\mathbf{x}''$ and $\mathbf{y}''$ containing the samples of $\mathbf{u}''$ in even and odd positions, respectively. Then, the normalized magnitudes of $\mathbf{x}''$ and $\mathbf{y}''$ are calculated using equation (1). Finally using the ration of $s_{x''}$ and $s_{y''}$, $z'' = \frac{s_{x''}}{s_{y''}}$, and the euclidean distance, the watermark data is extracted as follows:

$$\hat{m} = \arg \min_{m \in \{1,0\}} \left| z'' - Q_m(z'') \right| \tag{4}$$

If the watermarked signal is multiplied by a constant gain, , α, due to the use of division function, it has no effect on the decoding process. Thus, our algorithm is invariant to scaling attack.

3. ANALYSIS AND PERFORMANCE EVALUATION OF THE PROPOSED METHOD

Here, we analysis the performance of the proposed method by obtaining the embedding distortion and the error probability.

3.1. Signal Modeling

The host signal is assumed to be independent, identically distributed (i.i.d). Since the low-frequency components of the most natural signal such as image and audio are modeled with a Gaussian distribution, we assume that the base signal is Gaussian with zero mean and variance σ. In this section, we will need to calculate the distribution of the ratio of two normal variables $a = \mathcal{N}\left(\mu_a, \sigma_a^2\right)$ and $b = \mathcal{N}\left(\mu_b, \sigma_b^2\right)$. The distribution of c, $c = \frac{a}{b}$, for the case of non zero μ_a and μ_b, and $\frac{\sigma_a}{\mu_a} \ll 1$, $\frac{\sigma_b}{\mu_b} \ll 1$ can be well approximated by a Gaussian distribution. The parameters μ_c and σ_c^2 of the approximated Gaussian distribution using Taylor series can be derived as follows:

$$\mu_c = \frac{\mu_a}{\mu_b}(1 + A), \qquad \sigma_c^2 = \frac{\mu_a^2}{\mu_b^2}(A + 2A^2) + \frac{\sigma_a^2}{\mu_b^4} \tag{5}$$

where $A = \frac{\sigma_b^2}{\mu_b^2}$.

3.2. Derivation of Embedding Distortion

In order to obtain the embedding distortion, we need to find $u_i' - u_i$. According to (3), we have:

$$u_i' - u_i = \left(\sqrt{\frac{z_q}{z}} - 1\right) u_{2i} + \left(\sqrt{\frac{z}{z_q}} - 1\right) u_{2i-1} \tag{6}$$

Considering the quantization noise to be ε, we have $z_q = z + \varepsilon$. Assuming Δ to be sufficiently small and using Taylor expansion for $\sqrt{\frac{z_q}{z}}$ and $\sqrt{\frac{z}{z_q}}$, we get:

$$u_i' - u_i \simeq \frac{\varepsilon}{2z} u_{2i} - \frac{\varepsilon}{2z} u_{2i-1} \tag{7}$$

Here, we neglect higher order terms in Taylor expansion. From above equation, the embedding distortion $E\left[\|u_i' - u_i\|^2\right]$ can be found as:

$$E\left[\|\mathbf{u}' - \mathbf{u}\|^2\right] \simeq E\left[\|\mathbf{u}\|^2\right] E\left[\varepsilon^2\right] E\left[\frac{1}{4z^2}\right] \tag{8}$$

where three terms inside the expectation are considered to be independent. In order to find $E[\frac{1}{z^2}]$, we need to calculate the distribution of $\frac{1}{z^2} = \frac{s_y^2}{s_x^2}$. Using Central Limit Theorem (CLT), s_x^2 and s_y^2 can be modeled with a Gaussian distribution; i.e., $s_x^2 \sim \mathcal{N}(\sigma^2, 4\sigma^4/N)$, $s_y^2 \sim \mathcal{N}(\sigma^2, 4\sigma^4/N)$. According to the suggested Gaussian distribution in 3.1, $E\left[\frac{1}{z^2}\right] \simeq 1 + \frac{4}{N}$. When the quantization step-size is small, the quantization noise ε can be considered to follow uniform distribution between $[-\Delta/2, \Delta/2]$. Therefore, we have:

$$E\left[\|\mathbf{u}' - \mathbf{u}\|^2\right] \simeq \sigma^2 \frac{\Delta^2}{48}(1 + \frac{4}{N}) \tag{9}$$

Also Document to Watermark Ratio (DWR) can be obtained as follows:

$$DWR = \frac{E\left[\|\mathbf{u}\|^2\right]}{E\left[\|\mathbf{u}' - \mathbf{u}\|^2\right]} = \frac{N48}{(N+4)\Delta^2} \tag{10}$$

As can be seen, DWR is independent of the host signal variance which leads to a watermark power proportional to the host signal power.

3.3. Derivation of the Error Probability

We conduct the analysis by considering the watermarked signal to be sent through Additive White Gaussian Noise (AWGN) channel, i.e., $\mathbf{u}'' = \mathbf{u}' + \mathbf{n}'$. At the receiver, z'' is calculated as follows:

$$z'' = \frac{s_{x''}}{s_{y''}} = \frac{\sqrt{\frac{2}{N}\sum_{i=1}^{\frac{N}{2}} (x_i' + n_{x_i}')^2}}{\sqrt{\frac{2}{N}\sum_{i=1}^{\frac{N}{2}} (y_i + n_{y_i}')^2}} \tag{11}$$

where n_{x_i}' and n_{y_i}' denote the odd and even indexed noise terms. Since z'' is not easy to handle, we continue our analysis using z''^2. Thus, we have:

$$z''^2 = \frac{s_{x'}^2 + \nu_{x'}}{s_{y'}^2 + \nu_{y'}} \tag{12}$$

where $s_{x'}^2 = \frac{2}{N}\sum_{i=1}^{N/2} x_i'^2$, $\nu_{x'} = \frac{2}{N}\sum_{i=1}^{N/2} 2x_i' n_{x_i}' + \frac{2}{N}\sum_{i=1}^{N/2} n_{x_i}'^2$, $s_{y'}^2 = \frac{2}{N}\sum_{i=1}^{N/2} y_i'^2$, and $\nu_{y'} = \frac{2}{N}\sum_{i=1}^{N/2} 2y_i' n_{y_i}' + \frac{2}{N}\sum_{i=1}^{N/2} n_{y_i}'^2$. Considering small noise terms, z''^2 can be estimated as:

$$\begin{aligned} z''^2 &\simeq \frac{s_{x'}^2}{s_{y'}^2}\left[(1 + \frac{\nu_{x'}}{s_{x'}^2})(1 - \frac{\nu_{y'}}{s_{y'}^2})\right] \\ &\simeq \underbrace{\frac{s_{x'}^2}{s_{y'}^2}}_{z_q^2} + \underbrace{(\frac{\nu_{x'}}{s_{y'}^2} - \frac{s_{x'}^2}{s_{y'}^2}\frac{\nu_{y'}}{s_{y'}^2} - \frac{\nu_{x'}}{s_{y'}^2}\frac{\nu_{y'}}{s_{y'}^2})}_{n_t} \end{aligned} \tag{13}$$

The first term in this equation z_q^2 is the clean term and the second one n_t is the noisy term. Here, error in detection occurs when the noisy term causes the clean term to fall into a wrong region. In order to calculate the probability of error, we need to find the distribution of

n_t. For this, the distribution of $s^2_{x'}, s^2_{y'}, \nu_{y'}$, and $\nu_{x'}$ should be first calculated. Using CLT and the independency of noise and signal, we have: $s^2_{x'} \sim \mathcal{N}(z_q\sigma^2, \frac{z_q^2 4\sigma^4}{N})$, $s^2_{y'} \sim \mathcal{N}(\frac{\sigma^2}{z_q}, \frac{4\sigma^4}{z_q^2 N})$, $\nu_{x'} \sim \mathcal{N}(\sigma_n^2, \frac{8\sigma^2\sigma_n^2 z_q + 4\sigma_n^4}{N})$, $\nu_{y'} \sim \mathcal{N}(\sigma_n^2, \frac{\frac{8\sigma^2\sigma_n^2}{z_q} + 4\sigma_n^4}{N})$.

The noisy term n_t consists of three terms:

$$n_t = \underbrace{\frac{\nu_{x'}}{s^2_{y'}}}_{\zeta_1} \underbrace{- \frac{s^2_{x'}}{s^2_{y'}}\frac{\nu_{y'}}{s^2_{y'}}}_{\zeta_2} \underbrace{- \frac{\nu_{x'}}{s^2_{y'}}\frac{\nu_{y'}}{s^2_{y'}}}_{\zeta_3} \tag{14}$$

Using the suggested Gaussian distribution in (3.1), the distribution of these three terms can be easily calculated. Because of the independency of noise and signal, all these three terms are independent of each other. Therefore, mean and variance of n_t are $\mu_{n_t} = \mu_{\zeta_1} + \mu_{\zeta_2} + \mu_{\zeta_3}, \sigma^2_{n_t} = \sigma^2_{\zeta_1} + \sigma^2_{\zeta_2} + \sigma^2_{\zeta_3}$.

It is straightforward to show that the probability of error when minimum distance decoder is used is given as:

$$P_e = \sum_{k=1}^{\infty} Pr\{t^2_{(k-1)/2} < z^2 < t^2_{(k+1)/2}\}$$
$$\times \sum_{m=-\lfloor \frac{k}{2} \rfloor}^{\infty} Pr\{v^2_{(2m+k)} < z''^2 < v^2_{(2m+k+1)}\} \tag{15}$$

where t_k and v_k are defined as:

$$t_k = k\Delta, \qquad v_k = \frac{t_{k/2} + t_{(k+1)/2}}{2} \tag{16}$$

Using the distribution of z^2 and n_t, (15) can be written as:

$$P_e = \sum_{k=1}^{\infty} \left(Q\left(\frac{t^2_{(k-1)/2} - \mu_{z^2}}{\sigma_{z^2}}\right) - Q\left(\frac{t^2_{(k+1)/2} - \mu_{z^2}}{\sigma_{z^2}}\right)\right) \times$$
$$\sum_{m=-\lfloor \frac{k}{2} \rfloor}^{\infty} \left(Q\left(\frac{v^2_{(2m+k)} - \mu_{z''^2}}{\sigma_{z''^2}}\right) - Q\left(\frac{v^2_{(2m+k+1)} - \mu_{z''^2}}{\sigma_{z''^2}}\right)\right) \tag{17}$$

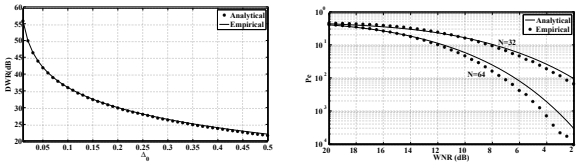
where $Q(\alpha) = 1/\sqrt{2\pi}\int_{\alpha}^{\infty} e^{-u^2/2}du$ is the Q-function, $\mu_{z''^2} = \mu_{n_t} + z_q^2$, and $\sigma_{z''^2} = \sigma_{n_t}$. It should be noted that z_q in the distribution of z'' is equal to $t_{k/2}$.

4. EXPERIMENTAL RESULTS

In this section, we test the performance of the proposed algorithm and the validity of analytical derivations by simulations on artificial signals and real images.

4.1. Simulation on Artificial Signals

We first conduct the experiment by simulation on synthetic signals. For this, we generate Gaussian i.i.d. data with $\mu = 0$ and $\sigma = 1$ to be used as the host signal. Then, data embedding is performed as described in section 2 and the rate of embedding for all simulations is 1/32, i.e., one bit in each 32 samples. The results are obtained by averaging over 100 simulations with 100,000 bits each. Fig. 1 (a) shows the analytical and empirical DWR for different Δ. As can be seen, the results are matched well which confirms the accuracy of analysis provided in section 3.2. In order to validate the analytical analysis for the error probability, data are extracted from the watermarked signal after adding white Gaussian noise with different variances at the decoder side. DWR is fixed at 30 dB. Fig. 1 (b) shows the empirical and analytical BER of the proposed method for various Watermark-to-Noise Ratios (WNRs). As seen, analytical results are closely matched to the empirical values. As expected, better improvement is achieved as N increase.



(a) (b)

Fig. 1. (a) Analytical and empirical DWR for different Δ. (b) Empirical and analytical BER of the proposed method for different WNRs (DWR=30dB).

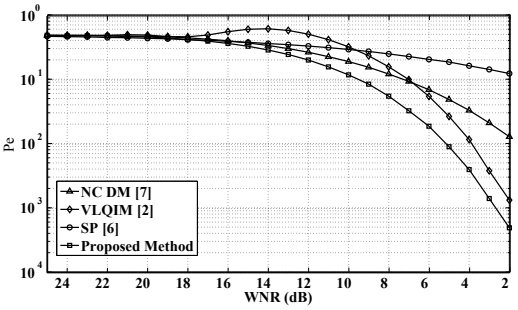


Fig. 2. Comparing the probability of error for the proposed method, SP, VLQIM, and NC-DM with DWR of 19 dB.

The robustness of our algorithm is also compared with three recently proposed method, SP, vector LQIM (VLQIM), and NC-DM. The methods are implemented on artificial signal in similar situations for DWR of 19 dB. The results are shown in Fig. 2 As we can see, our approach outperforms SP and NC-DM.

4.2. Simulation on Real Images

To show the performance of the proposed method in the real application of image watermarking and compare it with VLQIM, and NC-DM, we implement all methods in the wavelat transform domain. For this, a three-level wavelet transform with HAAR filter is applied to the host image. Then, the approximation coefficients of the third level are used for quantization, resulting in embedding of 4096 coefficients in a 512×512 host image. Length of host vectors N for all approaches are set to 32 which results in 128 bits for each image. For this study, we use ten well-known images of size 512×512: *Lenna*, *Baboon*, *Couple*, *Pirate*, *Barbara*, *Bridge*, *Plane*, *Boat*, *Peppers*, and *Goldhill*. Parameters of all methods are set in a way that PSNR for all images is equal to 48 dB. We also use the mean structural similarity index (MSSIM) [8] which is highly matched with the human visual system for verifying the imperceptibility of the watermarked images. Table 1 demonstrates the MSSIM of the watermarked images for all methods. As seen in Table 1, VLQIM achieves superior performance which is due to applying a logarithmic function to improve the perceptual quality. Furthermore, the MSSIM of

Table 1. the mean structural similarity index (MSSIM) of the watermarked images for three methods

Image	Methods		
	VLQIM	NC-DM	Proposed
Lenna	0.9970	0.9937	0.9959
Baboon	0.9991	0.9980	0.9987
Couple	0.9980	0.9961	0.9974
Pirate	0.9982	0.9961	0.9971
Barbara	0.9984	0.9961	0.9975
Goldhill	0.9983	0.9962	0.9970
Bridge	0.9992	0.9981	0.9987
Peppers	0.9980	0.9951	0.9957
Plane	0.9972	0.9935	0.9956
Boat	0.9983	0.9951	0.9958

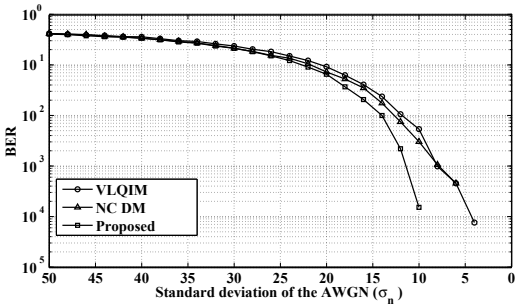


Fig. 3. BER of watermark extraction under AWGN attack for various noise variances. The results are averaged over ten well-known images.

our scheme is higher than NC-DM which shows its better perceptual advantage. Fig. 3 presents the bit error rate (BER) of watermark extraction for all methods under AWGN attack with different standard deviations. It is seen that our algorithm outperforms than other methods. The robustness of all methods is also investigated against JPEG compression with different quality factors. The results are shown in Fig. 4. The superiority of the proposed approach over VLQIM and NC-DM is obvious. Finally, the BER results of all methods under median and Gaussian lowpass ltering are obtained and summarized in Table 2.

5. CONCLUSION

In this paper, a novel quantization watermarking method robust against valumetric scaling attack was proposed. In this algorithm, after partitioning into two separate parts, the ratio of the magnitude of the vectors was quantized for data embedding. Data extraction was performed blindly using minimum distance decoder. Embedding distortion and the probability of error were analytically derived and assessed by simulations on arti cial signals. As shown in simulation results, the performance of the proposed method was better than VLQIM and NC-DM. Furthermore, as shown, the proposed method had perceptual advantage over NC-DM.

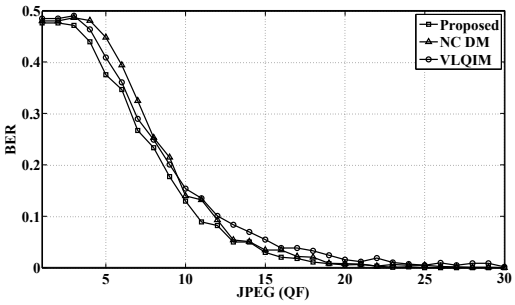


Fig. 4. BER Results of extracted watermark under JPEG compression. The results are averaged over ten well-known images.

Table 2. BER(%) of extracted watermark under Gaussian and Median ltering. The window size for Gaussian lter is 3×3

Methods	Gaussian Filter(σ^2)			Median Filter	
	$\sigma^2 = 1$	$\sigma^2 = 1.5$	$\sigma^2 = 2$	3×3	5×5
VLQIM	0.15	0.23	0.31	6.25	19.37
NC-DM	0.00	0.07	0.15	1.48	10.00
Proposed	**0.00**	**0.00**	**0.00**	**1.32**	**8.35**

6. REFERENCES

[1] B. Chen and G. W. Wornel, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Information Theory*, vol. 47, no. 4, pp. 1423-1443, May 2001.

[2] N. K. khademi and S. M. Ahadi, "A logarithmic quantization index modulation for perceptually better data hiding," *IEEE Trans. Image processing*, vol. 19, no. 6, June 2010.

[3] F. Perez-Gonzalez, C. Mosquera, M. Barni, and A. Abrardo "Rational dither modulation:A high-rate data-hiding method invariant to gain attacks," *IEEE Trans. Signal Process*, vol. 53, no. 10, pp. 39603975, Oct. 2005.

[4] P. Guccione and M. Scagliola, "Hyperbolic rdm for nonlinear valumetric distortions," *IEEE Trans. Inf. Forensics Security*, vol. 4, no. 1, pp. 2535, Mar. 2009.

[5] F. Ourique, V. Licks, R. Jordan, and F. Perez-Gonzalez, "Angle QIM: A novel watermark embedding scheme robust against amplitude scaling distortions," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Philadelphia, PA, Mar. 2005, vol. 2, pp. 797800.

[6] M. A. Akhaee, S. M. E. Sahraeian, and C. Jin, "Blind image watermarking using a sample projection approach," *IEEE Trans. Information Forensics and Security.*, vol. 6, no. 3, pp. 883-893 , Sep. 2011.

[7] Xinshan Zhu and Shuoling Peng, "A novel quantization watermarking scheme by modulating the normalized correlation," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Kyoto, Mar. 2012, pp. 1765-1768.

[8] 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 , Apr. 2004.