THE DIVISIVE NORMALIZATION TRANSFORM BASED REDUCED-REFERENCE IMAGE QUALITY ASSESSMENT IN THE SHEARLET DOMAIN

Wu Dong[a,b], Hongxia Bie[a], Likun Lu[b,1], and Yeli Li[b]

[a]School of Information and Communication Engineering, Beijing University of Posts and Telecommunication, Beijing, China
[b]Beijing Key Laboratory of Signal and Information Processing for High-end Printing Equipment, Beijing Institute of Graphic Communication, Beijing, China

ABSTRACT

Reduced-reference (RR) image quality assessment (IQA) metric aims to employ less partial information about the original reference image to achieve higher evaluation accuracy. In this paper, we propose a novel RRIQA metric based on the divisive normalization transform (DNT) in the discrete nonseparable shearlet transform (DNST) domain. In this metric, the coefficients in the DNST domain are normalized employing the Gaussian scale mixture statistical model, and then the marginal distribution of the coefficients changes into approximate Gaussian distribution. A set of statistical features is extracted from DNT-domain representations of the reference and distorted images, respectively. The weighting of these features is performed based on the characteristics of the human visual system. Structural similarity comparison of these features is conducted as an objective quality score of the distorted image. The proposed metric is evaluated on the public LIVE database and demonstrates fairly good performance across a wide range of image distortions.

Index Terms— reduced-reference image quality assessment, discrete nonseparable shearlet transform, divisive normalization transform, human visual system, structural similarity

1. INTRODUCTION

Over the past years, there has been a vast proliferation in great demand of image and video services. The multimedia content is delivered over both wireline and wireless networks and suffers from various kinds of distortions on its way to the destination at the same time. So there is considerable concern regarding how the quality of service of image and video being delivered can be managed [1]. A large number of image quality assessment (IQA) metrics have been designed to identify and quantify the quality of a distorted image. Full-reference image quality assessment (FRIQA) is not useful since the reference images are not accessible at the receiver side in most present and emerging practical visual communication environments. Reduced-reference (RR) IQA metrics are designed by employing a limited number features of the reference image to predict the quality degradation of the distorted image. The features extracted from the reference image should represent the reference image efficiently and be sensitive to a variety of distortions. Wang proposed the wavelet-domain natural image statistic metric (WNISM) [2, 3], in which a generalized Gaussian model is employed to summarize the marginal distribution of wavelet subband coefficients of the reference image and the parameters of the fitting model are used as RR features. Gao proposed an RR-IQA framework based on multiscale geometric analysis, which employs a series of transforms including wavelet, curvelet, contourlet, etc [4]. But the strong dependencies between the neighboring transform coefficients are ignored in [2-4]. Li proposed an RRIQA metric based on a divisive normalization image representation in the wavelet domain [5]. Yet, the wavelet transform only is very efficient in dealing with pointwise singularities and cannot perform as well to two dimensional data of the image.

In this paper, we propose a novel RRIQA metric based on the divisive normalization transform (DNT) in the discrete nonseparable shearlet transform (DNST) domain. The DNST forms a tight frame of well-localized waveforms, at various scales and orientations, and is optimally sparse in representing images [6]. The DNT has been recognized as an effective mechanism to model the perceptual sensitivity in biological perceptual system and also provides a useful model to describe the psychophysical visual masking effect, which refers to the reduction of the visibility of an image component in the presence of neighboring components. The DNT has been applied in 3D reduced-reference image quality assessment in the contourlet domain [7]. In addition, the contrast sensitivity function (CSF) and the oblique effect of the human visual system (HVS) also are taken into account in this paper.

[1]Corresponding author (*Likun Lu, Email:* lulikun.bigc@hotmail.com). This work is supported by the Science and Technology Project of Beijing Education Committee under grant no. SQKM201410015007, and the Green Printing and Publication Technology of Cooperation Innovation Center under grant no. PXM2016_014223_000025.

2. PROPOSED RRIQA METRIC

2.1. The discrete nonseparable shearlet transform

The DNST is associated with compactly supported shearlets generated by a nonseparable generator which provides good localization in space domain as well as improved directional selectivity [6]. The DNST is an optimal transform with sparse coefficients and has no limitation on the number of directions for shearing in comparison to the wavelet transform and the contourlet transform.

In the DNST, the classical theory of affine systems is employed. In dimension $n = 2$, the affine systems with composite dilations are the collections of the form as follows:

$$\mathcal{A}_{AB}(\psi) = \{\psi_{j,l,k}(x) = |detA|^{\frac{j}{2}}\psi(B^l A^j x - k): j, l \in \mathbb{z}, k \in \mathbb{z}^2\} \quad (1)$$

Where $\psi \in L^2(R^2)$, A, B are 2×2 invertible matrices and $|det\ B|=1$. The dilations matrices A^j are associated with scale transformations, and the matrices B^l are associated to area-preserving geometrical transformations, such as rotations and shear.

$A = A_0$ is the anisotropic dilation matrix and $B = B_0$ is the shear matrix, which are given by

$$A_0 = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix},\ B_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad (2)$$

The analyzing function of the shearlet transform is anisotropic and is defined at different scales, locations and orientations. The DNST is well localized, highly directional sensitivity and has optimally sparse representation to images. With these good properties, the DNST can provide more additional information about the distorted image than the traditional wavelet transform and the contourlet transform. So, the DNST is suitable for the RRIQA work.

2.2. Image statistics in divisive normalization representation

A DNT is built upon a linear transform model. The DNST provides a convenient framework for localized representation of images simultaneously in space, scale (frequency) and orientation. So, in this paper, the DNST is employed as the linear transformation. After the DNST is applied, one image is decomposed into a set of subbands at different scales and different orientations. Let c represents a DNST coefficient in a subband, and then a normalized coefficient $\tilde{c}$ is computed as $\tilde{c} = c/p$, where p is a positive divisive normalization factor which is derived through a local statistical image model. In this paper, the Gaussian scale mixtures (GSM) model of the DNST coefficients is employed as the local statistical model.

A length-N random vector Y is a GSM if it can be expressed as $Y = zU$, where U is zero-mean Gaussian random vector with covariance C_U, and z is a scalar random variable called a mixing multiplier [5]. Suppose that the mixing density is $p_z(z)$, then the density $p_Y(Y)$ of Y can be shown as follows:

$$p_Y(Y) = \int \frac{1}{[2\pi]^{\frac{N}{2}}|z^2 C_U|^{1/2}} \exp(-\frac{Y^T C_U^{-1} Y}{2z^2}) p_z(z) dz \quad (3)$$

To simplify the model, z only takes a fixed value at each location and varies over space of the subband. Thus Y is simply a zero-mean Gaussian vector with covariance $z^2 C_U$ and the normalization factor p in the DNT becomes as an estimate $\tilde{z}$ of the multiplier z from the neighboring coefficient vector Y. $\tilde{z}$ is obtained by a maximum-likelihood estimation as follows:

$$\tilde{z} = arg\ max_z\{logp(Y|z)\} = \sqrt{Y^T C_U^{-1} Y/N} \quad (4)$$

where the covariance matrix $C_U = \mathrm{E}[UU^T]$ is estimated from the entire DNST subband. In this paper, the vector Y of each DNST coefficient is composed of the adjacent coefficients in the same subband, the coefficient in the same spatial location of the coarse scale, and the coefficients in the same scale and spatial location but different orientation.

In Fig.1, the "womanhat" image that comes from the LIVE database [8] is used. In Fig.1(c), the distribution of the subband coefficients of the original natural image in the DNST domain is shown. This distribution has sharp peaks at zero amplitude and heavy tails on both sides of the peak. The kurtosis of Fig.1(c) is much higher than 3 which is the kurtosis of the Gaussian distribution. In Fig.1(e), the DNST coefficients distribution is compared with the Gaussian distribution with the same variance. Obviously, the marginal distribution of the DNST subband is highly non-Gaussian. In Fig.1(d), the kurtosis of the distribution of the same DNST coefficients after the DNT is very close to 3. By contrast in Fig.1(f), the distribution of the DNST coefficients after the DNT can be very fitted with a Gaussian with small error.

To different distortion types of the image [8], the statistics of the DNST coefficients after the DNT are investigated, which are provided in Fig.2. It is observed that the distortions can dramatically change the original near-Gaussian distribution. In addition, different distortion types have different effect.

2.3. The contrast sensitivity function

The HVS has different sensitivity to different spatial frequencies and orientations. Here we apply a model of the contrast sensitivity function (CSF) $H(f, \theta)$ [9, 10], which is shown as follows:

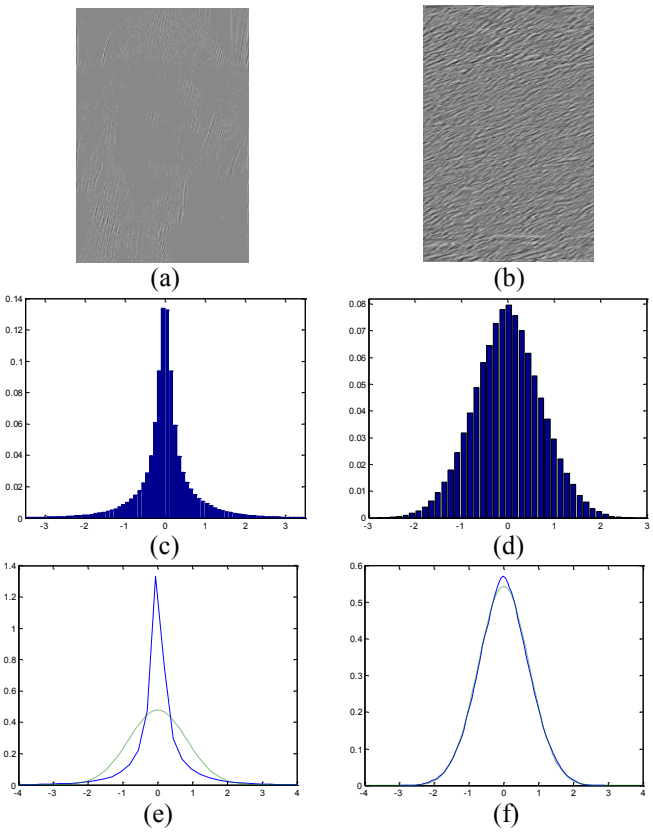

Fig.1. (a) original DNST subband coefficients; (b) the DNST subband coefficients after the DNT; (c) marginal statistics of Fig.1(a) and its kurtosis is 22.52; (d) marginal statistics of Fig.1(b) and its kurtosis is 3.01; (e) the histogram of Fig.1(a) (solid curve) and the fitting Gaussian curve with the same variance (dashed curve); (f) the histogram of Fig.1(b) (solid curve) and the fitting Gaussian curve with the same variance (dashed curve)

$$H(f,\theta)=\begin{cases}2.6(0.0192+\lambda f_\theta)\exp[-\lambda f_\theta] & f \geq f_{peak} \\ 0.981 & \text{otherwise}\end{cases} \quad (5)$$

$$f_\theta = f/[0.15\,cos(4\theta)+0.85] \quad (6)$$

where f denotes the radial spatial frequency, and θ denotes the orientation, and f_θ accounts for the oblique effect.

2.4. Feature extraction and quality pooling

In this paper, firstly the DNST is utilized to decompose the reference image at the sender side and a set of subbands at different scales and different orientations is obtained. Then the DNT representation for the coefficients of each subband is computed. To the subband at the *m*th scale and the *n*th orientation, the probability density function (PDF) of the DNT coefficients can be well approximated with a zero-mean Gaussian function $p^g_{mn}(\mathrm{x})$ with the standard deviation σ_{mn}.

The fitting error between the true PDF $p_{mn}(\mathrm{x})$ of the coefficients of the subband at the *m*th scale and the *n*th orientation and the Gaussian model is accounted for employing the Kullback-Leibler distance (KLD), which is given by:

$$d(p^g_{mn}||p_{mn}) = \int p^g_{mn}(x) log \frac{p^g_{mn}(x)}{p_{mn}(x)} dx \quad (7)$$

At the receiver side, the same DNT is applied to a distorted image. The distortion of the subband at the *m*th scale and the *n*th orientation is estimated by employing the KLD between the PDF $p_{mn}(x)$ of the original image and the PDF $q_{mn}(x)$ of the distorted image:

$$\hat{d}(p_{mn}||q_{mn}) = d(p^g_{mn}||q_{mn}) - d(p^g_{mn}||p_{mn}) \quad (8)$$

where $d(p^g_{mn}||q_{mn})$ is the KLD between $p^g_{mn}(x)$ and $q_{mn}(x)$.

The structure between the distortions across different subbands should be taken into account. Such distortion structure is a critical issue behind the philosophy of the structure similarity (SSIM) approach [11, 12]. Here, we apply the SSIM philosophy into the measurement of statistical features extracted from the DNT coefficients at different subbands, which is given as follows:

$$g'_\sigma(\sigma'_r,\sigma'_d) = \frac{||\sigma'_r||^2+||\sigma'_d||^2+c_1}{2(\sigma'_r\cdot\sigma'_d)+c_1} \quad (9)$$

$$g'_s(s'_r,s'_d) = \frac{||s'_r||^2+||s'_d||^2+c_2}{2(s'_r\cdot s'_d)+c_2} \quad (10)$$

$$g'_k(k'_r,k'_d) = \frac{||k'_r||^2+||k'_d||^2+c_3}{2(k'_r\cdot k'_d)+c_3} \quad (11)$$

where σ'_r and σ'_d represent the vectors containing the weighted standard deviations of the DNT coefficients from each subband in the reference and distorted images, respectively. Similarly, s'_r, s'_d represent vectors containing the weighted sknewnesses, and k'_r, k'_d represent the vectors containing the weighted kurtosises. $\sigma'_r\cdot\sigma'_d$, $s'_r\cdot s'_d$ and $k'_r\cdot k'_d$ represent the dot product between the two corresponding vectors. c_1, c_2 and c_3 denote the constants to avoid instability.

The CSF $H(f,\theta)$ is employed to weight the standard deviation σ_{mn}, the sknewness s_{mn} and the kurtosis k_{mn} of the subband at the *m*th scale and the *n*th orientation. Here, in $H(f,\theta)$, f denotes the scale m and θ denotes the orientation n.

Finally, the weighted sum of all subbands distortions is computed as the overall evaluation as follow:

$$D = g'_\sigma(\sigma'_r,\sigma'_d)g'_s(s'_r,s'_d)g'_k(k'_r,k'_d)\log(1+\frac{1}{D_0}\sum_m\sum_n H(f,\theta)|_{f=m,\theta=n}\cdot|\hat{d}(p_{mn}||q_{mn})|) \quad (12)$$

where D_0 is a constant to control the scale of the distortion measure.

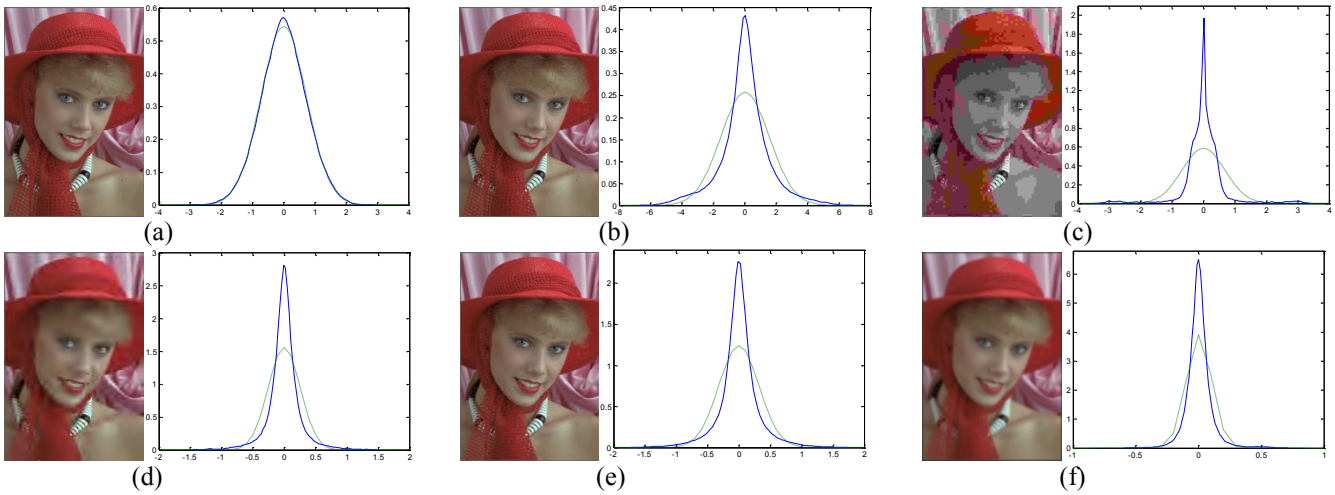

Fig.2. Histograms of the coefficients in the DNT domain under different distortion types. The solid and dashed curves are histograms of the coefficients in the DNT domain of the original and distorted images, respectively. In each figure, the dashed curve has the same variance as the solid curve. (a) original image; (b) white noise; (c) JPEG compression; (d) JPEG2000 compression; (e) fast fading channel distortion; (f) Gaussian blur.

3. IMPLEMENTATION

In this paper, the five-level DNST is applied to decompose the image into subbands with 16, 16, 16, 8, 8 directions from finer to coarser scales. To each subband, the DNT is applied using 17 neighboring coefficients, including 9 from the same subband, 1 from the subband at the coarse scale and 7 from the same spatial location in the other orientation subbands at the same scale. Four features $d(p_{mn}^{g}||p_{mn})$, σ_{mn}, s_{mn} and k_{mn} are extracted from each subband of the reference image and are sent to the receiver side for the quality evaluation of the distorted image.

The Live database [8] is used to test the proposed metric. This database contains 982 distorted images with 5 distortion types including JPEG compression, JPEG2000 compression (JP2K), white noise (WN), Gaussian blur (GB) and fast fading channel distortion (FF). Three criteria are employed to evaluate the performance of objective metrics [13, 14]: correlation coefficient after a nonlinear mapping, Spearman rank-order correlation coefficient and outlier ratio. The results are shown in Table 1.

Table 1: Performance evaluation result of objective metrics

	JP2K	JPEG	WN	GB	FF	All
	Correlation Coefficient					
PSNR	0.8928	0.9683	0.9578	0.7721	0.8847	0.8723
SSIM	0.9661	0.9776	0.9597	0.9434	0.9540	0.9446
WNISM	0.9323	0.9273	0.8991	0.8762	0.9273	0.8321
RRED	0.9501	0.9691	0.9712	0.9419	0.7734	0.9108
Proposed	0.9587	0.9814	0.9497	0.9512	0.9497	0.9391
	Spearman Rank-order Correlation Coefficient					
PSNR	0.8954	0.8809	0.9854	0.7823	0.8907	0.8756
SSIM	0.9320	0.9561	0.9491	0.8678	0.9431	0.9471
WNISM	0.9384	0.8620	0.8639	0.9145	0.9162	0.8437
RRED	0.9432	0.9272	0.9371	0.9571	0.8523	0.9149
Proposed	0.9345	0.9645	0.9581	0.9620	0.9397	0.9507
	Root Mean Square Error					
PSNR	11.882	17.009	9.221	13.430	13.408	13.360
SSIM	6.726	6.692	6.928	5.901	8.732	9.211
WNISM	8.012	8.132	7.350	7.191	6.402	12.912
RRED	9.280	17.740	10.581	5.623	11.034	17.618
Proposed	7.354	8.054	6.856	6.145	10.548	10.125

In Table 1, we compare the proposed metric with state-of-the-art RRIQA metrics, namely WNISM, RRED [15]. In addition, two FRIQA metrics, namely SSIM [12] and PSNR, are also been compared. The results show that the proposed metric significantly outperforms PSNR and is competitive as compared to SSIM. Meanwhile, it outperforms the already existing RRIQA metrics: WNISM, RRED.

4. CONCLUSION

In this paper, we propose a novel RRIQA metric employing the DNT based image statistical properties in the shearlet domain and the SSIM philosophy. Experimental verifications suggest that the proposed RRIQA metric exhibits good correlations with subjective evaluations of image quality over a wide variety of distortions. The proposed metric has the good potential to be employed for a wide range of applications.

5. REFERENCES

[1] Z. Wang and A. C. Bovik, "Reduced- and no-reference image quality assessment," *IEEE Signal Process Mag*, vol. 28, no. 6, pp. 29-40, 2011

[2] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E. Yang and A. C. Bovik, "Quality-aware images," *IEEE Trans. Image Process*, vol. 15, no. 6, pp. 1680-1689, 2006

[3] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," *Proc. SPIE*, vol. 5666, pp. 149-159, 2005

[4] X. Gao, W. Lu, D. Tao and X. Li, "Image quality assessment based on multiscale geometric analysis," *IEEE Trans. Image Process,* vol. 18, no. 7, pp. 1409-1423, 2009

[5] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE J. Sel. Top. Sign. Proces*, vol. 3, no. 2, pp. 202-211, 2009

[6] W. Lim, "Nonseparable shearlet transform," *IEEE Trans. Image Process*, vol. 22, no. 5, pp. 2056-2065, 2015

[7] X. Wang, Q. Liu, R. Wang and Z. Chen, "Natural image statistics based 3D reduced reference image quality assessment in contourlet domain," *Neurocomputing*, vol. 151, no. P2, pp. 683-691, 2015

[8] H. R. Sheikh, Z. Wang, A. C. Bovik and L. K. Cormack, "Image and video quality assessment research at LIVE", http://live. ece. utexax.edu/research/quality

[9] S. A. Golestaneh and L. J. Karam, "Reduced-reference quality assessment based on the entropy of DNT coefficients of locally weighted gradients," *ICIP*'15, pp. 4118-4120, 2015

[10] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *J. Electron. Imaging*, vol. 19, no. 1, pp. 0110061-011006121, 2010

[11] A. Rehman and Z. Wang, "Reduced-reference image quality assessment by structural similarity estimation," *IEEE Trans. Image Process,* vol. 21, no. 8, pp. 3378-3389, 2012

[12] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process,* vol. 13, no. 4, pp. 600-612, 2004

[13] H. R. Sheikh, M. F. Sabir and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process*, vol. 15, no. 11, pp. 3440-3451, 2006

[14] P. Corriveau et al., "Video quality expert group: current results and future directions," *Proc. SPIE Int Soc Opt Eng*, vol. 4067, pp. 742-753, 2000

[15] R. Soundararajan and A. C. Bovik, "RRED indices: reduced reference entropic differencing for image quality assessment," *IEEE Trans. Image Process*, vol. 21, no. 2, pp. 517–526, 2012