No-reference visually significant blocking artifact metric for natural scene images

By: Shan Suthaharan

S. Suthaharan (2009), "No-reference visually significant blocking artifact metric for natural scene images," Signal Processing 89, pp. 1647-1652.

Abstract:
Quantifying visually annoying blocking artifacts is essential for image and video quality assessment. This paper presents a no-reference technique that uses the multi neural channels aspect of human visual system (HVS) to quantify visual impairment by altering the outputs of these sensory channels independently using statistical "standard score" formula in the Fourier domain. It also uses the bit patterns of the least significant bits (LSB) to extract blocking artifacts. Simulation results show that the blocking artifact extracted using this approach follows subjective visual interpretation of blocking artifacts. This paper also presents a visually significant blocking artifact metric (VSBAM) along with some experimental results.
Keywords: Natural scene images, JPEG compression, Neural channels, Blocking artifacts

Article:
1. Introduction
We can find several block impairment metrics in the image and video processing literature, however, only very few metrics are no-reference (NR) metrics. The NR-metrics assume no knowledge of the original image when estimating the blocking artifacts. Most recently, authors of paper [1] considered eight NR-metrics and presented extensive comparison results. They compared the following eight NR-metrics: mean squared difference of slope (MSDS) [2], boundary discontinuity metric (BDM) [3], phase correlation (PCM) [4], blocking artifact metric (BAM) [5], generalized block impairment metric (GBIM) [6], power spectrum metric (PSM) [7], DCT-step metric (DSM) [8] and perceptually significant block impairment metric (PSBIM) [9]. The first metric MSDS uses a new concept called "mean square difference slope" which characterizes the level of block-edge artifact as a change in the intensity slope along the block boundaries. BDM defines the block-edge artifact using the shape of the blocky noise and the discontinuity along the block boundary. The minimum mean square error (MSE) is then used to estimate blocking artifacts. PCM uses the phase correlation and defines the block detector metric as a ratio between the measure of inter-block and intra-block similarity [4]. BAM is based on the homogeneous image regions in the compressed image [5]. GBIM uses the intensity changes along the adjacent block boundaries and incorporates contrast masking in the compressed domain. PSM smoothes the power spectrum to extract the frequencies associated with the blocking artifact. DSM uses shifted block concept, edge information in the DC coefficients of shifted blocks and human visual system (HVS) characteristics. PSBIM generates perceptual weights using a stimulus called "gradient image" and then uses these weights to measure the blocking artifacts. The experimental results in [1] show that the quality metric GBIM performs better than others by satisfying most of the expectations that they defined. In recent years significant attention has been given to the quality measurement of natural scene images [10,11]. Most importantly natural scene statistics have been used in the development of a blind quality metric for JPEG2000 compressed images [12]. This paper presents a NR blocking artifacts quantifier (BAQ) for natural scene images and it is presented in Section 2. It consists of two units. The first unit measures the visibility of distortions as a combination of blocking artifacts and undistorted image edges. This is achieved by using a multi neural channels concept of HVS and the standard score (SS) formula. It is discussed in detail in Section 3. The second unit uses bit patterns of the least significant bits (LSBs) to identify image regions that are affected by JPEG compression. This process is

discussed in Section 4. A normalized visually significant blocking artifact metric (VSBAM) is presented in Section 5. The proposed VSBAM is compared with GBIM and differential mean opinion scores (DMOS) in Section 6. A conclusion is presented in Section 7.

2. Proposed model

Natural scene images are formed by several groups of pixels that show closely tight relationships in visual characteristics such as brightness, contrast and color [11]. Hence a JPEG compressed image displays a combination of primary edges, undistorted image edges and blocking artifacts, where blocking artifacts consist of distorted image edges and block edges. Hence mathematically we can write

$$CE = PE + UE + BA \text{ and } BA = DE + BE$$

where CE, PE, UE, DE, BE and BA represent edges in a compressed image, primary edges, undistorted edges, distorted edges, block-edges and blocking artifacts, respectively. The terms primary, undistorted, distorted and block edges are defined as follows:

- Primary edges: The edges that are derived from the DC-only image which is constructed using only DC values of iterative blocks of size 8 × 8 pixels as described in [8]. These edges primarily provide global description of the image and they are not affected by the compression schemes.

- Undistorted edges: The edges that are not primary edges as well as not affected by the compression scheme. These edges primarily provide local description of the image and they are not affected by the compression schemes.

- Distorted edges: They are mainly present in high activity areas (such as the image regions with high statistical variance) and along the sharp edges (such as high contrast edges).

- Block edges: They are mainly present in less activity areas (such as the image regions with medium or low statistical variance) and homogeneous areas (such as the image regions with near zero statistical variance).

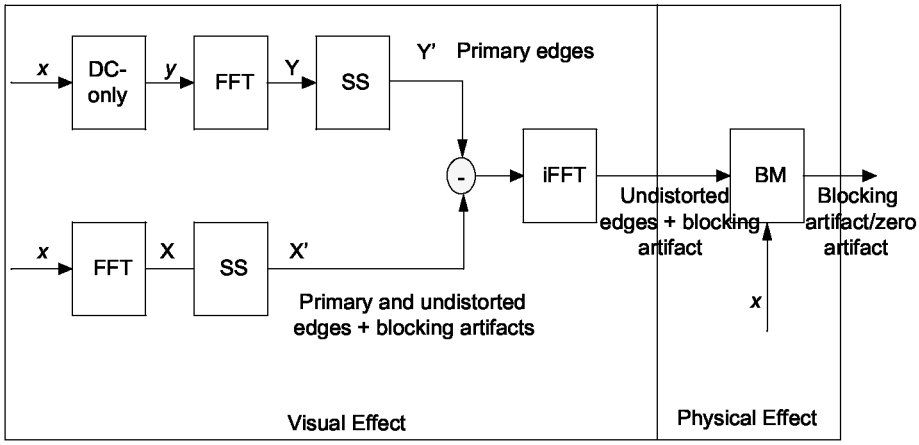

Fig. 1. Processes of the BAQ approach to quantify blocking artifacts.

The goal of this paper is to estimate PE and UE and then filter them out from CE to obtain an estimate for BA. Fig. 1 illustrates the proposed BAQ approach to estimate blocking artifacts. It shows two units that are labeled as the visual effect and the physical effect. The visual effect unit has two processes. The first process displays both primary and undistorted image edges as well as blocking artifacts. In this process image x is first transformed (FFT) into Fourier components X. Then neural channels are generated using different ranges of spatial frequencies. Subsequently Fourier components are modified (X') using the standard score (SS) formula in Eq. (1). The second process displays primary edges of the input image. The primary edges are detected using DC-only image y that is generated using the iterative (shifted) blocks of size 4 × 4 as explained in [8]. The image y is transformed into Fourier components Y and then the same SS formula is applied. The modified

Fourier components are denoted by *Y′*. In *iFFT* module, inverse FFT is applied to the absolute difference of *X′* and *Y′*. The physical effect unit uses the output of *iFFT* module and the bit patterns in the LSBs of the image *x* to output a BAM. It identifies three types (e.g. see Fig. 2(b)) of bit patterns in the LSBs: (i) blocks with random bit patterns (it is called random bit blocks) (ii) blocks with line-like patterns and (iii) homogeneous blocks. It uses the knowledge that the original scene images have more random bit blocks than their compressed images. Subsequently it considers edges in the image regions with random bit blocks as undistorted and filters them out from the output of *iFFT* module.

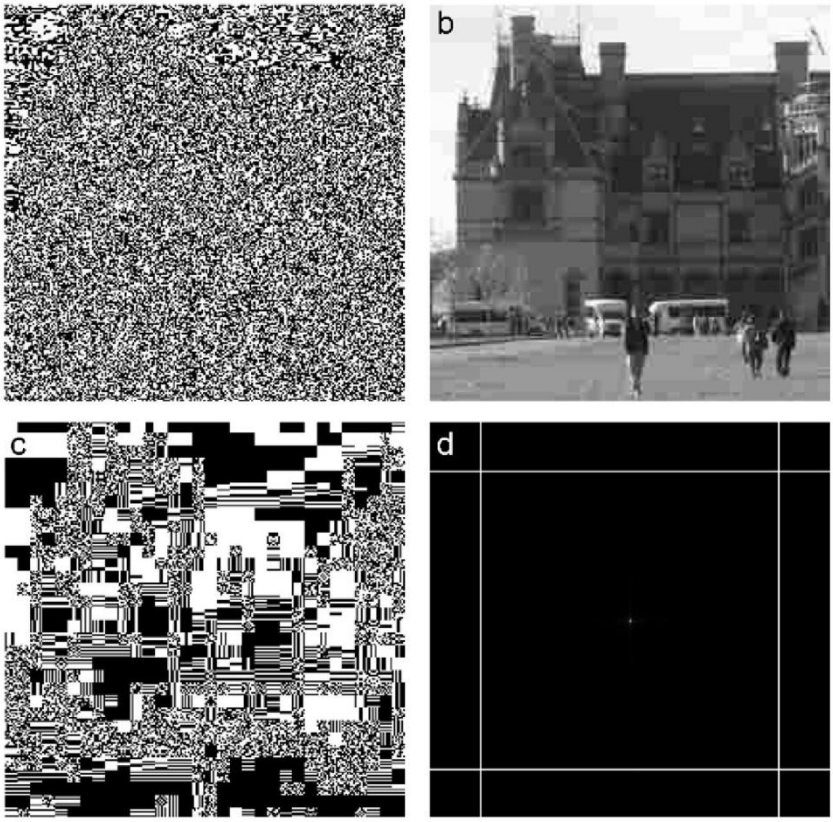

Fig. 2. (a) Random bit blocks in the original image. (b) Compressed "Biltmore Estate" image showing blocking artifacts. (c) Bit patterns in the LSBs of the compressed image. (d) Spatial frequencies of 2 vertical-channels and 2 horizontal-channels.

3. Effect of SS

The proposed BAQ is based on the findings in [13] that the HVS consists of a number of spatial-frequency channels and their outputs are detected independently. In recent years neural channels are defined using blocks of neighbor frequencies and combined with divisive normalization to reduce statistical dependencies of local image structures [14]. Innovatively, in the proposed BAQ, the neural channels are defined using relative vertical and horizontal frequencies and combined with statistical "standard score" formula to reduce statistical dependencies of local image structures.

Suppose $x(i,j)$ represents luminance value at (i,j)th pixel of the input image x of size $N \times N$ and $X(u,w)$ represents its Fourier component at (u,w)th frequency, where u and w represent horizontal and vertical frequencies, and $1 \leq i, j, u, w \leq N$. Then the proposed neural channels approach assumes that wth vertical-channel consists of all vertical spatial frequencies of the fixed wth horizontal frequency and that wth horizontal-channel consists of all horizontal spatial frequencies of the fixed wth vertical frequency. That is, the vertical-channel "w" and the horizontal-channel "w" consist of the following frequencies, respectively:

$$VF_w = \{(w, 1), (w, 2), \ldots, (w, N)\} \text{ and } HF_w = \{(1, w), (2, w), \ldots, (N, w)\}$$

For example, Fig. 2(d) illustrates some of those vertical-channel and horizontal-channel frequencies using vertical and horizontal white lines in Fourier domain. Hence the Fourier components (i.e. called channel outputs) of these vertical- and horizontal-channel frequencies are selected as follows:

$VS_w = \{X(w, 1), X(w, 2), \ldots, X(w, N)\}$ and $HS_w = \{X(1, w), X(2, w), \ldots, X(N, w)\}$

These channel outputs are altered independently assuming every channel w is sensitive to its own spatial frequencies. Hence the proposed BAQ uses the SS formula to alter spatial frequencies of a neural channel in order to standardize its sensitivity output such that it is independent from other channels. It is evidenced from [13] that the HVS contains several neural channels and each channel outputs a range of spatial frequencies. Subsequently each Fourier component in the set VS_w can be modified as follows (note: similarly Fourier component in the set HS_w can be modified):

$$VS'_w = \{X'(w, 1), X'(w, 2), \ldots, X'(w, N)\}$$

where

$$X'(w, u) = \left(\frac{X(w, u) - \mu_w}{\sigma_w / N}\right) \quad \text{and} \quad u = 1, \ldots, N \qquad (1)$$

The parameters μ_w and σ_w are the mean and the standard deviation of the Fourier components in the set VS_w. This modification affects the luminance values and provides meaningful edges. Eq. (1) plays a major role in the proposed approach as used in the following section (see Eq. (3)) to extract meaningful edges.

4. Effect of compression

The input image is divided into 8 x 8-pixels blocks to study the effect of JPEG compression on the LSBs of a natural scene image. A common property that is found in the original natural scene images used in this paper is that more than 91% of the blocks are random bit blocks. For example Fig. 2(a) shows the random bit blocks of the original "Biltmore Estate" image. A JPEG compressed version of this image is given in Fig. 2(b). The randomness property is tested using the "runs test" approach in [15]. Fig. 2(a) illustrates a large number of random bit blocks. The original image is captured using a Canon PowerShot A430 camera. When JPEG compression is applied line-like and homogeneous patterns are introduced in LSBs. Fig. 2(c) shows these LSB properties. Let us denote the set of random bit blocks by R and the sets of LSB blocks with line-like and homogeneity patterns by L and H, respectively. If R represents the ratio between the number of random bit blocks and the total number of LSB blocks then R increases proportional to the compression ratio. We use R as an estimate to the compression ratio and define a global mask as follows

$$m(i,j) = \begin{cases} 1 - R & x(i,j) \in \mathbf{R} \\ 1 & x(i,j) \in \mathbf{L} \text{ or } \mathbf{H} \end{cases} \qquad (2)$$

This mask is used to separate blocking artifact (as shown in Eq. (6) later) from the distortions calculated using the concept of neural channels with predetermined spatial frequencies.

5. Visual quality metric

Suppose $x(i,j)$ represents luminance value at (i,j)th pixel of the input image x of size $N \times N$ and $X(u,w)$ represents its Fourier component at (u,w)th frequency. Similarly, suppose $y(i,j)$ represents luminance value at (i,j)th pixel of the DC-only image y and $Y(u,w)$ represents the Fourier component at (u,w)th frequency. We then define (using Eq. (1)) vertical blocking artifact with undistorted edges in frequency domain as:

$$A_1(u,w) = \left(\frac{X(u,w) - \mu_u}{\sigma_u / N}\right) - \left(\frac{Y(u,w) - \mu'_u}{\sigma'_u / N}\right) \qquad (3)$$

The parameters μ_u, σ_u, μ'_u and σ'_u are defined as follows:

$$\mu_u = \sum_{v=1}^{N} X(u,v) \Big/ N, \quad \sigma_u = \sum_{v=1}^{N} (X(u,v) - \mu_u)^2 \Big/ N \qquad (4)$$

$$\mu'_u = \sum_{v=1}^{N} Y(u,v) \Big/ N \quad \text{and} \quad \sigma'_u = \sum_{v=1}^{N} (Y(u,v) - \mu'_u)^2 \Big/ N \qquad (5)$$

The first and the second terms on the RHS of Eq. (3) are derived from Eq. (1). The first term represents a combination of primary and undistorted edges as well as the blocking artifact and the second term represents the primary edges. Similarly the horizontal blocking artifact with undistorted edges $A_2(u,w)$ can be defined. Then

the blocking artifacts with undistorted edges in spatial domain is the inverse FFT of $(A_1(u,w)+A_2(u,w))/2$ and it is denoted by $B(i,j)$ and it is defined as follows:

$$b(i,j) = m(i,j) \otimes B(i,j) \quad (6)$$

where $m(i,j)$ is the mask defined in Eq. (1) and the operator $\otimes$ represents pixel-by-pixel multiplication. This equation is called BAQ. Using this quantifier the following normalized visually significant blocking artifacts metric is defined:

$$VSBAM = \left| \frac{\sum_i \sum_j b(i,j)}{\sum_i \sum_j (B(i,j) + y(i,j)) - \sum_i \sum_j b(i,j)} \right| \quad (7)$$

The numerator and denominator represent blocking artifacts and, primary and undistorted edges. The main motivation behind this metric is to normalize the strength of blocking artifacts between 0 and 1.

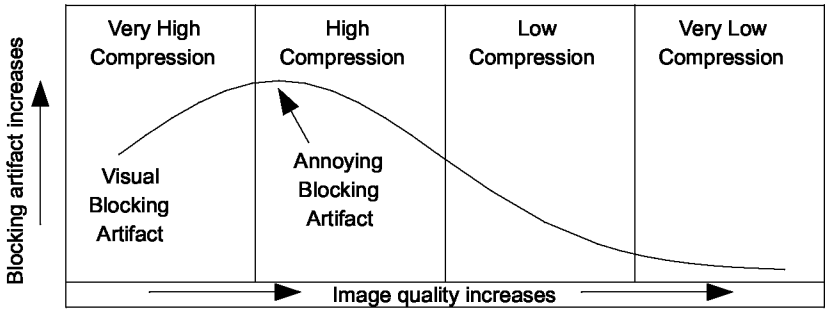

Fig. 3. It shows a theoretical visual interpretation of the blocking artifacts.

6. Subjective tests results

We have used the analogy of annoying blocking artifacts in Fig. 3 to interpret the results of our experiments. It is a simple model for the visual interpretation of the blocking artifacts and it is derived using our observation with the LIVE image database [16]. JPEG compression introduces blur and blocking artifacts. These artifacts are visible when image quality is low. In this case annoyance of blocking artifact is less effective due to blur interference. When image quality is medium, blocking artifacts dominate hence it is visibly annoying. When image quality is high, both of these artifacts are invisible. This analogy is validated using the DMOS values of JPEG compressed images in the LIVE database. For example, the DMOS values of the "Stream" compressed images in LIVE are 64.4151, 75.2790, 63.6490, 52.5575, 29.7388, 21.3458, and 21.0245 with bit rates of 0.20131, 0.29203, 0.41781, 0.57467, 1.0008, and 1.6833, respectively. It is clear that these DMOS values follow the shape of the curve in Fig. 3 as image quality increases. As the next step, we com-pressed the "Biltmore Estate" image using JPEG with quantization parameter, indicating 1 as low quality and 31 as high quality. This gives us 31 degraded images of "Biltmore Estate" that are adjacent to each other. BAQ values are calculated for these 31 compressed images and plotted against image quality increase in Fig. 4. We can clearly see that the shape of this graph also follows the theoretical shape of the curve in Fig. 3. Similarly 9 more natural scene images (captured using Canon PowerShot A430 camera) were tested and their BAQs also follow the shape of the curve in Fig. 3. This is one of the empirical supports to confirm that the proposed BAQ follows the subjective (DMOS) interpretation of blocking artifacts. In addition to these 10 natural scene images, we used DMOS values of JPEG compressed versions of ''I1: Lighthouse (6)", ''I2: Stream (6)", ''I3: Churchcapital (5)", "I4: Flower (6)", ''I5: House (7)", ''I6: Buildings (7)" and "I7: Studentsculpture (6)" images in LIVE database. Table 1 uses the labels I1–I7 as column headings. The integers in parentheses represent the number of compressed images (or number of DMOS values) available for that image. For each set of these images, BAQ values are calculated. Then the correlations between these BAQs and their DMOS values are calculated and presented in row 2 of Table 1. The high correlation values support that BAQ accurately represents the subjective (DMOS) interpretation of blocking artifacts. Next the performance of VSBAM is compared with that of GBIM. To achieve this, VSBAM and GBIM values of the same 7 sets (I1–I7) of LIVE images are calculated and compared with DMOS values. Table 1 shows the correlations of VSBAM and GBIM with their corresponding DMOS values. Over all, the higher correlation values of VSBAM indicate that the VSBAM represents subjective (DMOS) interpretation of blocking artifact more accurately than GBIM. Fig. 5 shows the average VSBAM values of our 10 natural scene images at each quality level (1–31). It exhibits the range of numerical values of VSBAM between 0 and 1 representing a value closer to 1 as high blocking artifacts and a

value closer to 0 as low blocking artifact. When blocking artifact is high, primary and undistorted edges in the denominator of Eq. (7) is the same as block edges. Hence VSBAM is closer to 1. When blocking artifact is low, the numerator of Eq. (7) is 0, hence VSBAM is closer to 0.

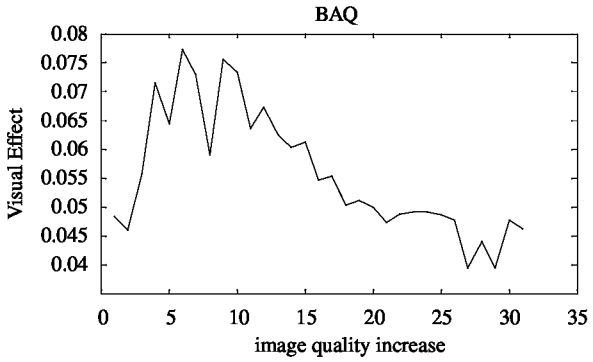

Fig. 4. Quantified visual blocking artifacts of the image in Fig. 2(b).

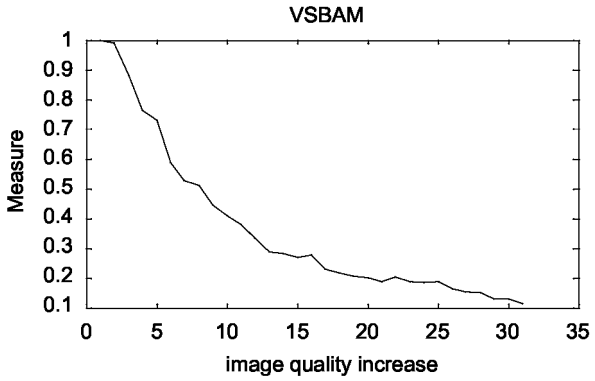

Fig. 5. The graph that represents the changes in the proposed VSBAM.

Table 1
Correlations with DMOS.

Metric	I1	I2	I3	I4	I5	I6	I7
BAQ	0.98	0.96	0.93	0.99	0.98	0.97	0.91
VSBAM	0.99	0.84	0.83	0.98	0.97	0.93	0.90
GBIM	0.98	0.56	0.95	0.71	0.79	0.63	0.86

7. Conclusion

A novel approach to quantify visually annoying blocking artifacts and a normalized metric to measure the visually significant blocking artifacts are proposed. The advantages of this approach and the metric are that they extract visually important blocking artifacts and provide a meaningful bounded range between 0 and 1 for the metric. The proposed techniques may have limitations when used with the images (original or lightly compressed images) that do not satisfy LSB bit patterns. The proposed blocking artifact quantifier and the visually significant blocking artifact metric can be used for both image and video quality assessment of natural scene images.

References

[1] A. Leontaris, P.C. Cosman, A.R. Reibman, Quality evaluation of motion-compensated edge artifacts in compressed video, IEEE Trans. Image Process. 16 (4) (April 2007) 943–956.

[2] S. Minami, A. Zakhor, An optimization approach for removing blocking effects in transform coding, IEEE Trans. Circuits Syst. Video Technol. 5 (2) (April 1995) 74–82.

[3] J. Yang, H. Choi, T. Kim, Noise estimation for blocking artifacts reduction in DCT coded images, IEEE Trans. Circuits Syst. Video Technol. 10 (7) (October 2000) 1116–1134.

[4] T. Vlachos, Detection of blocking artifacts in compressed video, IEE Electron. Lett. 36 (13) (June 2000) 1106–1108.

[5] W. Gao, C. Mermer, Y. Kim, A de-blocking algorithm and a blockiness metric for highly compressed images, IEEE Trans. Circuits Syst. Video Technol. 12 (12) (December 2002) 1150–1159.

[6] H.R. Wu, M. Yuen, A generalized block-edge impairment metric for video coding, IEEE Signal Process. Lett. 4 (11) (November 2000) 317–320.

[7] Z. Wang, A.C. Bovik, B.L. Evans, Blind measurement of blocking artifacts in images, in: Proceedings of the IEEE International Conference on Image Processing, vol. 3, 2000, pp. 981–984.

[8] S. Liu, A.C. Bovik, Efficient DCT-domain blind measurement and reduction of blocking artifacts, IEEE Trans. Circuits Syst. Video Technol. 12 (12) (December 2002) 1139–1149.

[9] S. Suthaharan, Perceptual quality metric for digital video coding, Electron. Lett. 39 (5) (March 2003) 431–433.

[10] C. Kayser, W. Einhauser, P. Konig, Temporal correlations of orientations in natural scenes, Neurocomputing 52–54 (2003) 117–123.

[11] Q. Ye, J. Jiao, J. Huang, H. Yu, Text detection and restoration in natural scene images, J. Vis. Commun. Image R 18 (2008) 504–513.
[12] H.R. Sheikh, A.C. Bovik, L. Cormack, Blind quality assessment of JPEG2000 compressed images using natural scene statistics, in: Proceedings of IEEE Asilomar Conference on Signals, Systems, and Computers, vol. 2, November 2003, pp. 1403–1407.
[13] M.B. Sachs, J. Nachmias, J.G. Robson, Spatial-frequency channels in human vision, J. Opt. Soc. Am. 61 (9) (1971) 1176–1186.
[14] M.J. Wainwright, O. Schwartz, E.P. Simoncelli, Natural image statistics and divisive normalization: modeling nonlinearities and adaptation in cortical neurons <http://www.cns.nyu.edu/pub/eero/wainwright00c.pdf>.
[15] L.C. Alwan, Statistical Process, McGraw-Hill, New York, 2000, pp. 119–126.
[16] H.R. Sheikh, Z. Wang, L. Cormack, A.C. Bovik, LIVE Image Quality Assessment Database Release 2 <http://live.cse.utexas.edu/re-search/quality>.