SIViP (2013) 7:403–410
DOI 10.1007/s11760-013-0443-4

ORIGINAL PAPER

Image distortion analysis based on normalized perceptual information distance

Nima Nikvand · Zhou Wang

Received: 16 November 2011 / Revised: 14 May 2012 / Accepted: 28 August 2012 / Published online: 16 March 2013

Abstract Image distortion analysis is a fundamental issue in many image processing problems, including compression, restoration, recognition, classification, and retrieval. Traditional image distortion evaluation approaches tend to be heuristic and are often limited to specific application environment. In this work, we investigate the problem of image distortion measurement based on the theory of Kolmogorov complexity, which has rarely been studied in the context of image processing. This work is motivated by the normalized information distance (NID) measure that has been shown to be a valid and universal distance metric applicable to similarity measurement of any two objects (Li et al. in IEEE Trans Inf Theory 50:3250–3264, 2004). Similar to Kolmogorov complexity, NID is non-computable. A useful practical solution is to approximate it using normalized compression distance (NCD) (Li et al. in IEEE Trans Inf Theory 50:3250–3264, 2004), which has led to impressive results in many applications such as construction of phylogeny trees using DNA sequences (Li et al. in IEEE Trans Inf Theory 50:3250–3264, 2004). In our earlier work, we showed that direct use of NCD on image processing problems is difficult and proposed a normalized conditional compression distance (NCCD) measure (Nikvand and Wang, 2010), which has significantly wider applicability than existing image similarity/distortion measures. To assess the distortions between two images, we first transform them into the wavelet transform domain. Assuming stationarity and good decorrelation of wavelet coefficients beyond local regions and across wavelet subbands, the Kolmogorov complexity may be approximated using Shannon entropy (Cover et al. in Elements of information theory. Wiley-Interscience, New York, 1991). Inspired by Sheikh and Bovik (IEEE Trans Image Process 15(2):430–444, 2006), we adopt a Gaussian scale mixture model for clusters of neighboring wavelet coefficients and a Gaussian channel model for the noise distortions in the human visual system. Combining these assumptions with the NID framework, we derive a novel normalized perceptual information distance measure, where maximal likelihood estimation and least square regression are employed for parameter fitting. We validate the proposed distortion measure using three large-scale, publicly available, and subject-rated image databases, which include a wide range of practical image distortion types and levels. Our results demonstrate the good prediction power of the proposed method for perceptual image distortions.

Keywords Kolmogorov complexity · Normalized information distance · image quality assessment · Perceptual information distance

N. Nikvand (✉) · Z. Wang
Department of ECE, University of Waterloo,
200 University Ave. W., Waterloo, ON N2L 3G1, Canada
e-mail: nnikvand@uwaterloo.ca

Z. Wang
e-mail: zhouwang@ieee.org

1 Introduction

One of the most successful recent developments in the theory of Kolmogorov complexity is the normalized information distance (NID) measure, which has been shown to be a valid and universal distance metric applicable to the similarity measurement of any two objects [1]. Similar to Kolmogorov complexity, the NID is non-computable and a practical solution is to approximate it using normalized compression distance (NCD) [1,2], which has led to impressive results in many applications such as construction of phylogeny trees using DNA sequences [1]. However, NCD did not achieve the same level of success in image similarity applications

[3,4]. A framework of normalized conditional compression distance (NCCD) was proposed in [3], which shows significantly wider applicability than existing image similarity/distortion measures. Kolmogorov complexity of an object may also be approximated using Shannon entropy, given that the object is from an ergodic stationary source [5]. The difficulty is that data that arise in practice in the form of images or complex video are generally non-stationary, and thus it is cumbersome to replace Kolmogorov complexity with Shannon entropy without any advanced transformation and modeling.

In this paper, we propose a framework which takes the reference image and the distorted image into the wavelet domain and assumes local independence among image subbands to approximate Kolmogorov complexity by Shannon's entropy. Inspired by [6], wavelet-domain Gaussian scale mixture (GSM) model is adopted for natural scene statistics (NSS), leading to a practical algorithm for entropy calculations.

2 Kolmogorov complexity-based information distances

The Kolmogorov complexity [7] of an object is defined to be the length of the shortest program that can produce that object on a universal Turing machine and halt:

$$K(x) = \min_{p:U(p)=x} l(p)\,. \tag{1}$$

In [1], the authors assume the existence of a general decompressor that can be used to decompress the presumably shortest program x^* to the desired object x. However, they note that due to the non-computability of this concept, a compressor that does the opposite does not have to exist.

The conditional Kolmogorov complexity of x relative to y is denoted by $K(x|y)$. An information distance between x and y can then be defined as:

$$\mathrm{ID}(x, y) = \max\{K(x|y), K(y|x)\} \tag{2}$$

which is the maximum of the length of the shortest program that computes x from y and y from x. To convert it to a normalized symmetric metric, a novel NID measure was introduced in [1]:

$$\mathrm{NID}(x, y) = \frac{\max\{K(x|y^*), K(y|x^*)\}}{\max\{K(x), K(y)\}}\,. \tag{3}$$

It was proved that NID is a valid distance metric that satisfies the identity and symmetry axioms and the triangular inequality [1]. The real-world application of NID is difficult because Kolmogorov complexity is a non-computable quantity [7]. By using the fact that $K(xy) = K(y|x^*) + K(x) = K(x|y^*) + K(y)$ (subject to a logarithmic term), and by approximating Kolmogorov complexity K using a practical data compressor C, a normalized compression distance (NCD) was proposed in [1] as:

$$\mathrm{NCD}(x, y) = \frac{C(xy) - \min\{C(x), C(y)\}}{\max\{C(x), C(y)\}}. \tag{4}$$

NCD has been proved to be an effective approximation of NID and achieves superior performance in bioinformatics applications such as the construction of phylogeny trees using DNA sequences [1].

When NCD was used to quantify image similarities, it did not achieve the same level of success as in other application fields. For example, it was reported in [4] that NCD works well when parts are added or subtracted from an image, but struggles when image variations involve form, material, and structure.

To overcome the difficulties in applying NCD for image similarity applications, normalized conditional compression distance (NCCD) was proposed in [3]. NCCD uses a general conditional compressor framework to describe the simplest transformation between two images:

$$K(y|x) \approx C_T(y|x) \quad \text{and} \quad K(x|y) \approx C_T(x|y). \tag{5}$$

This leads to a normalized conditional compression distance (NCCD) measure given by

$$\mathrm{NCCD}(x, y) = \frac{\max\{C_T(x|y), C_T(y|x)\}}{\max\{C(x), C(y)\}}\,. \tag{6}$$

and the conditional compressor C_T is defined as follows: Let $\{T_i|i = 1, \ldots, N\}$ be the set of transformations, let $T_i(x)$ represents the transformed image when applying the i-th transform to image x, and let $p(T_i, x)$ denotes the parameters used in the transformation. Each type of transformation is also associated with a parameter compressor, and C_i^p denotes the parameter compressor of the i-th transformation. The conditional compressor can be defined as:

$$C_T(y|x) = \min_i\{C[y - T_i(x)] + C_i^p[p(T_i, x)] + \log_2(N)\}\,, \tag{7}$$

where C remains to be a practical image compressor which encodes the difference between y and the transformed image $T_i(x)$, and the $\log_2(N)$ term computes the number of bits required to encode the selection of one out of N potential transformations.

3 Proposed method

Assuming stationarity and ergodicity for a source with probability measure μ, the following theorem applies [5]:

$$\lim_{n\to\infty} \frac{K(x^n)}{n} = H(\mu) \tag{8}$$

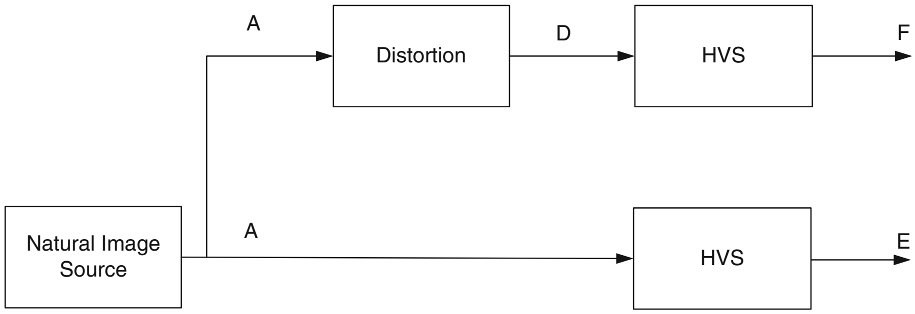

Fig. 1 Natural image scene statistics (NSS) is corrupted by a distortion channel (D) and then passes through the human visual system (HVS). Mutual information between A and E quantifies the amount of information extracted by the HVS from an original image, and mutual information between A and F quantifies the amount of information extracted from a distorted image [6].

Where x^n is an object of the source X and n shows the length of the object, with $K(x^n)$ being Kolmogorov complexity of the object. However, stationarity and ergodicity are not good assumptions of image signals in spatial domain; thus, we are interested in transform domain representations of images, where more reasonable assumptions may be made. Specifically, we transform the reference and distorted images into the wavelet domain and assume stationarity, decorrelation, and local independence among the image subbands. Based on these assumptions, we have:

$$K(x|y) = \sum_{i=1}^{n} K(x_n|y_n) \tag{9}$$

where K stands for Kolmogorov complexity of the images and x_n and y_n are the corresponding wavelet subbands of the reference and distorted image, respectively. Based on (9), the NID can be reformulated into:

$$\text{NID}(x, y) = \frac{\max\{\sum_{i=1}^{n} K(x_n|y_n), \sum_{i=1}^{n} K(y_n|x_n)\}}{\max\{\sum_{i=1}^{n} K(x_n), \sum_{i=1}^{n} K(y_n)\}} \tag{10}$$

Based on the stationarity and independence assumptions for wavelet coefficients, we may rewrite (10) in Shannon entropy framework based on (8):

$$\text{NID}_s(x, y) = \frac{\max\{\sum_{i=1}^{n} H(x_n|y_n), \sum_{i=1}^{n} H(y_n|x_n)\}}{\max\{\sum_{i=1}^{n} H(x_n), \sum_{i=1}^{n} H(y_n)\}} \tag{11}$$

Since GSM is capable of modeling important statistical features of natural images, such as heavy-tailed marginal distributions of the wavelet coefficients of natural images and nonlinear dependencies among them [8], we adopt a wavelet-domain Gaussian scale mixture (GSM) model for natural scene statistics. This is also consistent with visual information fidelity (VIF) method introduced in [6]. We also use the same distortion and human visual system (HVS) models used in [6].

A Gaussian scale mixture (GSM) is a random field which is represented as a product of a zero-mean Gaussian random vector U, and an independent scalar random field, S. In this sense, the GSM, A is defined to be $A = \{\overrightarrow{A}_i : i \in I\}$, where I is the set of spatial indices for the random field [6], and:

$$A = SU = \{S_i \overrightarrow{U}_i : i \in I\} \tag{12}$$

where $S = \{S_i : i \in I\}$ is a field of random scalars and $U = \{\overrightarrow{U}_i : i \in I\}$ is the random field of Gaussian vectors. It is easy to show that with the above assumption, the $\overrightarrow{A}_i$ vectors are normally distributed and independent given the random scalar s, which makes GSM easier to use in modeling clusters of coefficients of image subbands. It has also been shown that GSM is capable of modeling important statistical features of natural images, such as heavy-tailed marginal distributions of the wavelet coefficients of natural images and nonlinear dependencies among them [6]. This makes the GSM model more appealing for our application.

In the case that A is a GSM representing a cluster of coefficients in a natural image, we have:

$$E = A + \mathcal{N} \tag{13}$$

$$F = D + \mathcal{N}' \tag{14}$$

where $\mathcal{N}$ and $\mathcal{N}'$ are random fields of uncorrelated multivariate Gaussian noise that represents the internal neural noise in the HVS. D is a distortion model comprised of a signal gain and additive noise:

$$D = gA + \nu = gsU + \nu \tag{15}$$

and E and F are the clusters of coefficients in the reference and the distorted images, respectively [6]. Figure 1 shows a graphical representation of this model.

Let $\overrightarrow{A}_j = (\overrightarrow{A_1}, \ldots, \overrightarrow{A_N})_j$ represents N elements of a subband A_j, and $\overrightarrow{D}_j$, $\overrightarrow{E}_j$, $\overrightarrow{F}_j$ be defined correspondingly. Assuming that $S^N = s^N$ and is given for all the variables,[1] then NIDS becomes:

[1] Hence $\cdot|s^N$ is dropped in all notations.

$$\mathrm{NID}_s(E,F) = \frac{\max\{\sum_{j=1}^{n} H(\overrightarrow{E}_j|\overrightarrow{F}_j), \sum_{j=1}^{n} H(\overrightarrow{F}_j|\overrightarrow{E}_j)\}}{\max\{\sum_{j=1}^{n} H(\overrightarrow{E}_j), \sum_{j=1}^{n} H(\overrightarrow{F}_j)\}} \quad (16)$$

which can be further simplified into:

$$\mathrm{NID}_s(E,F) = 1 - \frac{\sum_j I(\overrightarrow{E}_j;\overrightarrow{F}_j)}{\max\{\sum_j H(\overrightarrow{E}_j), \sum_j H(\overrightarrow{F}_j)\}} \quad (17)$$

Since both $\overrightarrow{E}_j$ and $\overrightarrow{F}_j$ are continuous random variables, direct calculation of their entropies is difficult while their differential entropies do not provide adequate measures of their information content. To overcome this problem, we use the information content contained in the source $\overrightarrow{A}_j$ as the baseline and replace $H(\overrightarrow{E}_j)$ and $H(\overrightarrow{F}_j)$ with $I(\overrightarrow{A}_j;\overrightarrow{E}_j)$ and $I(\overrightarrow{A}_j;\overrightarrow{F}_j)$, respectively, which quantify the information content that is perceived by the HVS in the original and distorted images. We then define a normalized perceptual information distance (NPID) as:

$$\mathrm{NPID}(E,F) = 1 - \frac{\sum_j I(\overrightarrow{E}_j;\overrightarrow{F}_j)}{\max\{\sum_j I(\overrightarrow{E}_j;\overrightarrow{A}_j), \sum_j I(\overrightarrow{F}_j;\overrightarrow{A}_j)\}} \quad (18)$$

A normalized perceptual information similarity (NPIS) can be defined as:

$$\mathrm{NPIS}(E,F) = 1 - \mathrm{NPID}(E,F) = \frac{\sum_j I(\overrightarrow{E}_j;\overrightarrow{F}_j)}{\max\{\sum_j I(\overrightarrow{E}_j;\overrightarrow{A}_j), \sum_j I(\overrightarrow{F}_j;\overrightarrow{A}_j)\}} \quad (19)$$

(19) can be further simplified by using the fact that $\overrightarrow{E}_j$ and $\overrightarrow{F}_j$ are Gaussian for given s and the mutual information of correlated Gaussians can be calculated based on the determinants of the covariances [9]:

$$I(\overrightarrow{E}_j;\overrightarrow{F}_j) = \frac{1}{2}\log\left[\frac{|\mathbf{C}_E||\mathbf{C}_F|}{|\mathbf{C}_{(F,E)}|}\right] \quad (20)$$

$$I(\overrightarrow{E}_j;\overrightarrow{A}_j) = \frac{1}{2}\log\left[\frac{|\mathbf{C}_A||\mathbf{C}_E|}{|\mathbf{C}_{(A,E)}|}\right] \quad (21)$$

$$I(\overrightarrow{F}_j;\overrightarrow{A}_j) = \frac{1}{2}\log\left[\frac{|\mathbf{C}_A||\mathbf{C}_F|}{|\mathbf{C}_{(A,F)}|}\right] \quad (22)$$

Covariance matrices of A, D, E, and F are, respectively, computed to be [9]:

$$\mathbf{C}_A = s^2\mathbf{C}_\mathbf{U} \quad (23)$$

$$\mathbf{C}_D = g^2s^2\mathbf{C}_\mathbf{U} + \sigma_v^2\mathbf{I} \quad (24)$$

$$\mathbf{C}_E = s^2\mathbf{C}_\mathbf{U} + \sigma_n^2\mathbf{I} \quad (25)$$

$$\mathbf{C}_F = g^2s^2\mathbf{C}_\mathbf{U} + (\sigma_v^2+\sigma_n^2)\mathbf{I} \quad (26)$$

and we also have:

$$|\mathbf{C}_{(E,F)}| = \begin{vmatrix} \mathbf{C}_E & \mathbf{C}_{EF} \\ \mathbf{C}_{FE} & \mathbf{C}_F \end{vmatrix}, \quad (27)$$

$$|\mathbf{C}_{(A,E)}| = \begin{vmatrix} \mathbf{C}_A & \mathbf{C}_{AE} \\ \mathbf{C}_{EA} & \mathbf{C}_E \end{vmatrix}, \quad (28)$$

$$|\mathbf{C}_{(A,F)}| = \begin{vmatrix} \mathbf{C}_A & \mathbf{C}_{AF} \\ \mathbf{C}_{FA} & \mathbf{C}_F \end{vmatrix}. \quad (29)$$

It can be easily shown that: $\mathbf{C}_A = s^2\mathbf{C}_U$, $\mathbf{C}_{EF} = \mathbf{C}_{FE} = gs^2\mathbf{C}_\mathbf{U}$, $\mathbf{C}_{AE} = \mathbf{C}_{EA} = s^2\mathbf{C}_\mathbf{U}$ and $\mathbf{C}_{AF} = \mathbf{C}_{FA} = gs^2\mathbf{C}_\mathbf{U}$. Thus, (27)–(29) can be written as:

$$|\mathbf{C}_{(E,F)}| = \left|\left[\left(\sigma_v^2+\sigma_n^2\right)s^2 + \sigma_n^2 g^2 s^2\right]\mathbf{C}_\mathbf{U} + \sigma_n^2\left(\sigma_v^2+\sigma_n^2\right)\mathbf{I}\right| \quad (30)$$

$$|\mathbf{C}_{(A,E)}| = \left|\sigma_n^2 s^2 \mathbf{C}_\mathbf{U}\right| \quad (31)$$

$$|\mathbf{C}_{(A,F)}| = \left|\left(\sigma_v^2+\sigma_n^2\right)s^2\mathbf{C}_\mathbf{U}\right| \quad (32)$$

Since $\mathbf{C}_\mathbf{U}$ is a symmetric matrix, a further eigenvalue decomposition can be applied, which gives $\mathbf{C}_\mathbf{U} = \mathbf{Q}\Lambda\mathbf{Q}^T$, where $\mathbf{Q}$ is an orthogonal matrix and Λ is a diagonal matrix with eigenvalues λ_k for $k = 1, 2, \ldots, K$ along the diagonal entries. (23), (25), (26) can then be expressed as:

$$\mathbf{C}_A = \mathbf{Q}\{s^2\Lambda\}\mathbf{Q}^T \quad (33)$$

$$\mathbf{C}_E = \mathbf{Q}\{s^2\Lambda + \sigma_n^2\mathbf{I}\}\mathbf{Q}^T \quad (34)$$

$$\mathbf{C}_F = \mathbf{Q}\{g^2s^2\Lambda + (\sigma_v^2+\sigma_n^2)\mathbf{I}\}\mathbf{Q}^T \quad (35)$$

(30) – (32) become:

$$|\mathbf{C}_{(E,F)}| = |\mathbf{Q}\{[\sigma_v^2 + (1+g^2)\sigma_n^2]s^2\Lambda + \sigma_n^2(\sigma_v^2+\sigma_n^2)\mathbf{I}\}\mathbf{Q}^T| \quad (36)$$

$$|\mathbf{C}_{(A,E)}| = |\mathbf{Q}\{\sigma_n^2 s^2\Lambda\}\mathbf{Q}^T| \quad (37)$$

$$|\mathbf{C}_{(A,F)}| = |\mathbf{Q}\{(\sigma_v^2+\sigma_n^2)s^2\Lambda\}\mathbf{Q}^T| \quad (38)$$

Since $\mathbf{Q}$ is orthogonal and the middle matrices between $\mathbf{Q}$'s in (36),(37), and (38) are diagonal, calculation of the desired determinants can be simplified to a great extent [9]:

$$|\mathbf{C}_{E,F}| = \prod_{k=1}^{K}\left\{\left[\sigma_v^2 + (1+g^2)\sigma_v^2\right]s^2\lambda_k + \sigma_n^2\left(\sigma_v^2+\sigma_n^2\right)\right\} \quad (39)$$

$$|\mathbf{C}_{A,E}| = \prod_{k=1}^{K}\left\{\sigma_n^2 s^2\lambda_k\right\} \quad (40)$$

$$|\mathbf{C}_{A,F}| = \prod_{k=1}^{K}\left\{\left(\sigma_v^2+\sigma_n^2\right)s^2\lambda_k\right\} \quad (41)$$

Thus, the mutual information relations (20), (21), and (22) are simplified:

$$I(\vec{A}_j;\vec{E}_j)=\frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{K}\log_2\left(1+\frac{s_i^2\lambda_k}{\sigma_n^2}\right) \tag{42}$$

$$I(\vec{A}_j;\vec{F}_j)=\frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{K}\log_2\left(1+\frac{g^2 s_i^2\lambda_k}{\sigma_n^2+\sigma_v^2}\right) \tag{43}$$

$$I(\vec{E}_j;\vec{F}_j)=\frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{K}\log_2 \times\left(\frac{\left[g^2 s_i^2\lambda_k+\left(\sigma_v^2+\sigma_n^2\right)\right]\left(s_i^2\lambda_k+\sigma_n^2\right)}{\left[\left(\sigma_v^2+\sigma_n^2\right)s_i^2+\sigma_n^2 g^2 s_i^2\right]\lambda_k+\sigma_n^2\left(\sigma_n^2+\sigma_v^2\right)}\right) \tag{44}$$

To finish the computation of NPIS, we conclude by estimating a set of parameters involved in the calculations, including $\mathbf{C_U}$, s^2, g and σ_v^2. We follow the same path taken by [6,9]:

$$\hat{\mathbf{C}}_\mathbf{U}=\frac{1}{N}\sum_{i=1}^{N}\mathbf{A}_i\mathbf{A}_i^T \tag{45}$$

where N is the number of evaluation windows in the corresponding subband and A_i is the ith neighborhood coefficient. The multiplier s is estimated using a maximum-likelihood estimator [9]:

$$\hat{s}^2=\frac{1}{K}\mathbf{A}^T\mathbf{C}_\mathbf{U}^{-1}\mathbf{A} \tag{46}$$

Least square optimization may be used to estimate the parameters g and σ_v^2:

$$\hat{g}=\arg\min_g\|\mathbf{D}-g\mathbf{A}\|_2^2 \tag{47}$$

Take the derivative of the squared error cost function and let it be zero, we have:

$$\hat{g}=\frac{A^T D}{A^T A} \tag{48}$$

by substituting (48) into (15), we can estimate ${\sigma_v}^2$ by:

$$\hat{\sigma}_v^2=\frac{1}{K}(\mathbf{D}^T\mathbf{D}-\hat{g}\mathbf{A}^T\mathbf{D}) \tag{49}$$

In practice, when calculating the NPIS, we apply a five-scale Laplacian pyramid decomposition [10] to the original and distorted images, and compute the respective mutual information according to (42), (43), and (44) using a sliding 3×3 window that runs across each subband. At each location, the window contains a spatial neighborhood of ten coefficients (3×3 neighboring coefficients and one parent coefficient, thus $K=10$).

In [9], the authors propose an optimal pooling strategy for IQA algorithms using a multi-scale information content weighting approach based on a GSM model of natural images. The information weighting scheme is based on estimating the total perceptual information content for the reference and distorted images from evaluation of local information content of the images. It is shown that information content weighting often leads to significant improvements in performance of IQA methods. To incorporate this scheme, we first define a local-NPIS (L-NPIS) measure by:

$$\text{L-NPIS}_i=\frac{I(\vec{E}_i;\vec{F}_i)}{\max\{I(\vec{E}_i;\vec{A}_i),I(\vec{F}_i;\vec{A}_i)\}} \tag{50}$$

We can then compute an information content-weighted NPIS (IW-NPIS) measure using:

$$\text{IW-NPIS}_j=\frac{\sum_i\omega_{j,i}\text{L-NPIS}_i}{\sum_i\omega_{j,i}} \tag{51}$$

where $w_{j,i}$ is the weight assigned to the L-NPIS value calculated at the ith location at jth scale, and the value of $w_{j,i}$ is calculated based on the information content model given in [9]. Using the coarse-to-fine scale weights $\{\beta_1,\beta_2,\beta_3,\beta_4,\beta_5\} = \{0.0448, 0.2856, 0.3001, 0.2363, 0.1333\}$ from [11], we have:

$$\text{IW-LNPIS}=\prod_{j=1}^{M}(\text{IW-LNPIS}_j)^{\beta_j} \tag{52}$$

4 Results

To evaluate the performance of the proposed method, we test it using Laboratory for Image and Video Engineering (LIVE) [12], Tampere Image Database 2008 (TID2008) [13], and Categorial Image Quality (CSIQ) [14] databases and compare the results with a series of widely known and state-of-the-art IQA algorithms, including peak signal-to-noise ratio (PSNR) [9], Structural Similarity Index Measure (SSIM) [15], Information-Weighted Structural Similarity Index Measure (IW-SSIM) [9], visual information fidelity (VIF) [6], visual signal-to-noise ratio (VSNR) [16], HVS-based PSNR [17], information-weighted PSNR (IW-PSNR) [9], and most apparent distortion (MAD) [18].

The LIVE database consists of 29 original reference images contaminated by five types of distortions at different distortion levels. The distortion types include JPEG compression, JPEG2000 compression, white noise, Gaussian blur, and fast fading channel distortion of JPEG2000 compressed bitstream. A total of 982 subject-rated images are created from these distortions and the subjective scores of all images are adjusted according to an alignment process in which a cross-comparison of mixed images from all distortion types was done [12,9].

The TID2008 database consists of 25 original reference images contaminated by 17 distortion types at 4 different distortion levels. A total of 1,700 distorted images are generated and rated by subjects. The distortion types include additive Gaussian noise, additive noise where the noise in color components is more intensive than the noise in luminance components, spatially correlated noise, masked noise, high-frequency noise, impulse noise, quantization noise, Gaussian blur, image denoising, JPEG compression, JPEG2000 compression, JPEG transmission errors, JPEG2000 transmission errors, non-eccentricity pattern noise, local block-wise distortions of different intensity, mean shift, and contrast change [13,9].

The CSIQ database consists of 866 distorted images created from 30 original reference images using six types of distortions at four to five distortion levels. CSIQ images are subjectively rated based on a linear displacement of the images across four calibrated LCD monitors placed side by side with equal viewing distance to the observer. The database contains 5000 subjective ratings from 35 different observers, and ratings are reported in the form of DMOS. The distortion types include JPEG compression, JPEG2000 compression, global contrast decrements, additive pink Gaussian noise, and Gaussian blurring [14,9].

Five evaluation metrics are used to compare the IQA measures. Pearson linear correlation coefficient (PLCC), mean absolute error (MAE), root mean-squared error (RMS), Spearman's rank correlation coefficient (SRCC), and Kendall rank correlation coefficient (KRCC). A thorough explanation of these metrics is found in [9]. Among these metrics, PLCC, MAE, and RMS are adopted to evaluate prediction accuracy, and SRCC and KRCC are employed to assess prediction monotonicity. A better objective IQA measure should have higher PLCC, SRCC, and KRCC while lower MAE and RMS values [9].

The test results using the three databases are reported in Tables 1, 2, and 3. Figure 2 shows a sample scatter plot of NPIS versus three different subjective quality evaluation databases, where the subjective scores are given by Mean Opinion Score (MOS) or Difference of Mean Opinion Score (DMOS) between a distorted image and its corresponding original reference image. Each point in the scatter plot represents one image in the corresponding database. Figure 3 shows a similar scatter plot for IW-NPIS. Tables 1, 2, and 3 provide a comparison of the proposed methods with PSNR and state-of-the-art methods. The results of the proposed methods are in bold face. It can be observed that the proposed methods perform significantly and consistently better than PSNR and are in general comparable to many state-of-the-art algorithms. It takes roughly 5.62 s to calculate IW-NPIS for a typical set of 512×512 images on an Intel Core i3 M380 CPU at clock speed of 2.53 GHz, which is a moderate increase in computation time over 2.16 s required to calculate VIF on the same machine. This is impressive as an early attempt to use Kolmogorov complexity and NID theories for image quality assessment, where many existing methods are in their mature stages.

Table 1 Performance comparison based on LIVE [12] database

Model	PLCC	MAE	RMS	SRCC	KRCC
PSNR	0.8723	10.51	13.36	0.8756	0.6865
SSIM [15]	0.9449	6.933	8.946	0.9479	0.7963
IW-SSIM [9]	0.9556	6.212	8.047	0.9570	0.8197
VIF [6]	0.9598	6.148	7.667	0.9632	0.8270
VSNR [16]	0.9229	8.089	10.52	0.9271	0.7610
PSNR-HVS-M [17]	0.9251	7.966	10.37	0.9295	0.7659
MAD [18]	0.9394	7.293	9.368	0.9438	0.7920
NPIS	0.9211	7.458	8.491	0.9093	0.7514
IW-NPIS	0.9339	7.013	8.011	0.9376	0.7891

Table 2 Performance comparison based on TID2008 [13] database

Model	PLCC	MAE	RMS	SRCC	KRCC
PSNR	0.5223	0.8683	1.1435	0.5531	0.4027
SSIM [15]	0.7732	0.6546	0.8511	0.7749	0.5768
IW-SSIM [9]	0.8579	0.5276	0.6895	0.8559	0.6636
VIF [6]	0.8090	0.5990	0.7888	0.7496	0.5863
VSNR [16]	0.6820	0.6908	0.9815	0.7046	0.5340
PSNR-HVS-M [17]	0.5519	0.8036	1.1190	0.5612	0.4509
MAD [18]	0.7480	0.6641	0.8907	0.7708	0.5734
NPIS	0.7855	0.6239	0.8111	0.7682	0.5013
IW-NPIS	0.8244	0.5846	0.7637	0.8167	0.5819

Table 3 Performance comparison based on CSIQ [14] database

Model	PLCC	MAE	RMS	SRCC	KRCC
PSNR	0.7512	0.1366	0.1733	0.8058	0.6084
SSIM [15]	0.8612	0.0992	0.1334	0.8756	0.6907
IW-SSIM [9]	0.9144	0.0801	0.1063	0.9213	0.7529
VIF [6]	0.9277	0.0743	0.0980	0.9195	0.7537
VSNR [16]	0.7355	0.1335	0.1779	0.8109	0.6248
PSNR-HVS-M [17]	0.7725	0.1290	0.1667	0.8222	0.6529
MAD [18]	0.8202	0.1258	0.1502	0.8988	0.7272
NPIS	0.8999	0.1024	0.1188	0.8643	0.5920
IW-NPIS	0.9023	0.0923	0.1070	0.8985	0.6413

5 Conclusion

In this paper, we extend the application of Kolmogorov complexity and NID to image distortion and quality assessment by employing a wavelet-domain Gaussian scale mixture model of images and by estimating Kolmogorov complexity based on Shannon entropy approach. We show that the resulting image similarity measure is competitive with respect to state-of-the-art image quality assessment algorithms when tested using publicly available subject-rated

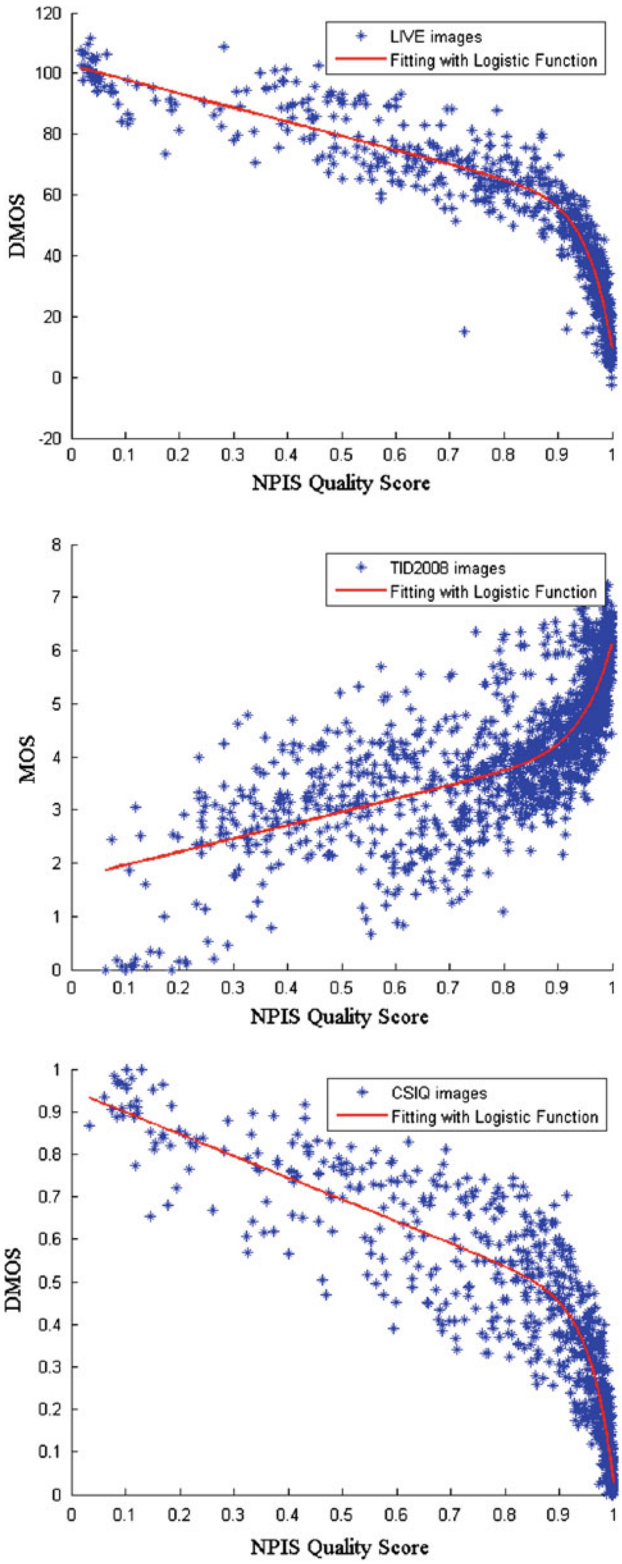

Fig. 2 Scatter plots of NPIS versus subjective scores for LIVE [12] TID2008 [13] and CSIQ [14] databases by NPIS

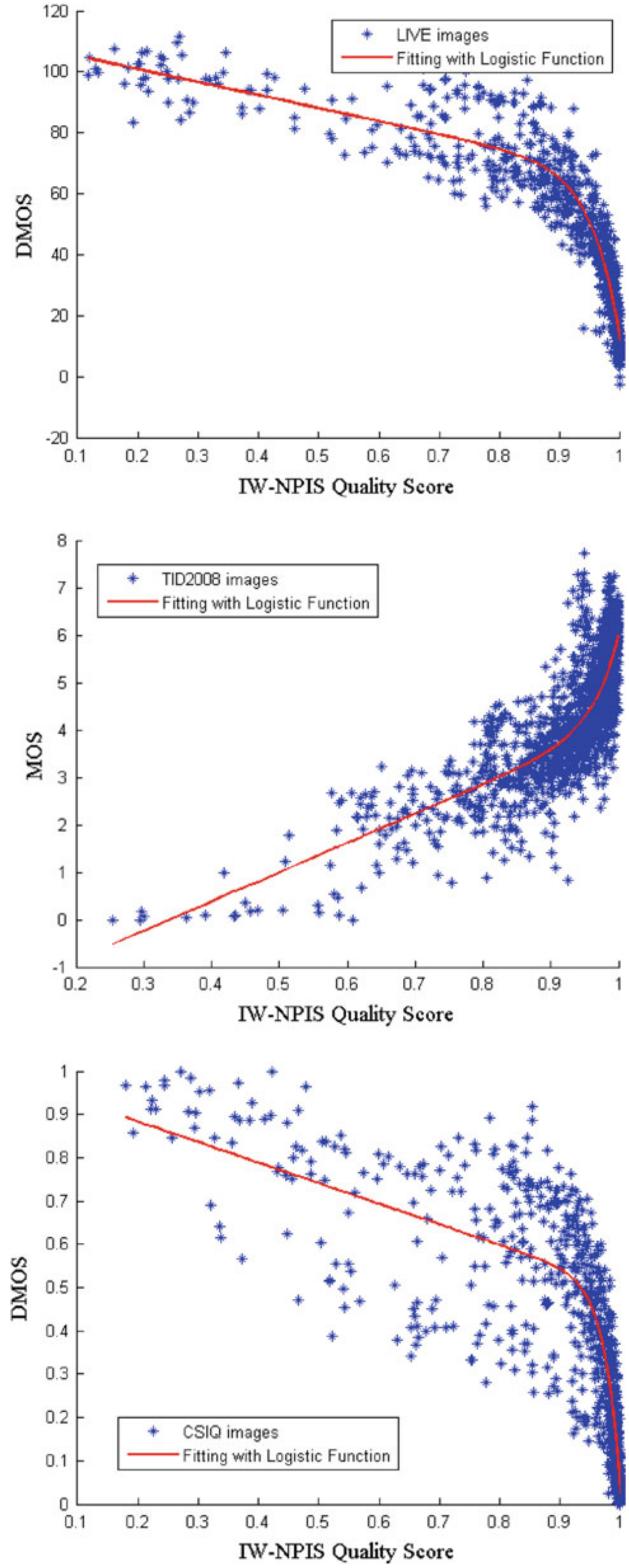

Fig. 3 Scatter plots of IW-NPIS versus subjective scores for LIVE [12] TID2008 [13] and CSIQ [14] databases by IW-NPIS

large-scale image databases. The proposed method draws some connections between the theories of Kolmogorov Complexity, NID, Shannon entropy, statistical image modeling, and real-world image processing applications.

Acknowledgments This research was supported in part by the Natural Sciences and Engineering Research Council of Canada, and in part by Ontario Early Researcher Award program, which are gratefully acknowledged.

References

1. Li, M., Chen, X., Li, X., Ma, B., Vitányi, P.M.B.: The similarity metric. IEEE Trans. Inf. Theory **50**, 3250–3264 (2004)
2. Cilibrasi, R., Vitányi, P.M.B.: Clustering by compression. In: IEEE Trans. Inf. Theory **51**, 1523–1545 (2005)
3. Nikvand, N., Wang, Z.: Generic image similarity based on kolmogorov complexity. In: Proceedings of International Conference on Image Processing (ICIP), 2010
4. Tran, N.: The normalized compression distance and image distinguishability. In: The 19th IS&T/SPIE Symposium on Electronic Imaging Science and Technology. Jan, San Jose (2007)
5. Sow, D.M., Eleftheriadis, A.: Complexity distortion theory. In: IEEE Trans. Inf. Theory **49**(3), 604–608 (2003)
6. Sheikh, H.R., Bovik, A.C.: Image information and visual quality. IEEE Trans. Image Process. **15**(2), 430–444 (2006)
7. Li, M., Vitányi, P.: An Introduction to Kolmogorov Complexity and Its Applications, 2nd edn. Springer, Berlin (1997)
8. Wainwright, M.J., Simoncelli, E.P.: Scale mixtures of gaussians and the statistics of natural images. In: Solla, S.A., Leen, T.K., Müller, K.-R. (eds.) Advances in Neural Information Processing Systems, vol. 12, pp. 855–861. MIT Press, Cambridge (2000)
9. Wang, Z., Li, Q.: Information content weighting for perceptual image quality assessment. IEEE Trans Image Process. **20**(5), 1185–1198 (2011)
10. Burt, P.J., Adelson, E.H.: The laplacian pyramid as a compact image code. IEEE Trans. Commun. **31**, 532–540 (1983)
11. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multi-scale structural similarity for image quality assessment. In: Proceedings of IEEE Asilomar Conference on Signals, Systems, and Computers (Pacific Grove, CA), Nov 2003, pp. 1398–1402
12. Sheikh, H.R., Seshadrinathan, K., Moorthy, K., Wang, Z., Bovik, A.C., Cormack, L.K.: Image and video quality assessment research at LIVE. (Online) Available http://live.ece.utexas.edu/research/quality
13. Ponomarenko, N., Egiazarrian, K.: Tampere image database 2008 TID2008. (Online) Available http://www.ponomarenko.info/tid2008.htm
14. Larson, E.C., Chandler, D.M.: Categorial image quality (CSIQ) database. (Online) Available http://vision.okstate.edu/csiq
15. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
16. Chandler, D.M., Hemami, S.S.: Vsnr: a wavelet-based visual signal-to-noise-ratio for natural images. IEEE Trans. Image Process. **16**, 2284–2298 (2007)
17. Ponomarenko, N., Silvestri, F., Egiazarian, K., Carli, M., Astola, J., Lukin, V.: On between-coefficient contrast masking of dct basis functions. In: 3rd International Workshop on Video Processing and functions. Scottsdale, Arizona, USA, Jan 2007
18. Larson, E.C., Chandler, D.M.: Most apparent distortion: full reference image quality assessment and the role of strategy. J. Electron. Imaging, **19**, 011006:1–21, Jan–Mar 2010
19. Cover, Thomas M., Thomas, Joy A.: Elements of Information Theory. Wiley-Interscience, New York (1991)