Presented at International Conference on Applied Mathematics, Modeling and Computational Science (AMMCS), Waterloo, ON, Canada, July 2011

PERCEPTUAL NORMALIZED INFORMATION DISTANCE FOR IMAGE DISTORTION ANALYSIS BASED ON KOLMOGOROV COMPLEXITY

Nima Nikvand and Zhou Wang

Dept. of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada
Email: nnikvand@uwaterloo.ca, zhouwang@ieee.org

ABSTRACT

Image distortion analysis is a fundamental issue in many image processing problems, including compression, restoration, recognition, classification, and retrieval. In this work, we investigate the problem of image distortion measurement based on the theories of Kolmogorov complexity and normalized information distance (NID), which have rarely been studied in the context of image processing. Based on a wavelet domain Gaussian scale mixture model of images, we approximate NID using a Shannon entropy based method. This leads to a series of novel distortion measures that are competitive with state-of-the-art image quality assessment approaches.

1. INTRODUCTION

Normalized Information Distance (NID) measure has been shown to be a valid and universal distance metric applicable to similarity measurement of any two objects [1]. Similar to Kolmgorov 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 arises in practice in the form of images or complex video signals can not generally be considered as stationary processes. In fact all images are non-stationary sources in spatial domain and 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], Gaussian Scale Mixture (GSM) model is adopted for Natural Scene Statistics (NSS), to further simplify 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: $K(x) = \min_{p:U(p)=x} l(p)$. 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 $\max\{K(x|y), K(y|x)\}$, 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]:

$$\text{NID}(x, y) = \frac{\max\{K(x|y^*), K(y|x^*)\}}{\max\{K(x), K(y)\}}\,. \tag{1}$$

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:

$$\text{NCD}(x, y) = \frac{C(xy) - \min\{C(x), C(y)\}}{\max\{C(x), C(y)\}}. \tag{2}$$

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: $K(y|x) \approx C_T(y|x) \quad and \quad K(x|y) \approx C_T(x|y)$. This leads to a normalized conditional compression distance (NCCD) measure given by

$$\text{NCCD}(x,y) = \frac{\max\{C_T(x|y), C_T(y|x)\}}{\max\{C(x), C(y)\}}, \tag{3}$$

and the conditional compressor C_T is defined as follows: Let $\{T_i|i=1,\cdots,N\}$ be the set of transformations, let $T_i(x)$ represent the transformed image when applying the i-th transform to image x, and let $p(T_i,x)$ denote 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: $C_T(y|x) = \min_i\{C[y - T_i(x)] + C_i^p[p(T_i,x)] + \log_2(N)\}$, 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 two theorems apply [5]: $\lim_{n\to\infty} \frac{K(x^n)}{n} = H(\mu)$, 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 the stationarity and ergodicity assumptions are not true for images and in fact all images are non-stationary sources in spatial domain, thus we are interested in transform domains where we can model the decomposed signal as stationary. Multiscale decomposition is one of the choices in this field. To continue with the problem, we take the reference and the distorted images into the wavelet domain, and assume stationarity, decorrelation and local independence among the image subbands. Based on these assumptions we have: $C(x|y) = \sum_{i=1}^n C(x_n|y_n)$, where C stands for Kolmogorov complexity of the images and x_n and y_n are the corresponding wavelet subbands of the reference image and distorted image. Thus the NID can be reformulated into:

$$\text{NID}(x,y) = \frac{\max\{\sum_{i=1}^n C(x_n|y_n), \sum_{i=1}^n C(y_n|x_n)\}}{\max\{\sum_{i=1}^n C(x_n), \sum_{i=1}^n C(y_n)\}}, \tag{4}$$

and based on the stationarity and independence assumptions for wavelet coefficients, we may re-write (4) in Shannon entropy framework:

$$\text{NID}(x,y) = \frac{\max\{\sum_{i=1}^n H(x_n|y_n), \sum_{i=1}^n H(y_n|x_n)\}}{\max\{\sum_{i=1}^n H(x_n), \sum_{i=1}^n H(y_n)\}}. \tag{5}$$

To further simplify the results we adopt a wavelet domain Gaussian Scale Mixture (GSM) model by natural scene statistics, which is similar to the model used in [6]. We also use the same distortion and human visual system (HVS) model used in [6]. In the case that C is a GSM representing the coefficients of a natural image we have $E = C + \mathcal{N}$, $F = D + \mathcal{N}'$, where $\mathcal{N}$ and $\mathcal{N}'$ are random fields of uncorrelated multivariate Gaussian noise that represent the internal neural noise in the HVS. D is a distortion model comprised of a signal gain and additive noise: $D = gC + \nu = gsU + \nu$, and E and F are the reference and the distorted images respectively [6]. Let $\overrightarrow{C}_j = (\overrightarrow{C_1}, \ldots, \overrightarrow{C_N})_j$ represent N elements of a subband C_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 NID becomes:

$$\text{NID}(E,F) = 1 - \frac{\sum_j I(\overrightarrow{E}_j; \overrightarrow{F}_j)}{\max\{\sum_j H(\overrightarrow{E}_j), \sum_j H(\overrightarrow{F}_j)\}}, \tag{6}$$

since $I(E;C)$ quantifies the information extracted by the HVS from an ideal image and $I(F;C)$ quantiefies the extracted information from a distorted image, they could be used to replace $H(E)$ and $H(F)$, respectively [6]. Since NID is a distance metric, 1-NID can be used as a similarity metric in our Image similarity and quality assessment applications. Here we define a Normalized Information Similarity (NIS) by 1-NID, which measures the perceptual image similarity by the HVS. According to our distortion and HVS models, the NIS for two images E and F, as perceived by the HVS, is defined as follows:

$$\text{NIS}(E,F) = \frac{\sum_j I(\overrightarrow{E}_j; \overrightarrow{F}_j)}{\max\{\sum_j I(\overrightarrow{E}_j; \overrightarrow{C}_j), \sum_j I(\overrightarrow{F}_j; \overrightarrow{C}_j)\}}. \tag{7}$$

(7) can be further simplified by using the fact that E and F are Gaussian for given s and the mutual information of correlated Gaussians can be calculated based on the determinants of the covariances [8]:

$$I(\overrightarrow{E}_j; \overrightarrow{F}_j) = \frac{1}{2}\log[\frac{|C_E||C_F|}{|C_{(F,E)}|}], \quad I(\overrightarrow{E}_j; \overrightarrow{C}_j) = \frac{1}{2}\log[\frac{|C_C||C_E|}{|C_{(C,E)}|}], \quad I(\overrightarrow{F}_j; \overrightarrow{C}_j) = \frac{1}{2}\log[\frac{|C_C||C_F|}{|C_{(C,F)}|}]. \tag{8,9,10}$$

[1] hence $.|s^N$ is dropped in all notations

Covariance matrices of C, D, E and F are respectively computed to be [8]:

$$C_C = s^2 C_{\mathbf{U}}, \quad C_D = g^2 s^2 C_{\mathbf{U}} + \sigma_\nu^2 \mathbf{I}, \quad C_E = s^2 C_{\mathbf{U}} + \sigma_n^2 \mathbf{I}, \quad C_F = g^2 s^2 C_{\mathbf{U}} + (\sigma_\nu^2 + \sigma_n^2)\mathbf{I}, \tag{11,12,13,14}$$

and we also have:

$$|C_{(E,F)}| = \begin{vmatrix} C_E & C_{EF} \\ C_{FE} & C_F \end{vmatrix}, \quad |C_{(C,E)}| = \begin{vmatrix} C_C & C_{CE} \\ C_{EC} & C_E \end{vmatrix}, \quad |C_{(C,F)}| = \begin{vmatrix} C_C & C_{CF} \\ C_{FC} & C_F \end{vmatrix}. \tag{15,16,17}$$

It can be easily shown that: $C_C = s^2 C_U$, $C_{EF} = C_{FE} = gs^2 C_{\mathbf{U}}$, $C_{CE} = C_{EC} = s^2 C_{\mathbf{U}}$ and $C_{CF} = C_{FC} = gs^2 C_{\mathbf{U}}$. Thus (15),(16) and (17) can be written as:

$$|C_{(E,F)}| = |[(\sigma_\nu^2 + \sigma_n^2)s^2 + \sigma_n^2 g^2 s^2] C_{\mathbf{U}} + \sigma_n^2(\sigma_\nu^2 + \sigma_n^2)\mathbf{I}|, \quad |C_{(C,E)}| = |\sigma_n^2 s^2 C_{\mathbf{U}}|, \quad |C_{(C,F)}| = |(\sigma_\nu^2 + \sigma_n^2) s^2 C_{\mathbf{U}}|, \tag{18,19,20}$$

since C_U is a symmetric matrix, a further decomposition of the covariance matrix into $C_{\mathbf{U}} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$ form can change (11), (13) and (14) into the following form [6]:

$$C_C = \mathbf{Q}\{s^2 \mathbf{\Lambda}\}\mathbf{Q}^T, \quad C_E = \mathbf{Q}\{s^2 \mathbf{\Lambda} + \sigma_n^2 \mathbf{I}\}\mathbf{Q}^T, \quad C_F = \mathbf{Q}\{g^2 s^2 \mathbf{\Lambda} + (\sigma_\nu^2 + \sigma_n^2)\mathbf{I}\}\mathbf{Q}^T, \tag{21,22,23}$$

and from (18), (19) and (20) we have:

$$|C_{(E,F)}| = \quad |\mathbf{Q}\{[\sigma_\nu^2 + (1+g^2)\sigma_n^2] s^2 \mathbf{\Lambda} + \sigma_n^2(\sigma_\nu^2 + \sigma_n^2)\mathbf{I}\}\mathbf{Q}^T|, \quad |C_{(C,E)}| = \quad |\mathbf{Q}\{\sigma_n^2 s^2 \mathbf{\Lambda}\}\mathbf{Q}^T|, \quad |C_{(C,F)}| = \quad |\mathbf{Q}\{(\sigma_\nu^2 + \sigma_n^2) s^2 \mathbf{\Lambda}\}\mathbf{Q}^T|. \tag{24,25,26}$$

Since $\mathbf{Q}$ is orthogonal and the middle matrices between Q's in (24),(25) and (26) are diagonal, calculation of the desired determinants can be simplified to a great extent [8]:

$$|C_{E,F}| = \prod_{k=1}^{K} \{[\sigma_\nu^2 + (1+g^2)\sigma_\nu^2] s^2 \lambda_k + \sigma_n^2(\sigma_\nu^2 + \sigma_n^2)\}, \quad |C_{C,E}| = \prod_{k=1}^{K} \{\sigma_n^2 s^2 \lambda_k\}, \quad |C_{C,F}| = \prod_{k=1}^{K} \{(\sigma_\nu^2 + \sigma_n^2) s^2 \lambda_k\}, \tag{27,28,29}$$

thus the mutual information relations (8), (9) and (10) are simplified:

$$I(\overrightarrow{C}_j; \overrightarrow{E}_j) = \frac{1}{2} \sum_{i=1}^{N} \sum_{k=1}^{K} \log_2\left(1 + \frac{s_i^2 \lambda_k}{\sigma_n^2}\right), \quad I(\overrightarrow{C}_j; \overrightarrow{F}_j) = \frac{1}{2} \sum_{i=1}^{N} \sum_{k=1}^{K} \log_2\left(1 + \frac{g^2 s_i^2 \lambda_k}{\sigma_n^2 + \sigma_\nu^2}\right), \tag{30,31}$$

$$I(\overrightarrow{E}_j; \overrightarrow{F}_j) = \frac{1}{2} \sum_{i=1}^{N} \sum_{k=1}^{K} \log_2\left(\frac{[g^2 s_i^2 \lambda_k + (\sigma_\nu^2 + \sigma_n^2)](s_i^2 \lambda_k + \sigma_n^2)}{[(\sigma_\nu^2 + \sigma_n^2)s_i^2 + \sigma_n^2 g^2 s_i^2]\lambda_k + \sigma_n^2(\sigma_n^2 + \sigma_\nu^2)}\right). \tag{32}$$

In computation of NIS, we estimate the required parameters ($\hat{C}$, $\hat{s}^2$, $\hat{g}$ and $\hat{\sigma_\nu}$) by following the same path taken by [6, 8].

4. RESULTS

To evaluate the performance of the proposed method, we validate the method against TID2008 [9] database and compare the results with those of PSNR [8], SSIM [10], IW-SSIM [8], VIF [6], VSNR [11], HVS based PSNR [12], IW-PSNR [8], and MAD [13].

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 [8]. 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 [8].

To incorporate the information content-based weighing approach proposed in [8] into the proposed NIS framework, we define a local NIS (LNIS) measure as:

$$\text{LNIS}_i = \frac{I(\overrightarrow{E}_i; \overrightarrow{F}_i)}{\max\{I(\overrightarrow{E}_i; \overrightarrow{C}_i), I(\overrightarrow{F}_i; \overrightarrow{C}_i)\}}. \tag{33}$$

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

$$\text{IW-NIS} = \frac{\sum_i \omega_i \text{LNIS}_i}{\sum_i \omega_i}, \tag{34}$$

where w_i is the weight assigned to the LNIS value calculated at the *i-th* location, and the value of w_i is calculated based on the information content model given in [8]. The test results using the TID2008 database are reported as IW-NIS in Table 1.

Table 1 provides a comparison of the proposed methods with PSNR and state-of-the-art methods. The results of the proposed method are in bold face. It can be observed that the proposed method perform significantly and consistently better than PSNR and are in general comparable to many state-of-the-art algorithms. 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. Table 1: Performance comparison based on TID2008 [9] database

Model	PLCC	MAE	RMS	SRCC	KRCC
PSNR	0.5223	0.8683	1.1435	0.5531	0.4027
SSIM [10]	0.7732	0.6546	0.8511	0.7749	0.5768
IW-SSIM [8]	0.8579	0.5276	0.6895	0.8559	0.6636
VIF [6]	0.8090	0.5990	0.7888	0.7496	0.5863
VSNR [11]	0.6820	0.6908	0.9815	0.7046	0.5340
PSNR-HVS-M [12]	0.5519	0.8036	1.1190	0.5612	0.4509
MAD [13]	0.7480	0.6641	0.8907	0.7708	0.5734
IW-NIS	**0.7811**	**0.6431**	**0.8431**	**0.7276**	**0.5601**

5. CONCLUSION

In this paper we extend the application of Kolmogorov complexity and NID to image distortion and image 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 large-scale image databases. The proposed method is the first information theoretic approach towards image quality assessment problem by using Normalized Information Distance (NID) and approximating Kolmogorov complexity using Shannon's entropy, and shows its potential for future research in this area.

6. ACKNOWLEDGMENT

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.

7. REFERENCES

[1] M. Li, X. Chen, X. Li, B. Ma, and P. M. B. Vitányi, "The similarity metric," *IEEE Trans. Info. Theory*, vol. 50, pp. 3250–3264, Dec. 2004.

[2] R. Cilibrasi and P. M.B. Vitányi, "Clustering by compression," *IEEE Trans. Info. Theory*, vol. 51, pp. 1523–1545, Apr. 2005.

[3] N. Nikvand and Z. Wang, "Generic image similarity based on kolmogorov complexity," in *Proceedings of International Conference on Image Processing (ICIP) 2010*.

[4] N. Tran, "The normalized compression distance and image distinguishability," in *The 19th IS&T/SPIE Symposium on Electronic Imaging Science and Technology*, San Jose, Jan. 2007.

[5] D. M. Sow and A. Eleftheriadis, "Complexity distortion theory," *IEEE Trans. Information Theory*, vol. 49, no. 3, pp. 604–608, Mar. 2003.

[6] H.R. Sheikh and A.C. Bovik, "Image information and visual quality," *IEEE Trans. Image Processing*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[7] M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*. Berlin, 2nd edition: Springer, 1997.

[8] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. on Image Processing*, vol. 20, no. 5, pp. 1185–1198, May 2011.

[9] N. Ponomarenko and K. Egiazarrian, "Tampere Image Database 2008 TID2008." [Online]. Available: {http://www.ponomarenko.info/tid2008.htm}

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

[11] D. M. Chandler and S. S, Hemami, "Vsnr: A wavelet-based visual signal-to-noise-ratio for natural images," *IEEE Trans. Image Processing*, vol. 16, pp. 2284–2298, Sept. 2007.

[12] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, "On between-coefficient contrast masking of dct basis functions," in *Third International Workshop on Video Processing and functions, (Scottsdale, Arizona,USA)*, Jan. 2007.

[13] E. C. Larson and D. M. Chandler, "Most apparent distortion: full reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, pp. 011 006:1–21, Jan.-Mar. 2010.