

Image Quality Assessment Based on Contourlet and ESD Method

Shuang Liang¹, Guan-Xiang Wang², Lei Sun³

¹*School of Mathematical Science, Peking University, Beijing P.R. China*

²*School of Mathematical Science, Peking University, Beijing P.R. China*

³*Army Aviation Institute, Beijing P.R. China*

¹liangshuang12@pku.edu.cn, ²gxwang@math.pku.edu.cn, ³waterprison@163.com

Abstract: In recent years, the development of the digital image processing promotes the research of the image quality assessment (IQA). A novel metric for full-reference image quality assessment is presented. The metric combined the contourlet transform with the energy of structural distortion (ESD), namely the CT-ESD. The calculation of the ESD is carried out in each subband of the contourlet transform. Then the comparisons between the reference and the distorted images on each subband are integrated by weighting sum. The superiority of the contourlet transform integrates well into new IQA metric. Experiments performed on the database TID2013 demonstrate that the CT-ESD can achieve high consistency with the subjective evaluation.

1. Introduction

Digital images are experiencing a tremendous growth in both theoretical developments and sophisticated applications. The primary issue in digital image processing is the image quality assessment (IQA) [1]. Image quality assessment is the procedure mapping the changes of the images to corresponding visual preference. It can be divided into subjective evaluation and objective evaluation. Obviously, humans, as the ultimate receivers of the images, are the best candidate to assess the quality of images. Nevertheless, the inconvenience and huge consumption of subjective evaluation leads to its less favor. And the objective image quality assessment has attracted increasing attentions from more and more researchers.

According to the dependence of the reference image, the objective image quality can fall into three categories: full-reference (FR), reduced-reference (RR) and no-reference (NR). Full-reference image quality assessment algorithms has the strongest dependence on the reference image. They take both a reference image and a distorted image as input and produce a scalar which measure the quality of the distorted image. Reduced-reference image quality assessment algorithms has less dependence on reference images. Instead of reference images, certain features of reference images are available to evaluate the quality. No-reference image quality assessment (also known as blind image quality assessment) algorithms perform image quality estimation with no reference image information but only the distorted images. In this paper we focus on the full-reference image quality assessment.

Conventional full-reference image quality assessment algorithms calculate the pixel-wise difference between a distorted image and its corresponding reference image, for example, the mean-squared error (MSE) and the peak signal-to-noise ratio (PSNR). Despite its intensive usage, these metrics are not preferable choice because they correlate poorly with human visual subjectivity.

More FR image quality assessment algorithms have been developed these years. As the ultimate objective of IQA is human visual preference, it is taken for granted to combine the characteristics of human visual system (HVS) with the IQA. The easy way to combine with HVS is weighting the frequency domain error with a contrast sensitivity function (CSF), which is called the weighting signal-to-noise ratio (WSNR) [2]. N. Damera Venkata et al. proposed the noise quality measure (NQM) which consists of the contrast pyramid processing and the signal-to-noise ratio computation [3]. Karen Egizarian et al. presented the IQA metric named as PSNR-HVS [4]. It removes the mean shift and then stretches the contrast by a scanning window before the calculation of the modified PSNR. The modification of PSNR is taking HVS into account when in the calculation of the MSE in PSNR. In the next year, these researchers proposed the PSNR-HVS-M metric [5] based on PSNR-HVS. The new metric is based on model of inter-coefficient masking of DCT basis functions and the modifications version of PSNR. Damon M. Chandler and Sheila S. Hemami presented an IQA metric based on near-threshold and suprathreshold properties of human vision called the visual signal-to-noise ratio (VSNR) [6].

HVS-based IQA metrics have a high dependence on the accuracy on the models of HVS. So many

researchers employ other models such as image statistics. Hamid Rahim Sheikh et al. took the IQA problem as an information fidelity problem. They proposed a new metric named as information fidelity criterion (IFC) [7] which measured the statistical information that a distorted image had of the reference image. Similarly, Hamid Rahim Sheikh and Alan C. Bovik presented visual information fidelity (VIF) [8] which was derived from a statistical model for natural scenes, a model for image distortions and a HVS model. In the other direction, Zhou Wang and Qiang Li found a novel weighting approach based on a Gaussian scale mixture (GSM) model. It had pretty well and consistent performance improvement of both PSNR and SSIM based IQA named as information content weighted PSNR (IW-PSNR) and SSIM (IW-SSIM) respectively [9].

Besides HVS and statistics based IQA, there are structure based IQA metrics. Though ‘image structure’ remains non uniform definition, this class of metric is of more preference. Among these metrics, the universal image quality index (UQI) [10] and the structural similarity index (SSIM) [11] are of essential importance. The UQI employs cross-correlation and measures of luminance and contrast differences to estimate quality. The SSIM is actually the extended version of the UQI. The difference is that the SSIM added small constants to the numerator and the denominator of each measure. There are a lot of IQA metrics based on SSIM. Take CW-SSIM [12] as an example, it combines a complex wavelet with the SSIM and it is insensitive to translation, scaling and rotation of images in comparison with SSIM. Except for these SSIM based metrics, there are other structure based IQA. The energy of structural distortion (ESD) [13] is an IQA metric of this kind. Cailing Wang et al. extended ESD to the spectral image in spectrum aspect and spatial aspect [14].

In this paper, we proposed a full-reference image quality assessment metric based on the contourlet transform and the energy of structural distortion (ESD). Section II presents the background knowledge of the contourlet transform and the ESD. Section III presents our proposed IQA metric named as contourlet transform based ESD (CT-ESD). Section IV presents the subjective validation details. Then we conclude the paper in Section V.

2. Theory Basis

It is known to all that 2-D wavelet transform is short of efficiency for image representation in spite of its excellent performance for 1-D signal processing. Therefore, to develop multiscale geometric analysis is necessary. Contourlet transform [15] is an attractive alternative. It is constructed originally in discrete domain, meanwhile, it has a precise connection with continuous domain expansions. Contourlet transform allows for a different number of directions at different scale while achieve nearly critical sampling. Its implement based on iterated filter banks leads to computationally efficient. Moreover, Contourlet

transform can provide a sparse representation for natural images with smooth contours. It possesses the qualities of directionality and anisotropy which are important for image representations. Thus we applied Contourlet transform in IQA metric design.

Contourlet transform is actually a double filter bank structure. The first filter bank is Laplacian pyramid (LP). It is aimed at capture the point discontinuities. Here we focus on the decomposition of LP. At each level, the input image is filtered by lowpass analysis filter H and downsampled by sampling matrix M to generate a lowpass version of the input. At the same time, a bandpass image is obtained by the subtraction between the input at this level and the prediction. The prediction is the lowpass version mentioned above with sequentially upsampling and filtering. The sampling matrix in upsampling is M and the filter is synthesis filter G . The second filter bank is directional filter bank (DFB). This filter bank links point discontinuities into linear structures. It decomposes the bandpass image from LP into 2^l subbands via l -level binary tree. Each subband has a wedge-shaped frequency partitioning. DFB has two building blocks: two-channel quincunx filter bank and shearing operator. The former one divides a 2-D spectrum into horizontal and vertical directions while the latter reorder the image samples. The combination of these two building blocks can leads to the desired 2-D spectrum division. The Fig 1 shows this two steps procedure of the contourlet transform. The contourlet transform has been applied to image quality assessment before, for example, in [16].

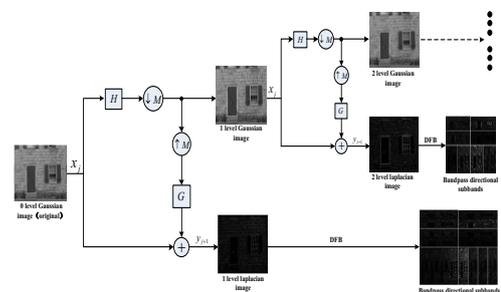


Figure 1. The flow diagram of the contourlet transform. First, the LP is employed to decompose the image into multiscale. Then in each bandpass channel, the DFB divides it into different directional subbands.

After we learn about the contourlet transform, we will continue to introduce the energy of structural distortion (ESD) [13].

The reference and distorted image are divided into blocks with size $H \times L$. The blocks are actually 2D vectors denoted by b_i and b'_i respectively for the reference and distorted image. The element of b_i is defined as $P_i(x, y)$. The structural characterization S_i of b_i is defined as

$$S_i = \frac{b_i}{\sqrt{\sum_{x,y} P_i(x, y)^2}} \quad (1)$$

It is noticeable that the square of S_i is 1, in other words, the energy of S_i is 1.

The inner product is set to be the simplest one. When $V = \{v_{i,j} | i=1, \dots, H, j=1, \dots, L\}$ and $W = \{w_{i,j} | i=1, \dots, H, j=1, \dots, L\}$, the inner product of them is

$$\langle V, W \rangle = \langle W, V \rangle = \sum_{i,j} v_{i,j} * w_{i,j}, i=1, \dots, H, j=1, \dots, L^4$$

Based on the inner product, we obtain the energy of structural information E_i of b_i by the following formula,

$$E_i = \langle b_i, S_i \rangle \quad (2)$$

The energy of structural information E'_i corresponding to b'_i in distorted image is calculated the same way as follows:

$$E'_i = \langle b'_i, S_i \rangle \quad (3)$$

It should be noted that in the calculation of E'_i , the object in the inner product is S_i .

The final score of ESD is defined as

$$ESD = \log \left(\sqrt{\sum_{i=1}^K (E_i - E'_i)^2} \right) \quad (4)$$

in which K is the number of blocks. The lower the score of ESD is, the higher image quality is. The fact that the distorted image is identical to the reference one will result in the value of ESD being negative infinity. So we insist that the calculation of ESD should add a small constant before the logarithm operation.

3. Metric Proposed

The ultimate objective of IQA is to obtain an evaluation of image quality which approximate the perception by human visual system (HVS). The multi-resolution, directionality, and anisotropy are characteristics possessed both by HVS and Contourlet transform. Moreover, the Contourlet transform is sparse and effective for image representation which is fundamental to IQA. Subsequently, it is naturally inclined to employ the Contourlet transform into the IQA metric. We propose an IQA metric combined the Contourlet transform and the energy of structural distortion (CT-ESD). The procedures are list below.

Input the distorted image I_D and the corresponding reference image I_R . These two images are decomposed by the Contourlet transform to get the coefficients $\{c_{j,k}^D\}$ and $\{c_{j,k}^R\}$. The superscript D indicates the distorted image and R indicates the reference image. The subscript j represents the scale index and k means the orientation index. For example, $c_{j,k}^D$ denotes the Contourlet coefficients of the distorted image at the j th level in the k th direction. And the low frequency subband has no direction division.

The Contourlet coefficients $\{c_{j,k}^D\}$ and $\{c_{j,k}^R\}$ are divided into blocks $\{b_{i,j,k}^D\}$ and $\{b_{i,j,k}^R\}$ with the size $H \times L$ (in this paper 8×8). The index i denotes the block index. Then obtain the structural characterization $\{S_{i,j,k}\}$ as the same way as the equation (1).

Calculate the energy of structural information of the distorted image and the reference image which are denoted as $\{E_{i,j,k}^D\}$ and $\{E_{i,j,k}^R\}$ respectively.

$$\{E_{i,j,k}^D\} = \langle b_{i,j,k}^D, S_{i,j,k} \rangle, \{E_{i,j,k}^R\} = \langle b_{i,j,k}^R, S_{i,j,k} \rangle \quad (5)$$

It should be noted that each subband of Contourlet transform generates an energy value matrix.

At each scale j , the variance of the energy is calculated

$$EV_j = \sum_{k=1}^K \sum_{i=1}^N (E_{i,j,k}^R - E_{i,j,k}^D)^2, j=1, \dots, J \quad (6)$$

The sum ensures EV_j is a scalar.

The final metric proposed is defined as

$$ESD = \log \left(\sqrt{\sum_j w_j * EV_j + \alpha} \right) \quad (7)$$

in which α is a constant to avoid the logarithm of zero and small enough to have no influence on the result. w_j is the weighting coefficient of the EV_j at scale j .

Moreover, the sum of all the w_j is one ($\sum_j w_j = 1$).

For different scale, its weighting coefficient can change according to its contribution to the HVS.

The flow chat of the CT-ESD is in Figure 2.

4. Experiments

At the beginning of this section, we need to introduce the image database briefly. According to [1] there are over twenty publicly available image databases. In consideration of the diversity of distortion patterns and the challenge of the database, we choose the image database TID 2013[17]. It is intended for analysis and verification of full-reference image quality assessment metrics. Actually the image database TID 2013 is the successor of the image database TID2008. Compared to TID2008, new types and one new level of distortions have been added to TID2013. In general, TID2013 contains 25 reference color images. It involves 24 types of image distortion in which 7 types are new. And 5 levels of distortion are in the database approximately corresponding to the PSNR equal to 21,24,27,30, and 33dB respectively. In [17] the distorted images of TID2013 can be divided into six subsets: Noise, Actual, Simple, Exotic, New, and Color. These subsets depend upon different applications. For example, the subset 'Noise' excludes the images distorted by the type of noise in conventional image processing. Other details about the subsets and TID2013 can refer to [17]

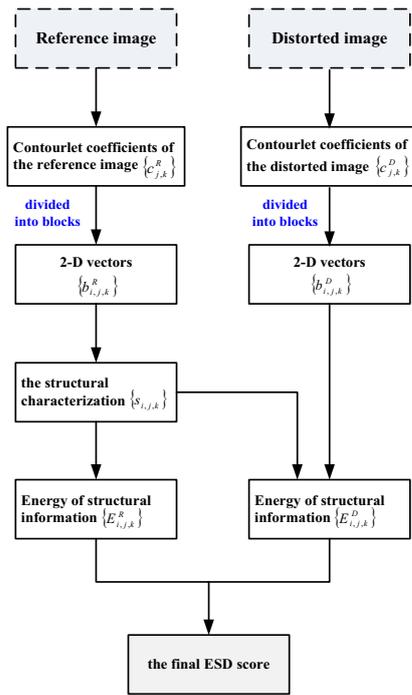


Figure 2. The flow chart of the IQA metric combined the Contourlet transform and the energy of structural distortion (CT-ESD).

In this paper, we choose the Spearman rank-order correlation coefficient (SROCC) and the Kendall rank-order correlation coefficient (KROCC) just as the authors of TID2013 do. They prefer to employ rank order correlation coefficients to avoid fitting procedures which may be not unique.

The value of the SROCC and the KROCC between the mean opinion score (MOS) and the proposed metric score is expected to be high which implies the new metric is a good substitution for HVS.

As there are there are parameters $w_j, j = 1, 2, \dots, J$ in our proposed metric, we choose the w_j which can lead to higher SROCC and KROCC. Experiments prove that high parameter corresponding to medium frequency subband can lead to high rank-order correlation coefficients. This phenomenon is essentially in agreement with the characteristic of HVS. In this paper we choose 3 level Laplacian decomposition in the contourlet. It produces coefficients of four cells corresponding to one low frequency subband and three pyramidal levels. And each pyramidal level contains 2^3 frequency bands. Here, we set the parameters as follows: $w_1 = 0.01$, $w_2 = 0.93$, $w_3 = 0.03$, $w_4 = 0.03$. They were assigned to the low frequency subband and three pyramidal levels. Obviously, w_2 is the biggest one which means the weighting coefficient of the medium frequency subband is the biggest.

The performance of the proposed CT-ESD metric will be validated and compared with some representative IQA metrics of the HVS based, statistics based and structural based IQA metrics respectively. These metrics are NQM [3], PSNR-HVS [4], VSNR[6], UQI[10], SSIM[11], IFC[7], VIF[8], IW-PSNR[9], and ESD [13].

And the comparisons to the conventional IQA such as PSNR and MSE are meaningful and necessary. The data is provided in [17] for most of these metrics. And it also can be calculated by Metrix MUX Visual Quality Assessment Package [18].

The following two tables show the SROCC and KROCC between the metrics and the MOS respectively for each subset. In the tables, each column represents a subset of the database TID2013 except for the last column. The last column 'Full' means the whole set of the database TID2013. Each row represents a metric. For example, in the Table 1, the value at the row 'CT-ESD' and the column 'Full' is 0.7250. It indicates that the metric CT-ESD gets 0.7250 for the SROCC with the MOS on the whole set. As a result of the challenging difficulty of the database TID2013, the values of the SROCC and KROCC have not appear the high values very close to 1.

From the tables, we can conclude that CT-ESD outperforms the other metrics in the table on the whole distorted images set. Though the CT-ESD fails to outperform the others on each subset, its performance is more evenly in different subset. Especially, the 'Exotic' subset contains distortions types which are not happen frequently but among the most difficult ones for IQA. In this subset the CT-ESD has the highest KROCC value 0.5021 of that column in the Table 2. In Table 1, the CT-ESD is 0.6894, only lower than the VSNR 0.7064 on this subset. On the 'New' subset and the 'Color' subset, which drag down the rank-order correlation coefficient for most metrics, the CT-ESD has preferable performance. It gets 0.6776 and 0.6021 for the SROCC which are both ahead in the corresponding column.

It is noteworthy that compared to ESD, CT-ESD is better on the subsets 'Noise', 'Actual', and 'Exotic'. These three subset relate to distortion types most common and difficult respectively. And CT-ESD has a significant improvement on the subset 'Exotic'. Together these subsets, CT-ESD outperforms the ESD in the full database TID2013. Whether the SROCC or the KROCC, it wins the other metrics at the 'Full' column with the highest values 0.7250 and 0.5514 respectively.

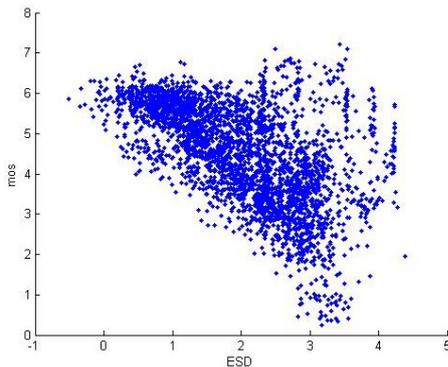
In Figure 3 we can see the scatter plot between the MOS and the metrics. From Figure 3(a) and (b), it is easy to conclude that the scatters in (b) are more compact and have more apparent tendency. Figure 3(b) represents the scatter plot between the MOS and the CT-ESD. So this means the CT-ESD has better evaluation of the image quality. Figure 3(c) and (d) show that different metrics may lead to the scatter plot into different shapes. From them we can know that the CT-ESD has better performance as it performs better on the compactness and tendency.

Table 1 the SROCC values of the metrics for the database TID2013

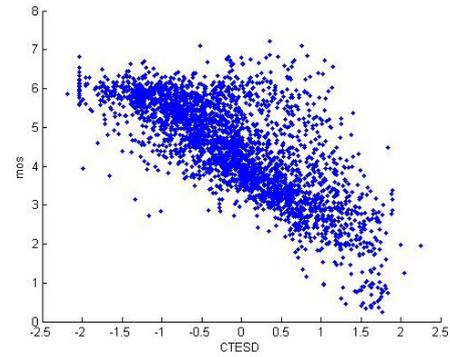
Metric	Noise	Actual	Simple	Exotic	New	Color	Full
PSNR	0.8217	0.8246	0.9134	0.5968	0.6190	0.5387	0.6395
MSE	0.7855	0.8048	0.8785	0.5468	0.7210	0.6696	0.6724
NQM	0.8362	0.8572	0.8752	0.5891	0.6258	0.5418	0.6349
PSNR-HVS	0.9172	0.9257	0.9507	0.6006	0.6471	0.5589	0.6536
VSNR	0.8691	0.8817	0.9121	0.7064	0.5888	0.5122	0.6809
UQI	0.6482	0.6904	0.7575	0.5313	0.4935	0.4440	0.5444
SSIM	0.7574	0.7877	0.8371	0.6320	0.5801	0.5057	0.6370
IFC	0.7218	0.7608	0.7792	0.3610	0.5444	0.4449	0.5400
VIF	0.8420	0.8589	0.9321	0.6282	0.5930	0.5210	0.6770
IW-PSNR	0.8961	0.9097	0.9237	0.6510	0.6470	0.5533	0.6888
ESD	0.7402	0.7893	0.8796	0.3438	0.7813	0.7719	0.5753
CT-ESD	0.8681	0.8698	0.8661	0.6894	0.6776	0.6021	0.7250

Table 2 the KROCC values of the metrics for the database TID2013

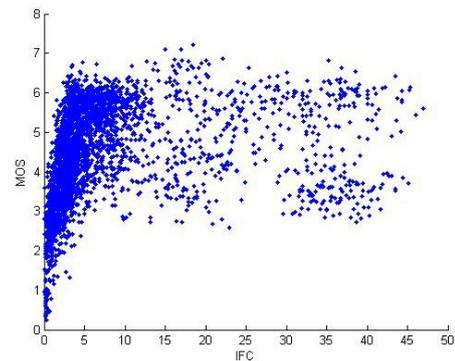
Metric	Noise	Actual	Simple	Exotic	New	Color	Full
PSNR	0.6236	0.6242	0.7452	0.4254	0.4728	0.4156	0.4700
MSE	0.5793	0.6007	0.6984	0.3822	0.5319	0.4874	0.4870
NQM	0.6413	0.6665	0.6812	0.4120	0.4831	0.4087	0.4662
PSNR-HVS	0.7547	0.7661	0.8092	0.4356	0.5169	0.4486	0.5077
VSNR	0.6761	0.6908	0.7312	0.5193	0.4374	0.3788	0.5077
UQI	0.4601	0.4976	0.5499	0.3776	0.3529	0.3154	0.3906
SSIM	0.5515	0.5768	0.6286	0.4548	0.4226	0.3823	0.4636
IFC	0.5273	0.5630	0.5740	0.2579	0.3982	0.3209	0.3959
VIF	0.6590	0.6729	0.7694	0.4634	0.4474	0.3998	0.5148
IW-PSNR	0.7240	0.7465	0.7705	0.4606	0.5185	0.4349	0.5250
ESD	0.5409	0.5935	0.7081	0.2364	0.5856	0.5673	0.4175
CT-ESD	0.6899	0.6916	0.6748	0.5021	0.5194	0.4509	0.5514



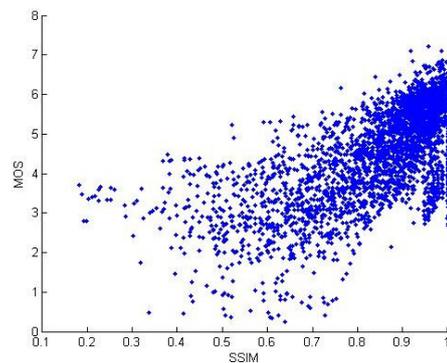
(a)



(b)



(c)



(d)

Figure 3. The scatter plot between the MOS and some metrics. (a) the plot between the MOS and the ESD;(b) the plot between the MOS and the CT-ESD;(c) the plot between the MOS and the IFC;(d) the plot between the MOS and the SSIM.

The data and the charts above show that it is meaningful to introduce the contourlet transform into the IQA metric. We attribute the preferable performance of the CT-ESD to the attendance of the contourlet transform. The multi-resolution and the sparse representation of the contourlet transform make the IQA metric more effective and more similarity to the HVS. These advantages lead to the success of the IQA.

5. Discussions

In this paper, we proposed a novel full-reference IQA combined the ESD with the contourlet transform named

as CT-ESD. It is a structural based IQA metric which measures the loss of the structure information to evaluate the image quality. The attendance of the contourlet lead to the improvement of the ESD image quality assessment. Moreover, its preferable performance on the database TID2013 proves its validity and practicability, especially for the 'Exotic' type distortions. Our proposed metric still has the potential of improvement. We will focus on the utilization of the multi-direction of the contourlet in the IQA and the IQA metric aimed at the 'Color' type distortion. We insist that the contourlet transform as a 'real' two dimensional transform will be a powerful tool in the IQA.

References

- [1] Dammon M.Chandler, Seven Challenges in Image Quality Assessment: Past, Present, and Future Research, ISRN Signal Processing, volume 2013
- [2] T.Mitsa, K.L.Varkur, Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms, in IEEE International Conference on Acoustics, Speech, and Signal Processing(ICASSP), Minneapolis,USA,1993,pp.301-304
- [3] Niranjan Damera-Venka, Thomas D.Kite, Wilson S.Geisler, Brian L.Evans, and Alan C.Bovik, Image Quality Assessment Based on a Degradation Model, IEEE Transactions on Image Processing, Vol.9, No.4, 2000
- [4] K.Egiazarian, J.Astola, N.Ponomarenko, V.Lukin, F.Battisti, M.Carli, New full-reference quality metrics based on HVS, in Second International Workshop on Video Processing and Quality Metrics, Scottsdale, USA, 2006,pp.1-4
- [5] N.Ponomarenko, F.Silvestri, K.Egiazarian, M.Carli,J.Astola,V.Lukin, On between-coefficient contrast masking of DCT basis functions, in Third International Workshop on Video Processing and Quality metrics, Scottsdale, USA,2007,pp.1-4
- [6] Damon M.Chandler, Sheila S.Hemami, VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images, IEEE Transactions on Image Processing, Vol.16, No.9, 2007
- [7] Hamid Rahim Sheikh, Alan Conrad Bovik, Gustavo de Veciana, An Information Fidelity Criterion for Image Quality Assessment Using Natural Scene Statistics, in IEEE Transaction on Image Processing, Vol.14,No.12,2005
- [8] Hamid Rahim Sheikh and Alan C.Bovik, Image information and visual quality, in IEEE Transactions on Image Processing, Vol.15,No.2,2006
- [9] Zhou Wang, Qiang Li, Information Content Weighting for Perceptual Image Quality Assessment, in IEEE Transactions on Image Processing , Vol.20, No.5, 2011
- [10]Zhou Wang, Alan C.Bovik, A Universal Image Quality Index, in IEEE Signal Processing Letters, Vol.9, No.3, 2002
- [11]Zhou Wang, Alan C.Bovik, Hamid Rahim Sheikh, Eero P.Simoncelli, Image Quality Assessment: From Error Visibility to Structural Similarity, in IEEE Transaction on Image Processing, Vol.13, No.4,2004
- [12]Zhou Wang, E.P.Simoncelli, Translation insensitive image similarity in complex wavelet domain, in IEEE International Conference on Acoustics, Speech, and Signal Processing(ICASSP'05), Philadelphia, USA,2005,pp.573-576
- [13]Jianxin Pang, Rong Zhang, Lu Lu, Zhengkai Liu, Image Quality Assessment Based on Energy of Structural Distortion, Advances in Multimedia Information Processing-PCM 2007,pp785-793
- [14]Cailing Wang, Xuebin Liu, Bingliang Hu, A New Spectral Image Assessment Based on Energy of Structural Distortion, in 2009 International Conference on Image Analysis and Signal Processing(IASP),2009
- [15]Minh N.Do and Martin Vetterli, The Contourlet Transform: An Efficient Directional multiresolution Image Representation, in IEEE Transactions on Image Processing, Vol.14,No.12,2005
- [16]Bin Wang, Image Quality Assessment Method Based on Contourlet Transform and Generalized Gamma Distribution, in 3rd International Conference on Material, Mechanical and Manufacturing Engineering (IC3ME2015), 2015
- [17]Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, C.-C. Jay Kuo, Image database TID2013: Peculiarities, results, and perspectives, in Signal Processing: Image Communication 30(2015),pp.57-77, 2015
- [18]M.Gaubatz, Metrix MUX Visual Quality Assessment Package, in http://foulard.ece.cornell.edu/gaubatz/metrix_mux/, 27 March 2014.