Highly Transparent Steganography Scheme of Speech Signals into Color Images Using Quantization Index Modulation

Diego Renza(✉), Dora M. Ballesteros L., and Jeisson Sanchez

Universidad Militar Nueva Granada, Bogotá, Colombia
{diego.renza,dora.ballesteros,u1400881}@unimilitar.edu.co

Abstract. A highly transparent steganography scheme of speech signals into color images based on Quantization Index Modulation (QIM) is presented in this paper. The proposed method takes advantage of the low distortion in the host image introduced by the scalar quantization of its wavelet coefficients. The stego image is highly similar to the host image, and the secret content is imperceptible. The secret message is recovered at the receiver with high correlation to the original speech signal. For the purpose of increasing the security of the system, an external mask is used to encrypt the secret content before the embedding process. Several tests were carried out in order to quantify the influence of the size of the quantizer (delta) on the quality of the recovered secret content and the transparency of the stego image.

Keywords: Steganography · Discrete wavelet transform · Quantization index modulation · Imperceptibility

1 Introduction

Nowadays, data hiding techniques are very popular due to the need for data privacy or copyright protection. In the first case, steganography consists of embedding a secret message into a host signal with the purpose of concealing the existence of its secret content [7]; in the second case, a mark is embedded in order to protect authorship of the host [11]. This means that in steganography the most important signal is the secret message while in the case of watermarking it is the host signal. Regardless of the data hiding technique, the secret messages and host signals can be text, audio, image or video [1,3].

Some proposals of data hiding embed data in the wavelet transform domain. For example, a speech steganography scheme based on the DWT is proposed in [13]. Authors found that the obtained stego signals are highly correlated (similar) to the host speech signals, and the transparency is guaranteed. These results have been confirmed in other studies [5,6], even in medical applications [8].

On the other hand, Quantization Index Modulation (QIM) has proved to be a good solution for reversible and robust data hiding schemes [4,12]. Nevertheless, some works have demonstrated that security of the systems based on QIM can

J.F. Martínez-Trinidad et al. (Eds.): MCPR 2016, LNCS 9703, pp. 241–250, 2016.
DOI: 10.1007/978-3-319-39393-3_24

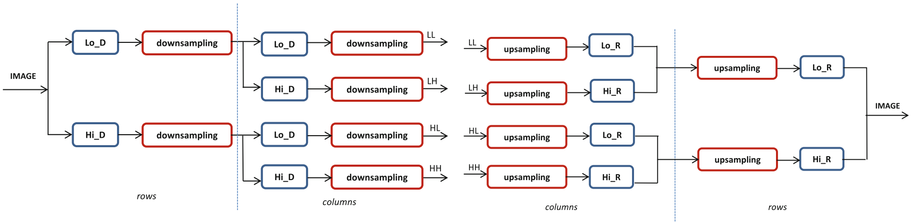
be weak and an attacker can reveal the secret content from the analysis of the stego information [9].

In this work we propose a steganography scheme for concealing a speech signal in a color image, by using QIM in the wavelet domain of the host signal. Additionally, we add a level of security by using an XOR operation between an external mask and the speech signal. The result is a stego image with high similarity to the host image, a recovered speech signal with high quality, and an additional level of security to reveal the secret content.

2 Methods and Algorithms

2.1 Discrete Wavelet Transform (DWT)

The DWT is a useful mathematic method for signal decomposition into different resolutions. In the case of images, the input image is passed through two filters to obtain four sub-bands: one approximation sub-band (*LL*) and three detail sub-bands. The detail coefficients are oriented in one of three directions, horizontal (*LH*), vertical (*HL*) or diagonal (*HH*). The *LL* sub-band keeps the highest similarity to the image, while *HL*, *LH* and *HH* keep information related to the specific orientation. Figure 1 shows the chart for the decomposition and reconstruction stages.



(a) Decomposition (b) Reconstruction

Fig. 1. Chart of the wavelet transform for images.

Filters used in the decomposition (Lo_D, Hi_D) and reconstruction process (Lo_R, Hi_R) must satisfy the condition of quadrature mirror filters [10].

2.2 Quantization Index Modulation

The QIM is a binary data concealing technique based on quantization rules. The QIM generally refers to data embedding with a specific quantizer or a sequence of quantizers on a host signal. The insertion process is based on the modification of the carrier value (pixel, wavelet coefficient, audio sample, etc.); the carrier value C is approximated to the nearest multiple of the static step value (Δ) obtaining the stego value S.

2.3 Quality Metrics

Different techniques of quality signal analysis such as Structural Similarity Index Metric (SSIM), Gray Value Degree (GVD) and Pearson's Correlation Coefficient (SPCC) were used to evaluate the imperceptibility of the stego image and the quality of the recovered audio signal. Next, the general characteristics of these methods are described.

Structural Similarity Index Metric. SSIM is an efficient method of quality analysis for similarity on image and videos. For two images U and V, the SSIM is given by Eq. 1 [14] and its ideal value is 1.

$$SSIM = \frac{(2\mu_U\mu_V + c_1)\,(2\sigma_{UV} + c_2)}{(\mu_U^2 + \mu_V^2 + c_1)\,(\sigma_U^2 + \sigma_V^2 + c_2)} \tag{1}$$

Where μ is the mean, σ is the standard deviation, c_1 and c_2 are stabilization variables.

Gray Value Degree. The Gray Value Degree, (GVD) uses histogram differences between the gray values of adjacent pixels to specify the texture characteristics. Its ideal value is zero. To obtain the GVD of a $U(M \times N)$ image, it is necessary to compute the gray value (GN) and the average neighborhood gray difference (AN), by means of the Eqs. 2–4.

$$GN\,(x,y) = 4U(x,y) - \sum_{i=0}^{3} U\left(x + \left\lfloor \frac{i}{2} \right\rfloor (-1)^i, y + \left\lfloor \frac{3-i}{2} \right\rfloor (-1)^i\right) \tag{2}$$

Using all the results of GN, the average neighborhood gray difference (AN) is calculated as follows:

$$AG = \sum_{x=2}^{M-1} \sum_{Y=2}^{N-1} GN\,(x,y) \tag{3}$$

Finally, the gray value degree is calculated as follows:

$$GVD = \frac{AG' - AG}{AG' + AG} \tag{4}$$

Where AG' and AG are the average neighbourhood gray difference of the original image and the modified image, respectively.

Pearson's Correlation Coefficient (SPCC). Audio and image testing were based on correlation coefficient between the original signal and the recovered signal. The correlation function $SPCC$ for two signals U and V is defined by means of Eq. 5, where σ is the deviation and cov the covariance [2].

$$SPCC = \frac{cov(U,V)}{\sigma_U\sigma_V} \tag{5}$$

3 Proposed Scheme

The proposed algorithm conceals an audio signal in an image. Color images are suitable since they can allow one to hide more information than a grayscale image. Moreover, the larger the image size, the greater the concealment capacity. The proposed scheme performs the audio embedding as follows:

3.1 Audio Embedding

The input parameters are the carrier image C and the audio signal A as the secret message. The outputs are the stego image S and the secret key. The audio embedding process is summarized in Fig. 2 and described below.

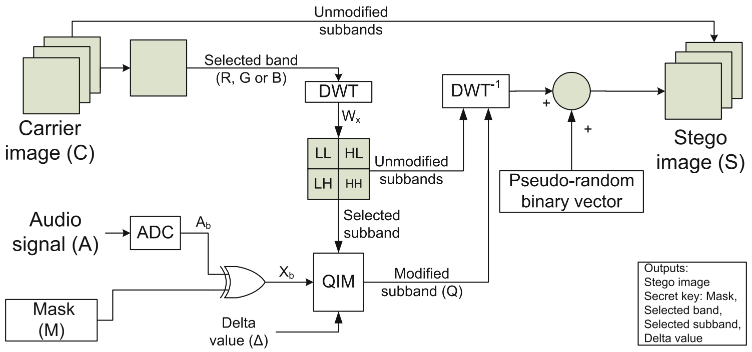


Fig. 2. Illustrative representation of the flow chart for audio embedding.

Step 1. Mask generation: the mask (M) is obtained through the conversion of a text string to a binary representation through the ASCII code. This text string is used to provide security to the embedded message. For example, if the text string is 32-characters length, the total number of bits in M is 256.

Step 2. Binary representation of the speech signal and masking operation: the speech signal (A) is converted to a binary representation (A_b). The total number of bits is equal to the total number of samples multiplied by the number of bits per sample. Then, an XOR operation is performed, $X_b = A_b \oplus M$. If the length of K is lower than the length of A_b, M is repeated as necessary. For example, if M is 256 bits length, and A is a speech signal with $f_s = 8\,KHz$, 2 s of duration and 16-bits per sample, the length of A_b is 256 K bits. In this case, M is repeated 1000 times, then an XOR operation is made (bit to bit).

Step 3. Extraction of color and wavelet sub-band: the host image is separated in three color bands (R, G, B) and one of them is selected. Then, the DWT is applied to the selected sub-band with one level of decomposition.

Step 4. Data insertion: the algorithm selects one of the wavelet sub-bands; then every bit of the X_b vector is inserted into the wavelet-subband through the

QIM function. Equation 6 represents the way to insert data; where Q is the quantized sub-band:

$$\begin{cases} Q(i,j) = \Delta \left\lfloor \frac{W_x(i,j)}{\Delta} \right\rfloor & if \quad X_b = 0 \\ \\ Q(i,j) = \Delta \left\lfloor \frac{W_x(i,j)}{\Delta} \right\rfloor + \frac{\Delta}{2} & if \quad X_b = 1 \end{cases} \tag{6}$$

For the purpose to recover the secret data at the receiver, delta value, the mask, the selected color band, and the selected wavelet sub-band are transmitted by a private channel. These data form the secret key.

Step 5. Addition of noise to the quantized sub-band: since after the QIM process, only will exist wavelet coefficients multiples of $\Delta/2$, an additive noise is inserted into the quantized sub-band with the purpose to resist attacks based on data distribution.

Step 6. Image reconstruction: the image is reconstructed from the wavelet sub-bands and the color bands. The result is the stego image (S).

3.2 Audio Extraction

The input parameters are the received stego image (Sr) and the parameters used in embedding phase (Δ value in QIM, color band, wavelet sub-band, and the mask). The output is the recovered audio signal (Ar). The audio extraction process is summarized in Fig. 3 and described below.

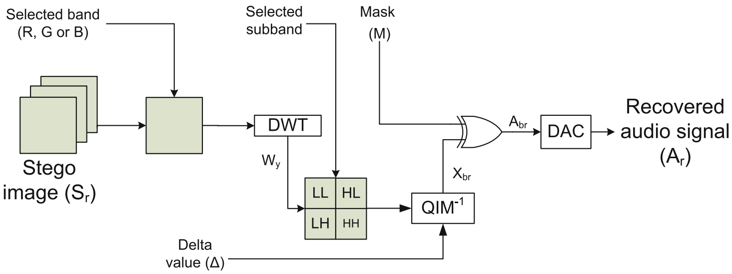


Fig. 3. Illustrative representation of the flow chart for audio extraction.

Step 1. Separation of the three bands of the stego image: the stego image is separated in the colors Red, Green and Blue.

Step 2. Application of the wavelet transform: firstly, one of the color bands is selected according to the information of the secret key. The DWT is applied to this band and then the sub-band specified by the secret key is selected properly.

Step 3. Data extraction: in every pixel of the selected sub-band the inverse QIM function is applied, according to Eq. 7. The result is the recovered masked data, X_{br}.

$$\begin{cases} X_{br} = 1 & if \quad \frac{\Delta}{4} < W_y(i,j) - \Delta \left\lfloor \frac{W_y(i,j)}{\Delta} \right\rfloor \leq \frac{3\Delta}{4} \\ X_{br} = 0 & else \end{cases} \tag{7}$$

Step 4. Unmask the audio: an xor operation between X_{br} and M is performed. M is obtained from the secret key. The result is the recovered bits of the speech signal, $A_{br} = X_{br} \oplus M$.

Step 5. Conversion to analog values: the binary values A_{br} are converted to analog values by using 16-bits data.

3.3 Insertion Capacity

In an information-hiding scheme, the insertion capacity (IC) is a measure of the quantity of information per unit that the method is able to hide. The insertion capacity of the proposed method is given by the Eq. 8,

$$IC = \frac{M \times N}{\left(2^L\right)^2} \, (bits/pixel) \tag{8}$$

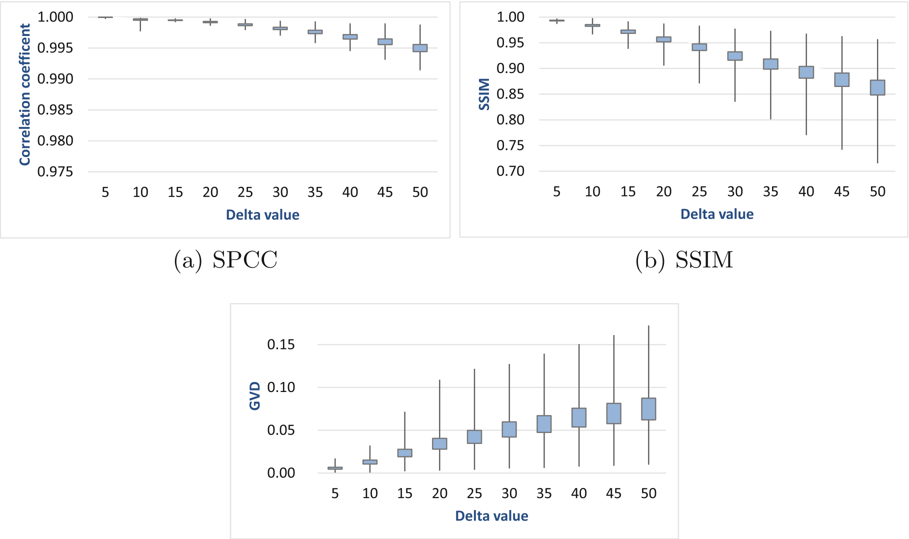
Where, M and N are the dimensions of the image and L is the number of decomposition levels. The above means that in a 1024 × 768 image, it is possible to hide 196608 bits. With an audio file sampled at 8 KHz (speech), it is possible to hide a file of 24.576 s of duration. With an audio file sampled at 44.1 KHz (music), it is possible to hide a file of 4.45 s of duration. In any case, it is possible to hide a little phrase or message.

4 Experimental Results and Analysis

To evaluate the effectiveness of the algorithm in terms of imperceptibility of the stego image and quality of the recovered signal we use five RGB images (three standard images and two proprietary images): Baboon (512 × 512 × 3), Lena (1600 × 900 × 3), Peppers (676 × 470 × 3), Penguins (1024 × 768 × 3) and Flower (1366 × 768 × 3). Besides, Haar wavelet base was selected and ten Delta values, four wavelet subbands and ten secret messages were used. A total of 2000 tests were done (5 images × 10 Delta values × four subbands × 10 secret messages). For imperceptibility measures (Figs. 4 and 5), we use the following parameters: SSIM, GVD and correlation. To evaluate quality of the recovered audio (Figs. 6 and 7), we use SPCC.

4.1 Imperceptibility of the Stego Image

Imperceptibility aims to quantify the differences between the carrier-image and the stego-image. In this case it is desired that no appreciable distortion exists in the stego image, and therefore that there is no evidence that the image has been modified. To do this, the values of SSIM, GVD and SPCC should be as



(a) SPCC (b) SSIM

(c) GVD

Fig. 4. Imperceptibility results for three metrics. Confidence range (95 %) by Delta value.

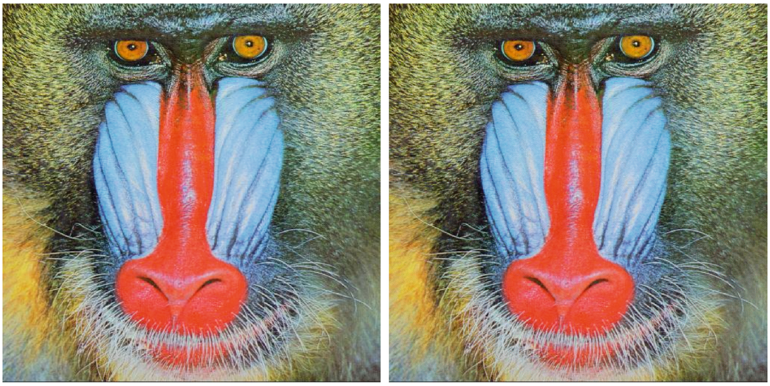
close to its ideal value, namely, 1, 0 and 1, respectively. To corroborate the imperceptibility of the method, the results for the three metrics above were consolidated by confidence intervals. That is, the ranges in which the 95 % of the values are concentrated were obtained, whereby it is possible to assess the trend of each metric. This evaluation was performed by varying the delta value from 5–50, in increments of 5, as shown in Fig. 4.

Figure 4(a), shows the variation of the correlation coefficient depending on the delta value. Here it is observed that the transparency decreases as the delta value increases, however all values are above 0.99, which means that in terms of correlation coefficient, the transparency is good. It also shows that delta values that present good imperceptibility results in terms of SSIM are lower than 20.

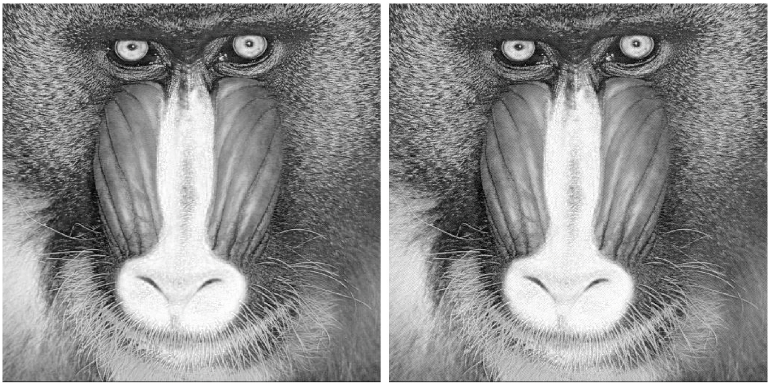
Regarding the SSIM index, Fig. 4(b), shows the trend of this metric at change the delta value. As in the correlation coefficient, the transparency decreases as the delta value increases, but this decrease is more pronounced compared to the previous index. However, for delta values lower than 20, SSIM tends to remain below 0.95.

The GVD meanwhile, gets values remains below 0.15, as shown in Fig. 4(c). As you increase the delta value, the GVD away from its ideal value, so in terms of GVD again small delta values are recommended.

Figure 5 shows the original image and the resulting one after using the proposed scheme. In this case, the HH sub-band, the red band and a delta value of 50 were used as input parameters. The quality metrics are showed in the caption. Although there are differences between the stego band and the original band, at



(a) Original RGB Image (b) Stego RGB Image



(c) Original band (d) Stego band

Fig. 5. Example of imperceptibility test. Quality metrics: SSIM=0.8240, GVD=0.0147 and SPCC=0.9810.

the end of the hiding process the stego RGB image is very similar to the original RGB image.

4.2 Quality of the Recovered Audio

Quality of the recovered audio signal quantifies the differences between the original audio signal and the recovered signal. In this case, should no exist appreciable distortion in the recovered signal. To do this, similarity between these two signals was calculated by means of the squared pearson correlation coefficient. SPCC values should be as close to its ideal value, namely, 0.

The Fig. 6 shows the confidence intervals for the SPCC between the original audio and the recovered audio signal. Here, unlike the evaluation of imperceptibility, increased delta value means a better quality of the output signal. However, according to the results of Fig. 6, if Delta is higher than 15, SPCC is close to 1 and then the quality of the recovered audio signal is very high.

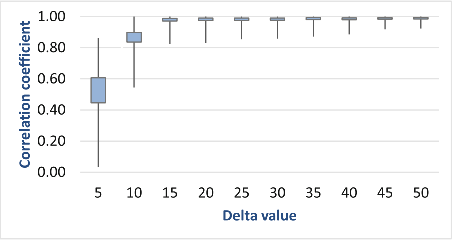
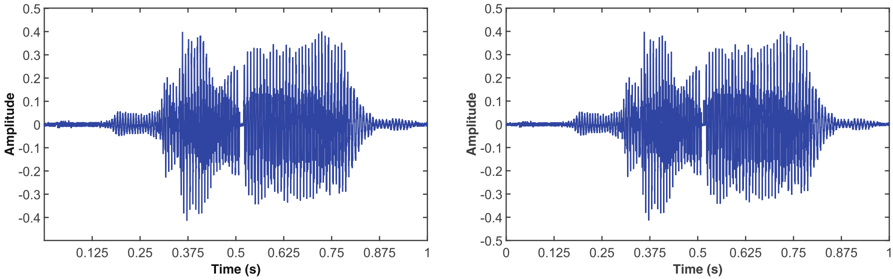


Fig. 6. Correlation coefficient for original and recovered audio signals. Confidence range (95 %) for SPCC by Delta value.

According to the above, the imperceptibility and quality of the recovered signal, have a tradeoff based on the delta value. Note that the higher the delta value, the lower the quality of stego image and lower the quality of the recovered audio. In any case, a delta value that presents a good compromise between imperceptibility and quality of the recovered signal is 15. In order to illustrate the performance of proposed system, Figs. 7(a) and 7(b) show the original audio signal and the recovered audio signal in the time domain. Here it is possible to confirm that the two signals have a high correlation degree.



(a) Original Audio Signal (b) Recovered Audio Signal

Fig. 7. Example of the quality of the recovered audio. Quality metric: SPCC=0.9969.

5 Conclusion

We propose a steganography scheme that allows to hide an audio signal into a RGB image. The process works on the wavelet domain with QIM method. Both imperceptibility and quality of the recovered audio depend on the Delta value. According to the results if Delta value increases, quality increases but imperceptibility decreases. A good tradeoff between them is obtained for a Delta value of 15.

Acknowledgment. This work is supported by the “Universidad Militar Nueva Granada-Vicerrectoría de Investigaciones” under the grant INV-ING-1768 of 2015.

References

1. Ballesteros L, D.M., Renza, D., Rincon, R.: Gray-scale images within color images using similarity histogram-based selection and replacement algorithm. J. Inf. Hiding Multimed. Sig. Process. **6**(6), 1156–1166 (2015)
2. Benesty, J., Chen, J., Huang, Y., Cohen, I.: Pearson correlation coefficient. Noise Reduction in Speech Processing. Springer Topics in Signal Processing, vol. 2, pp. 1–4. Springer, Heidelberg (2009)
3. Bhattacharyya, D., Dutta, P., Balitanas, M.O., Kim, T., Das, P.: Hiding data in audio signal. In: Chang, C.-C., Vasilakos, T., Das, P., Kim, T., Kang, B.-H., Khurram Khan, M. (eds.) ACN 2010. CCIS, vol. 77, pp. 23–29. Springer, Heidelberg (2010)
4. Chen, B., Wornell, G.W.: Quantization index modulation methods for digital watermarking and information embedding of multimedia. J. VLSI Signal Process. Syst. Signal Image video Technol. **27**(1–2), 7–33 (2001)
5. Delforouzi, A., Pooyan, M.: Adaptive digital audio steganography based on integer wavelet transform. Circuits Syst. Signal Process. **27**(2), 247–259 (2008)
6. Hemalatha, S., Acharya, U.D., Renuka, A.: Wavelet transform based steganography technique to hide audio signals in image. Procedia Comput. Sci. **47**, 272–281 (2015)
7. Hopper, N.J., Langford, J., von Ahn, L.: Provably secure steganography. In: Yung, M. (ed.) CRYPTO 2002. LNCS, vol. 2442, pp. 77–92. Springer, Heidelberg (2002)
8. Jero, S.E., Ramu, P., Ramakrishnan, S.: Discrete wavelet transform and singular value decomposition based ECG steganography for secured patient information transmission. J. Med. Syst. **38**(10), 1–11 (2014)
9. Li, S.B., Tao, H.Z., Huang, Y.F.: Detection of quantization index modulation steganography in G.723.1 bit stream based on quantization index sequence analysis. J. Zhejiang Univ. SCIENCE C **13**(8), 624–634 (2012)
10. Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. **11**(7), 674–693 (1989)
11. Marvel, L.M.: Information hiding: steganography and watermarking. In: Javidi, B. (ed.) Optical and Digital Techniques for Information Security. Advanced Sciences and Technologies for Security Applications, pp. 113–133. Springer, New York (2005)
12. Phadikar, A.: Multibit quantization index modulation: A high-rate robust data-hiding method. J. King Saud Univ. Comput. Inf. Sci. **25**(2), 163–171 (2013)
13. Rekik, S., Guerchi, D., Selouani, S.A., Hamam, H.: Speech steganography using wavelet and fourier transforms. EURASIP J. Audio Speech Music Process. **2012**(1), 1–14 (2012)
14. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)