Optik 125 (2014) 6205–6209

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Multi-scale SSIM metric based on weighted wavelet decomposition

Fang Qian [a,b,*], Jin Guo [a], Tao Sun [a], Tingfeng Wang [a]

[a] Changchun Institute of Optics, Fine Mechanics and Physics Chinese Academy of Sciences, State Key Laboratory of Laser Interaction with Matter, Changchun, Jilin 130033, China
[b] University of the Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Article history:
Received 5 November 2013
Accepted 5 June 2014

OCIS:
110.0110
110.3000
140.0140
140.3330

Keywords:
Image quality assessment performance
Image quality study
Subject quality assessment
Multi-resolution analysis
Error sensitivity

ABSTRACT

Image quality assessment aims to use computational models to assess the image quality consistently with subjective evaluations. This paper proposes a new metric composed of weighted wavelet multi-scale structural similarity (WWMS-SSIM). Four-level 2-D wavelet decomposition is performed for the original and disturbed images, respectively. Each image can be partitioned into one low-frequency subband (LL) and a series of octave high-pass subbands (HL, LH and HH). Different subbands are processed with different weighting factors. Based on the results of the above, we can construct a modified WWMS-SSIM. Comparison experiments show that the correlation, prediction accuracy and consistency of the proposed metric are respectively 5.8%, 5.2% and 4.8% higher than the PSNR metric. The correlation, prediction accuracy and consistency of the proposed metric are respectively 0.7%, 1.6% and 2.6% higher than the SSIM metric. In terms of the experiment results, the WWMS-SSIM metric shows good feasibility comparing with PSNR and SSIM methods.

1. Introduction

Image quality assessment (IQA) aims to design quality measures that can automatically predict the image quality. The simplest and most widely used objective quality metrics are the mean squared error (MSE) and the peak signal-to-noise ratio (PSNR). They are simple to calculate and have clear physical meanings, but they do not correlate well with the subjective evaluations. The well known structural-similarity (SSIM) metric brings IQA from pixel-based stage to structure-based stage [1]. On the basis of SSIM, there have been many improved algorithms. The multi-scale structural similarity (MS-SSIM) produces better results than single-scale structural similarity [2]. The three-component weighted structural similarity (three-SSIM) metric assigns different weights to the SSIM scores according to the local region type, such as edge, texture and smooth area [3]. In [4], the authors presented an attention shift mechanism based image quality assessment (PP-SSIM). But unfortunately, they are inconsistent with the subjective rating on cross-distortion image assessment.

Based on the above analysis, in this paper we propose a novel weighted wavelet multi-scale structural similarity, named WWMS-SSIM. We first carry out the wavelet transforms upon the both original and the distorted images for the subbands. Thereafter, the changes between the original subbands and the distorted subbands are measured by SSIM for the SSIM maps. The SSIM maps are weighted by taking characteristics of human visual system into account. Finally, we use a mean SSIM index to evaluate the overall SSIM maps. To examine the performance of the proposed WWMS-SSIM metric, MSE, PSNR, SSIM and WWMS-SSIM are used to evaluate the image quality and the experiment is focused on the correlation, accuracy and consistency, respectively.

2. Weighted wavelet multi-scale SSIM

Human visual system has different response to different colors, directions and spatial frequency. Consequently, the multi-resolution analysis is very important in the image processing. It means the general structure is observed in the low-resolution and the detail information is observed in the high-resolution. Based on the above analysis, a weighted wavelet multi-scale structural similarity (WWMS-SSIM) metric is proposed. The original and distorted images are decomposed into four wavelet levels firstly. The images are divided into different scales and spatial frequency subbands. After that, a luminance comparison is calculated in the

* Corresponding author at: Changchun Institute of Optics, Fine Mechanics and Physics Chinese Academy of Sciences, State Key Laboratory of Laser Interaction with Matter, Changchun, Jilin 130033, China.
E-mail address: qfmail@sina.cn (F. Qian).

http://dx.doi.org/10.1016/j.ijleo.2014.06.134
0030-4026/

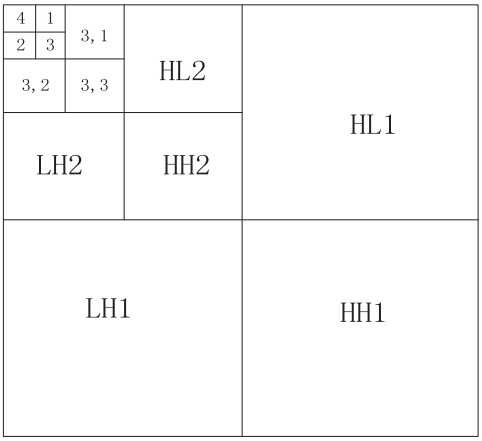

Fig. 1. A four-level wavelet transform map.

low-frequency subband and the contrast comparisons and structure comparisons are calculated in the high-frequency subbands. The different subbands are weighted by taking characteristics of human visual system into account. The mean weighted subband value is the final result. WWMS-SSIM metric evaluates the image quality based on characteristics of the human visual system. It is better matched to perceived visual quality than MSE and PSNR. WWMS-SSIM metric considers the different spatial frequency impacting on the image quality, so it is more accurate than SSIM in evaluating the cross-distorted images.

2.1. Wavelet decomposition

The multi-resolution analysis from rough to fine commonly uses octave division. The image is decomposed on a logarithmic scale, in different frequency bands with equal bandwidth [5]. SSIM metric evaluates the image quality through the loss of structure in the image. It provides better approximation of perceived image quality than MSE and PSNR, but it ignores that the visual system concerns the different detail with different spatial frequency. So SSIM cannot give the accurate assessment in evaluating the cross-distorted images. Wavelet decomposition is used octave division, which is close to the human visual system information processing mechanism, so it can be used in the image quality assessment in order to get more accurate value. Wavelet decomposition not only has the ability to characterize the local signal characteristics in the time domain and frequency domain, but also has the multi-scale analysis ability and the direction-sensitive property [6,7]. Therefore, the proposed WWMS-SSIM metric based on the wavelet decomposition can achieve comparable performance with the human subjective perception.

After the wavelet decomposition, the image is decomposed into a set of subband images which resolutions are gradually reduced. As shown in Fig. 1, the low level subband image (LL) is decomposed into different directions and frequencies secondary subband images. LL subband represents the low frequency information of the image including most of the energy of the image. LH subband represents the horizontal edge information (horizontal low frequency, vertical high frequency). HL subband reflects the vertical edge information (horizontal high frequency, vertical low frequency). HH subband reflects the diagonal edge information including the horizontal and vertical high frequency detail [8].

Five-level or six-level wavelet decomposition is more consistent with human visual characteristics, but the computation is too complex. Four-level wavelet decomposition is efficient in practice [9,10]. In this paper, four-level wavelet decomposition by "sym8" filters is used to get 13 different bands.

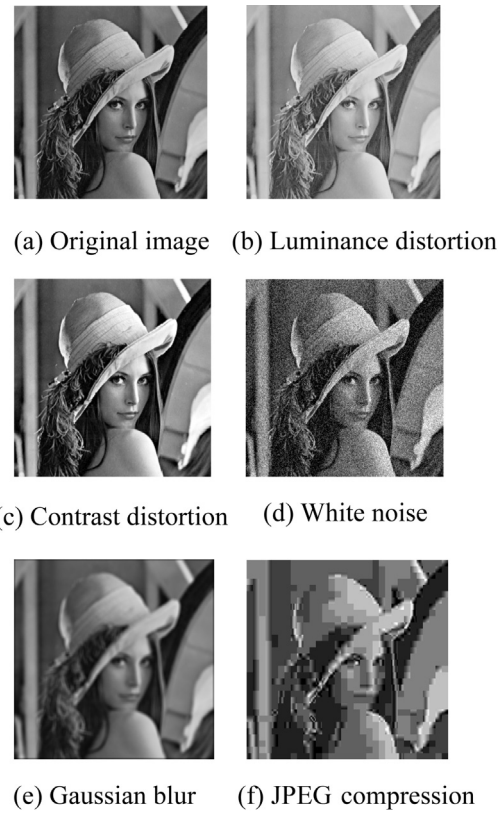

Fig. 2. Lena images with different types of distortions.

2.2. Wavelet multi-scale SSIM metric

Distorted images are classified into two categories, namely non-structural distortion and structural distortion. After the wavelet decomposition, the images are decomposed into different directions and different spatial frequency subband images. According to the characteristics of the human visual system, the different spatial frequency components in the image have different sensitivity toward non-structural or structural distortion, so the different distorted types have different effects on image quality. Fig. 2 shows different types of distorted images, (a) is an original image, (b) is a luminance distorted image, and (c) is a contrast distorted image. They are all non-structural distorted images. (d)–(f) are white noise, Gaussian blur, JPEG compression distorted images, respectively. They are all structural distorted images. For convenience, original and distorted images are decomposed into subbands by one-level wavelet decomposition. Each image is decomposed into a low-frequency subband (LL) and three high-frequency subbands (HL, LH, and HH), combining the high-frequency subbands obtains:

$$highmap = \sqrt{HL^2 + LH^2 + HH^2} \tag{1}$$

Luminance comparison, contrast comparison and structural comparison of the low-frequency LL and high-frequency *highmap* are calculated using Eqs. (2)–(4).

Luminance comparison is defined as

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{2}$$

Contrast comparison is defined as

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{3}$$

Table 1
Image quality assessment of different frequency.

Comparison	Frequency	Non-structure distortion		Structure distortion		
		Luminance distortion	Contrast distortion	White noise	Gaussian blur	JPEG compression
l	Low	0.8887	0.9898	0.9998	0.9979	0.9993
	High	0.9998	0.9998	0.9514	0.9984	0.9898
c	Low	0.0435	0.0430	0.0435	0.0434	0.0434
	High	0.0077	0.0045	0.0078	0.0079	0.0076
s	Low	1	1	0.9980	0.9934	0.9895
	High	0.9929	0.9993	0.6013	0.8492	0.7576

Structure comparison is defined as

$$s(x,y)=\frac{\sigma_{xy}+C_3}{\sigma_x^2\sigma_y^2+C_3} \tag{4}$$

where μ_x and μ_y represent the mean gray value of the subband images x and y. σ_x and σ_y represent the standard deviation of x and y. σ_{xy} represents the covariance between x and y. C_1, C_2 and C_3 is small enough to be ignored. After calculation, the comparison functions of the low and high frequency in various distorted images are given in Table 1.

From the underlined values in Table 1, we can see that the changes of the luminance have much effect on the low frequency components, but the changes of the contrast and the structure have little effect. The high frequency components are just the reverse of the low frequency. It is not sensitive to the distortion of the luminance, but it is sensitive to the contrast and structure. Consequently, wavelet multi-scale structural similarity (WMS-SSIM) metric only calculates the luminance comparison in the low-frequency subband and calculates the contrast and structure comparison in the high-frequency subbands. Four-level low-frequency wavelet subband structural similarity is defined as

$$\text{WMS-SSIM}_4^{\text{LL}}(x,y)=[l_4^{(\text{LL})}(x,y)]^{\alpha} \tag{5}$$

Four-level high-frequency wavelet subband structural similarity is defined as

$$\text{WMS-SSIM}_j^{(i)}(x,y)=[c_j^{(i)}(x,y)]^{\beta}[s_j^{(i)}(x,y)]^{\gamma}\quad (i)=\{\text{LH, HL, HH}\},\quad j=\{1,2,3,4\} \tag{6}$$

where l, c and s is defined in Eqs. (2)–(4), j is the wavelet decomposition level, i is the wavelet decomposition direction. Where $\alpha>0$, $\beta>0$ and $\gamma>0$ are parameters used to adjust the relative importance of the three components. Set $\alpha=\beta=\gamma=1$ in this paper.

2.3. Weighted wavelet coefficients

It has been reported that human visual system has different sensitivity to signals with different frequency. Contrast sensitivity function (CSF) has different response to different spatial frequency, so it can be used to compensate the lack of the sensitivity because of the too large or too small spatial frequency. Among those previous efforts, CSF can be demonstrated as an effective tool to calculate the sensitivity of human visual system in different frequency.

In the basic CSF, it is defined

$$H(r)=2.6\times(0.0192+0.114r)\exp(-(0.114r)^{1.1})\quad r=\sqrt{R_f^2+C_f^2} \tag{7}$$

Where r stands for the spatial frequency, R_f and C_f are the spatial frequency in the horizontal and vertical direction respectively. Which can be computed by Eq. (8) as introduced in [10,11]:

$$R_f=\sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=2}^{N}[f(i,j)-f(i,j-1)]^2}$$
$$C_f=\sqrt{\frac{1}{MN}\sum_{i=2}^{M}\sum_{j=1}^{N}[f(i,j)-f(i-1,j)]^2} \tag{8}$$

Consider an M × N image $f(i,j)$, where M is the number of rows and N is the number of columns. when $R_f=C_f$, the spatial frequency in the diagonal direction can be easily computed as $r_d=\sqrt{2}\times R_f$.

$$H(r)=2.6\times(0.0192+0.114\times\sqrt{2}R_f)\exp(-(0.114\times\sqrt{2}R_f)^{1.1}) \tag{9}$$

For four-level wavelet decomposition, images should be decomposed into 13 subbands, the CSF curves are divided into five bands (HL and LH are considered as the same type). Fig. 3 shows the CSF curve of HL/LH subbands with weighting factors. Fig. 4 is the corresponding CSF curve in HH direction.

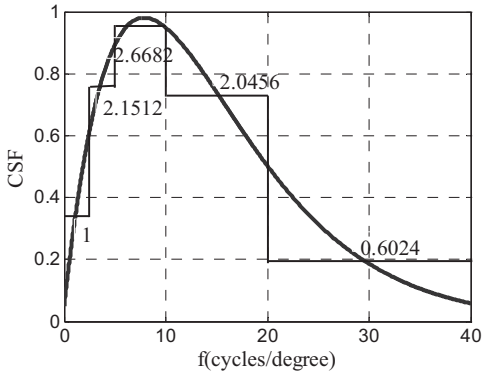

Fig. 3. A weighted CSF curve in LH/HL direction.

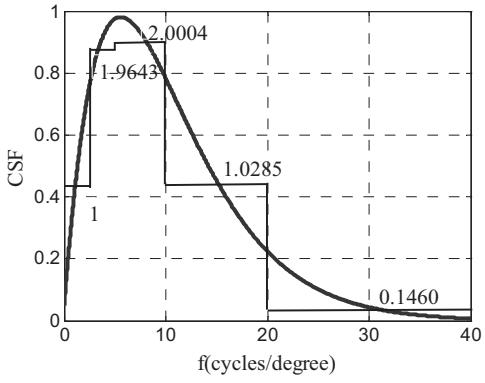

Fig. 4. A weighted CSF curve in HH direction.

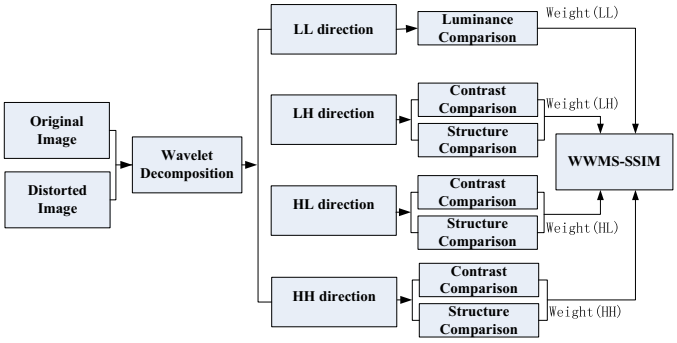

Fig. 5. Diagram of the measurement system.

2.4. Weighted wavelet multi-scale structural similarity metric

Based on the above analysis, weighted wavelet structural similarity (WWMS-SSIM) metric is defined as

$$\text{WWMS-SSIM} = \frac{\sum_{(i)}\sum_{j=1}^{4} W(j) \times \text{WMS-SSIM}_j^{(i)}(x,y) + \text{WMS-SSIM}_4^{\text{LL}}(x,y)}{\sum_{j=1}^{4} W(j) + 1} \quad (i) = \{\text{LH, HL, HH}\} \tag{10}$$

where WMS-SSIM is defined in Eqs. (5) and (6).

Four-level 2-D wavelet decomposition is performed for the original and distorted images, respectively. Each image can be decomposed into one low-frequency subband (LL) and a series of octave high-frequency subbands (HL, LH and HH). Wavelet coefficients of the original image are then compared with that of the distorted image by WMS-SSIM. The nonlinear band-pass characteristic of CSF is applied to weight WMS-SSIM value. Different subbands are processed with different weighting factors. The procedures to calculate WWMS-SSIM are illustrated in Fig. 5.

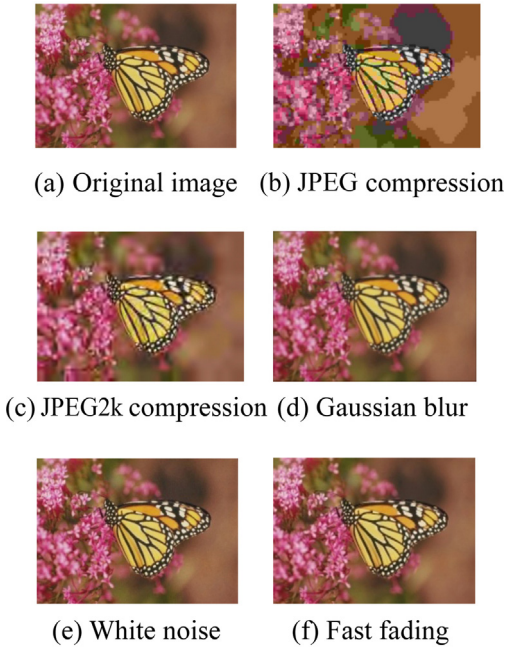

Fig. 6. Monarch images with different types of distortions.

Table 2
Images quality assessment using different algorithms.

	b	c	d	e	f
MSE	174.978	117.149	58.188	26.813	46.474
PSNR	25.701	27.443	30.482	33.847	31.458
SSIM	0.8365	0.8698	0.9539	0.9482	0.9686
WWMS-SSIM	0.6923	0.7071	0.8061	0.9787	0.9304

3. Experimental result and analysis

3.1. Cross-distorted image quality assessment

We compare the cross-distorted image quality in LIVE image database database. Fig. 6(a) is an original image. Fig. 6(b)–(f) are JPEG compression, JPEG2K compression, Gaussian blur, white noise, fast fading distorted images.

In this experiment, we examined the performance of the MSE, PSNR, SSIM and WWMS-SSIM on the cross-distortion image quality assessment in Fig. 6. The results are given in Table 2.

Fig. 6(e)–(f)–(d)–(c)–(b) image quality is decreased in the order from human subjective feeling. MSE should increase gradually and PSNR, SSIM, WWMS-SSIM should reduce gradually. From Table 2, the trend of MSE, PSNR and SSIM are inconsistent with the subjective assessment. But WWMS-SSIM predicts the image quality consistently with human subjective evaluations. Therefore, it can reflect the image visual quality more accurately.

3.2. Image database quality assessment

In order to provide quantitative measures on the performance of the objective quality assessment, we follow the performance evaluation procedures employed in the video quality experts group (VQEG). Logistic functions are used in a fitting procedure to provide

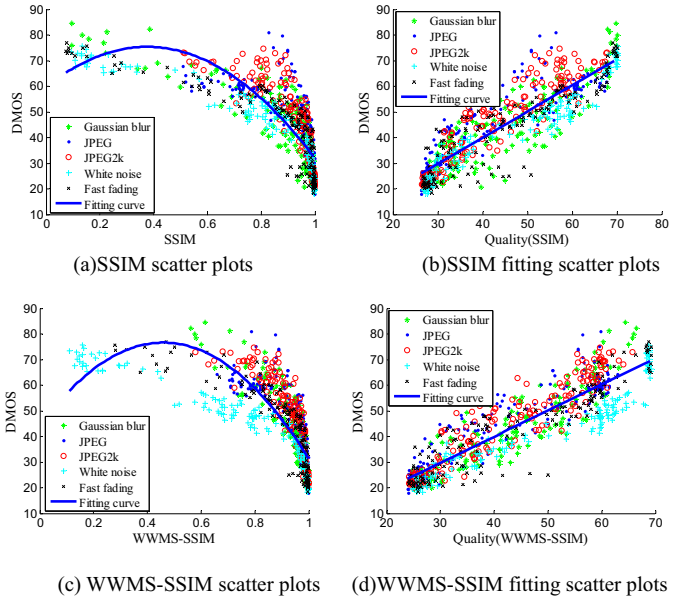

Fig. 7. Scatter plots for different algorithms.

Table 3
Performance comparison of image quality assessment algorithms.

	CC	MAE	RMSE	OR
PSNR	0.8716	7.3613	9.2909	0.4018
SSIM	0.9157	5.7204	7.4669	0.0039
WWMS-SSIM	0.9222	5.6095	7.2556	0.0013

a nonlinear mapping between the objective and subjective scores. The mapping function is Eq. (11).

$$\text{Quality}(x) = \frac{\beta_1 - \beta_2}{1 + e^{-((x-\beta_3)/(|\beta_4|))}} + \beta_2 \tag{11}$$

Fig. 7 shows the scatter plots of subjective DMOS versus the predicted scores by WWMS-SSIM and SSIM indices on the LIVE database. Fig. 7(a) is SSIM scatter plots. The curve shown in Fig. 7(b) is obtained by nonlinear fitting according to Eq. (11). Fig. 7(c) is WWMS-SSIM scatter plots. The curve shown in Fig. 7(d) is obtained by nonlinear fitting according to Eq. (11). From Fig. 6, we can see that the objective scores predicted by WWMS-SSIM are more consistently with the subjective evaluations than SSIM.

In this experiment, we use the correlation coefficients (CC) score, which is a widely accepted evaluation measure for image quality assessment metrics. The CC between objective and subjective scores after non-linear regression analysis provides an evaluation of prediction correlation. The outlier ratio (percentage of the number of predictions outside the range of ±2 times of the standard deviations) after the non-linear regression analysis measures the prediction consistency. In addition to these, we also calculate the root mean square error (RMSE) and mean absolute prediction error (MAE) to measure the prediction accuracy. The evaluation results for all the models being compared are summarized in Table 3.

The CC value is the larger the better. The RMSE, MAE and OR values are the smaller the better. From Table 3, we can see that the correlation, prediction accuracy and consistency of the WWMS-SSIM metric are 5.8%, 5.2% and 4.8% higher than PSNR, respectively. The correlation, prediction accuracy and consistency of the proposed metric are respectively 0.7%, 1.6% and 2.6% higher than SSIM. WWMS-SSIM performs better than all of the other metrics being compared.

4. Conclusion

In this paper, we proposed a novel image quality assessment metric based on weighted wavelet decomposition, namely weighted wavelet multi-scale structural similarity (WWMS-SSIM) metric. The principle of the wavelet decomposition represents the idea of the multi-resolution. The CSF weighted factors are used in WWMS-SSIM and they represent the characteristic of the human visual system. Particularly, the objective scores predicted by WWMS-SSIM are more consistent with the subjective evaluations. The WWMS-SSIM metric has better correlation, accuracy and consistency than PSNR or SSIM.

Acknowledgement

This work was funded by the Independent basic research of State Key Laboratory (grant no.: SKLLIM1203-01).

References

[1] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (2004) 6–7.
[2] Z. Wang, E.P. Simoncelli, A.C. Bovik, Multi-scale structural similarity for image quality assessment, in: Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, 2004, IEEE, 2003, pp. 1–4.
[3] C. Li, A.C. Bovik, Three-component weighted structural similarity index, Proc. SPIE 7242 (2009) 1–9.
[4] H. Zhang, Y.X. Xia, W.H. Zhou, Attention shift mechanism based image quality assessment, J. Sci. Instrum. 31 (2010) 2056–2061.
[5] M. Barkowsky, B. Eskoer, J. Bialkowski, et al., Temporal trajectory aware video quality measure, IEEE J. Sel. Topics Signal Process. 3 (2009) 266–279.
[6] K.Q. Huang, Z.Y. Wu, S.K.F. George, et al., Color image denoising with wavelet thresholding based on human visual system model, Signal Process. Image Commun. 20 (2005) 115–127.
[7] G.W. Cermak, Consumer opinions about frequency of artifacts in digital video, IEEE J. Sel. Topics Signal Process. 3 (2009) 336–343.
[8] M.J. Tsai, C.H. Shen, Wavelet tree group modulation (WTGM) or digital image watermarking, in: IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007, vol. 2, IEEE, 2007, pp. 173–176.
[9] D.M. Chandler, S.S. Hemami, A wavelet based visual signal-to-noise ratio for natural images, IEEE Trans. Image Process. 16 (2007) 2284–2298.
[10] M.A. Saad, A.C. Bovik, C. Charrier, A DCT statistics-based blind image quality index, IEEE Signal Process. Lett. 17 (2010) 583–586.
[11] M.J. Nadenau, J. Reichel, M. Kunt, The effects of a visual fidelity criterion on the encoding of images, IEEE Trans. Inform. Theory 20 (1974) 527–528.