IEEE Inter. Conf. Image Processing, Hong Kong, China, Sept. 26-29, 2010.

REDUCED-REFERENCE SSIM ESTIMATION

Abdul Rehman and Zhou Wang

Dept. of Electrical & Computer Engineering, University of Waterloo, Waterloo, ON, Canada
Email: a5rehman@uwaterloo.ca, zhouwang@ieee.org

ABSTRACT

The structural similarity (SSIM) index has been shown to be a good perceptual image quality predictor. In many real-world applications such as network visual communications, however, SSIM is not applicable because its computation requires full access to the original image. Here we propose a reduced-reference approach that estimates SSIM with only partial information about the original image. Specifically, we extract statistical features from a multi-scale, multi-orientation divisive normalization transform and develop a distortion measure by following the philosophy analogous to that in the construction of SSIM. We found an interesting linear relationship between our reduced-reference SSIM estimate and full-reference SSIM when the image distortion type is fixed. A regression-by-discretization method is then applied to normalize our measure between image distortion types. We use the LIVE database to test the proposed distortion measure, which shows strong correlations with both SSIM and subjective evaluations. We also demonstrate how our reduced-reference features may be employed to partially repair a distorted image.

Index Terms— reduced-reference image quality assessment, structural similarity, natural image statistics, divisive normalization transform, regression by discretization, image repairing

1. INTRODUCTION

Multimedia contents delivered over networks suffer from various types of distortions on its way to the destination. It is highly desirable to measure the perceptual similarity of the received content with the original. The structure similarity (SSIM) index [1] was shown to correlate well with perceived image quality and has found a wide variety of applications, ranging from image coding, restoration and fusion, to watermarking and biometrics [2]. However, direct SSIM evaluation is not possible in practical visual communication applications, because it is a full-reference (FR) measure that requires the original image at the receiver [2]. On the other hand, the lack of knowledge of natural scene statistics and the human visual system (HVS) creates great challenge for no-reference image quality assessment (NR-IQA), especially for the general-purpose case. Reduced-reference (RR) IQA, a compromise between FR and NR, is designed to employ only a set of RR features extracted from the reference image for quality evaluation of the distorted image at the receiver [2].

To the best of our knowledge, only a few schemes have been presented in the literature for general purpose RR-IQA. In [3], the marginal distribution of wavelet subband coefficients is modeled using a generalized Gaussian density function, and the variations of marginal distributions are used to quantify image distortion. This led to an effective RR-IQA method with low RR data rate. This scheme was further improved in [4] by making use of a divisive normalization transform (DNT). An RR video SSIM metric was proposed in [5] for quantifying visual degradations caused by channel transmission error. It is based on local spatial statistical features and uses distributed source coding techniques to reduce the required bandwidth to transmit RR features, though the resulting RR data rate is still much higher than those in [3] and [4].

In this paper, we propose a new approach for the design of low data rate general-purpose RR-IQA method. Instead of directly constructing an RR algorithm to predict subjective quality evaluations, we develop our method as an attempt to estimate SSIM. The benefits of this approach are twofold. First, the successful design principle in the construction of SSIM can be naturally incorporated into the development of our algorithm. Second, when the algorithm design involves a supervised machine learning stage, it is much easier to obtain training data, because SSIM can be readily computed, as opposed to the expensive and time-consuming subjective tests. Our experiments using the LIVE database [6] show that this is a useful approach, as the resulting RR-SSIM estimator exhibits good performance in predicting not only FR-SSIM, but also subjective scores. Moreover, we also use a simple image deblurring example to show that the RR features employed in our approach can be employed to partially repair a distorted image.

2. RR SSIM ESTIMATION

The proposed RR-SSIM estimation algorithm starts from a feature extraction process of the reference image based on a multi-scale multi-orientation divisive normalization transform (DNT). Divisive normalization was found to be a simple but effective mechanism to account for many neuronal behaviors in biological perceptual systems [7]. In [7], a DNT is defined by using a Gaussian scale mixture (GSM) model of image wavelet coefficients. A vector Y of length N is a GSM if it can be represented as the product of two independent components: $Y \doteq zU$, where z is a scalar random variable called mixing multiplier, and U is a zero-mean Gaussian distributed random vector with covariance C_U. It was found that the histogram of normalized wavelet coefficient vector, $\nu = Y/\hat{z}$, can be modeled by a zero-mean Gaussian density function [7], where $\hat{z}$ is a local estimation of the multiplier z using a maximum-likelihood estimator [7]:

$$\hat{z} = \sqrt{Y^T C_U^{-1} Y / N}. \quad (1)$$

As a result, the DNT coefficient distribution of each subband is characterized by a single parameter σ, the standard deviation of the Gaussian distribution. This provides a very efficient summary of the reference image. In addition to σ, the Kullback-Leibler divergence (KLD) between model Gaussian distribution, $p_m(x)$, and the true probability distribution of the DNT-domain coefficients, $p(x)$, denoted by $d(p_m||p)$ is extracted as the second feature for each subband. The subband distortion of the distorted image can be evaluated

by the KLD between the probability distribution of the original image, $p(x)$, and that of the distorted image, $q(x)$:

$$\hat{d}(p||q) = d(p_m||q) - d(p_m||p)\,, \tag{2}$$

where $d(p_m||q)$ is the KLD between the model Gaussian distribution and the distribution computed from the distorted image. As demonstrated in [3, 4], different types of distortions affect the statistics of the reference image in a different manner, but are all summarized in Eq. (2) to a single distortion measure.

By assuming independence between subbands, the subband-level distortion measure of Eq. (2) can be combined to provide an overall distortion assessment of the whole image [4]

$$D = \log\left(1 + \frac{1}{D_0}\sum_{k=1}^{K}\left|\hat{d}^k(p^k||q^k)\right|\right), \tag{3}$$

where K is the total number of subbands, p^k and q^k are the probability distributions of the k-th subband of the reference and distorted images, respectively, $\hat{d}^k$ represents the KLD between p^k and q^k, and D_0 is a constant to control the scale of the distortion measure.

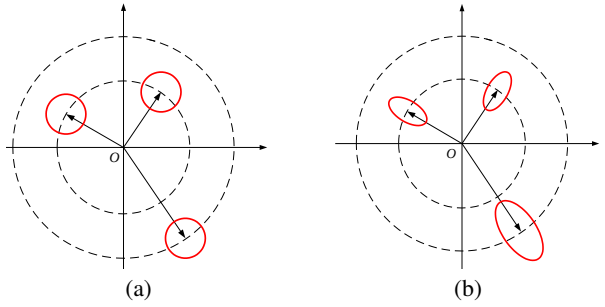

Fig. 1. Equal-distortion contours with respect to the central reference vectors. (a) MSE measure; (b) SSIM measure.

The limitation of the measure in Eq. (3) is that it does not take into account the relationship (or structures) between the distortions across different subbands. Such distortion structure is a critical issue behind the philosophy of the SSIM approach [1], which attempts to distinguish structural and non-structural distortions. Figure 1 provides a graphical explanation in the vector space of image components, where the image components can be pixels, wavelet coefficients, or extracted features from the reference image. For the purpose of illustration, two-dimensional diagrams are shown here. However, the actual dimensions may be equal to the number of pixels or features being compared. In the graphs for both MSE and SSIM measures, we use three vectors to represent three reference images, and the contour around each vector represents the set of images that have the same level of distortion with respect to the reference. Unlike the MSE metric, SSIM is totally adaptive according to the reference signal. In particular, if the distortion is consistent with the underlying reference signal (the reference vector direction), we call it a non-structural distortion, which is much less objectional than structural distortions (for example, the distortions perpendicular to the reference vector direction). This is reflected in the shapes of the equal-distortion contours. Here we make a first attempt to extend this idea for RR IQA by applying it to the subband standard deviation measures of the reference and distorted images. This is intuitively sensible because in the case that the distorted image is a globally contrast scaled (contrast reduction or enhancement) version of the reference image, then the standard deviations of all subbands should scale by the same factor, which is considered consistent non-structural distortion and is less objectional than the case that the subband standard deviations change in different ways.

Let $\boldsymbol{\sigma}_r$ and $\boldsymbol{\sigma}_d$ be the vectors containing the standard deviation σ values of the DNT coefficients from each subband in the reference and distorted images, respectively. We define a new RR distortion measure as

$$D_n = g(\boldsymbol{\sigma}_r, \boldsymbol{\sigma}_d)\log\left(1 + \frac{1}{D_0}\sum_{k=1}^{K}\left|\hat{d}^k(p^k||q^k)\right|\right), \tag{4}$$

where the key feature is the function $g(\boldsymbol{\sigma}_r, \boldsymbol{\sigma}_d)$ added in front of Eq. (3). This function should serve the purpose of identifying and distinguishing the consistent non-structural distortion directions in the feature vector space of subband σ values, so as to scale the distortion measure D in a way that structural distortions are penalized more than non-structural distortions. Motivated by the successful normalized correlation formulation in SSIM [1], we define $g(\boldsymbol{\sigma}_r, \boldsymbol{\sigma}_d)$ as

$$g(\boldsymbol{\sigma}_r, \boldsymbol{\sigma}_d) = \frac{|\boldsymbol{\sigma}_r|^2 + |\boldsymbol{\sigma}_d|^2 + C}{2|\boldsymbol{\sigma}_r \cdot \boldsymbol{\sigma}_d| + C}, \tag{5}$$

where $\boldsymbol{\sigma}_r \cdot \boldsymbol{\sigma}_d$ represents the dot product between the two vectors, and the constant C is included to avoid instability when $\boldsymbol{\sigma}_r \cdot \boldsymbol{\sigma}_d$ is close to $\mathbf{0}$. This function is lower-bounded by 1, when $\boldsymbol{\sigma}_r$ and $\boldsymbol{\sigma}_d$ are fully correlated, or in other words, when their orientations in the feature vector space are completely consistent. With the decrease of correlation, $g(\boldsymbol{\sigma}_r, \boldsymbol{\sigma}_d)$ increases, thus gives more penalty to inconsistent structural distortions.

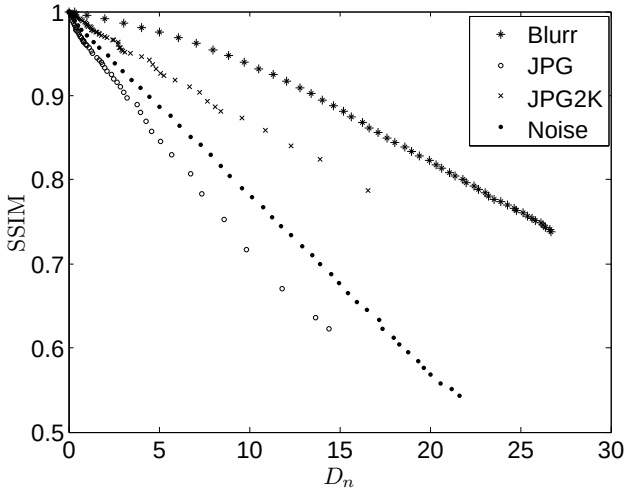

Fig. 2. Relationship between SSIM and D_n for blur, JPEG compression, JPEG2000 compression, and noise contamination distortions.

Figure 2 shows the D_n results computed for 4 different distortion types at different distortion levels, and compares them with the corresponding SSIM values calculated for the distorted images. Interestingly, for each fixed distortion type, D_n exhibits a nearly perfect linear relationship with SSIM. We regard this as an outcome of the similarity between their design principles, even though the principle is applied to completely different domains of signal representation. The clean linear relationship also helps us to design an

SSIM predictor based on D_n because the remaining job is just to estimate the normalization slope factor across distortion types. More specifically, an RR-SSIM estimator can be written as

$$\hat{S} = 1 - \alpha D_n, \tag{6}$$

where α is the slope factor that needs to be learned from training images. In particular, we adopted a regression-by-discretization approach [8], which is a regression scheme that employs a classifier on a copy of the data that has the class attribute discretized, and the predicted value is the expected value of the mean class value for each discretized interval. A decision tree classifier was built using $|\sigma_o - \sigma_d|$ and $|k_r - k_d|$ as the attributes, where k_r and k_d are the kurtosis values of the DNT coefficients computed from the original and distorted images, respectively.

3. IMPLEMENTATION AND VALIDATION

To extract RR features, the reference image is first decomposed into 12 subbands using a three-scale four-orientation steerable pyramid transform [9]. Division normalization is then applied using 13 neighboring coefficients, including 9 spatial neighbors from the same subband, 1 from parent subband, and 3 from the same spatial location in the other orientation bands at the same scale. Three features, σ_r, k_r and $d(p_m||p)$, are extracted for each subband, resulting in a total of 36 RR features for a reference image. These RR features are used for SSIM estimation of the distorted image using the approach described in Section 2.

A training process is needed to determine the slope factor α based on the observed differences between subband standard deviation and kurtosis. Our training data included 29 reference images altered with 50 levels of distortions for five types of distortions, including Gaussian Blur, JPEG2000 compression, JPEG compression, fast fading channel distortion of JPEG2000 compressed bitstream and white Gaussian noise. Decision trees were built using the open source data mining tool WEKA [10].

The proposed scheme is tested using the LIVE database [6], which contains seven data sets with a total of 779 distorted images. Figure 3 shows the scatter plot, and Table 1 computes the mean absolute error (MAE) and Pearson linear correlation coefficient (PLCC) between FR SSIM and our RR SSIM estimate. It can be seen that the proposed SSIM estimator achieves high prediction accuracy across various types of distortions.

To further validate the proposed algorithm, we compare three objective IQA algorithms, namely peak signal-to-noise-ratio (PSNR), SSIM, and our RR SSIM estimate, with subjective quality evaluations (in particular, the differences of mean opinion scores) available in the LIVE database [6]. Four metrics are employed for evaluation, which include PLCC and MAE after nonlinear mapping between subjective and objective scores, Spearman's rank correlation coefficient (SRCC), and Kendall's rank correlation coefficient (KRCC). The results are shown in Table 2. It can be observed that in general the proposed method performs inferior to SSIM (which is as expected) and significantly outperforms PSNR. It needs to be mentioned that the comparison is unfair to the proposed method, because the other two are FR measures. However, It outperforms the already existing general purpose RR measures in the literature [3] [4].

4. IMAGE REPAIRING USING RR FEATURES

Since the RR features reflect certain statistical properties about the reference signal, they may be used to partially "repair" the distorted image. Here we provide an example that uses RR features to correct a blurred image. Since blur reduces energy at mid and high frequencies, the subband standard deviation σ_d of DNT coefficients in the distorted image is smaller than that of the reference image σ_r. The most straightforward way to enforce a corrected image to have the same statistical property as the reference image is to scale up all the DNT coefficients in each subband i of the distorted image by a fixed scale factor $s^i = \sigma_r^i/\sigma_d^i$:

$$\nu^i_{repaired} = s^i \nu^i_d \,. \tag{7}$$

Figure 4(d) compares the histograms of the reference, distorted and repaired DNT coefficients. It can be observed that the histogram of scaled DNT coefficients is very close to that of the reference image.

To reconstruct the repaired image, it remains to invert the DNT transform, where the critical issue is to estimate the local scalar multiplier $\hat{z}$. Based on Eq. (1), the scalar multiplier for inverse DNT is given by

$$\begin{aligned} \hat{z}_{\rm inv} &= \sqrt{(sY)^T (C_U^{-1}/s^2)(sY)/N} \\ &= \sqrt{Y^T C_U^{-1} Y/N} = \hat{z} \,. \end{aligned} \tag{8}$$

This largely simplifies the inversion, as we have already calculated $\hat{z}$. We can then compute the wavelet coefficients using $\hat{z}_{\rm inv}\nu_{repaired}$,

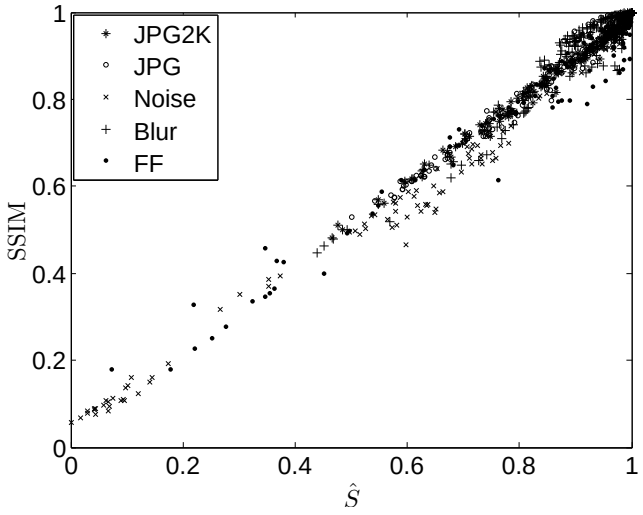

Fig. 3. SSIM versus RR SSIM estimation $\hat{S}$ for LIVE database.

Table 1. MAE and PLCC between SSIM and RR SSIM estimation $\hat{S}$ for LIVE database

	MAE	PLCC
JP2 (1)	0.0107	0.9829
JP2 (2)	0.0098	0.9894
JPG (1)	0.0147	0.9603
JPG (2)	0.0111	0.9877
Noise	0.0178	0.9816
Blur	0.0156	0.9624
FF	0.0206	0.9760
All data	0.0155	0.9802

(a) (b) (c) (d)

Fig. 4. Repairing blurred image using RR features. (a) Original "building" image; (b) Blurred image, SSIM = 0.674, $\hat{S}$ = 0.662; (c) Repaired, image SSIM = 0.918, $\hat{S}$ = 0.928; (d) DNT coefficient histograms of original, distorted and repaired images.

Table 2. Performance comparison of IQA measures using the LIVE database

	PLCC			MAE			SRCC			KRCC		
	PSNR	SSIM	$\hat{S}$	PSNR	SSIM	$\hat{S}$	PSNR	SSIM	$\hat{S}$	PSNR	SSIM	$\hat{S}$
JP2 (1)	0.9331	0.9687	0.9597	6.5033	4.7620	4.9860	0.9264	0.9637	0.9555	0.7600	0.8332	0.8140
JP2 (1)	0.8740	0.9691	0.9632	9.9656	5.2016	5.2320	0.8549	0.9604	0.9539	0.6640	0.8290	0.8163
JPG (1)	0.8866	0.9667	0.9449	8.6900	4.7096	5.6854	0.8779	0.9637	0.9493	0.7026	0.8364	0.8096
JPG (2)	0.9167	0.9851	0.9761	10.013	4.6077	5.7997	0.7699	0.9215	0.8979	0.5776	0.7774	0.7240
Noise	0.9879	0.9830	0.9773	3.4195	4.2499	4.8172	0.9854	0.9694	0.9642	0.8939	0.8523	0.8345
Blur	0.7840	0.9483	0.9154	9.0550	4.6651	7.5136	0.7823	0.9517	0.8692	0.5847	0.8010	0.7158
FF	0.8897	0.9552	0.9316	9.9898	6.1810	8.0113	0.8907	0.9556	0.9138	0.7069	0.8207	0.7473
All	0.8721	0.9449	0.9212	10.5248	6.9325	8.3641	0.8755	0.9479	0.9214	0.6864	0.7963	0.7561

followed by an inverse wavelet transform to construct the repaired image. An example is given in Fig. 4, where the blurred image is successfully repaired, and the effect is reflected by both SSIM and the proposed RR SSIM measures.

5. CONCLUSIONS

We propose an RR SSIM estimation algorithm by incorporating DNT-domain image statistical properties and the design principle of the SSIM approach. Our experiments show that the proposed SSIM estimation has good correlations with not only FR SSIM, but also subjective evaluations of image quality. We also demonstrate that the RR features being used can be employed to partially repair a distorted images. The proposed method has a fairly low RR data rate and is applicable to various types of distortions. It has good potentials to be employed in real-world visual communications systems for quality monitoring and resource allocation purposes. It may also be a useful tool in image quality optimization problems when the reference image is not fully available.

6. ACKNOWLEDGMENT

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada in the form of Discovery and Strategic Grants, and in part by Ontario Ministry of Research & Innovation in the form of an Early Researcher Award, which are gratefully acknowledged.

7. REFERENCES

[1] 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, 2004.

[2] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*, Morgan & Claypool Publishers, March 2006.

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

[4] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE Journal on Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 202–211, 2009.

[5] A. Albonico, G. Valenzise, M. Naccari, M. Tagliasacchi, and S. Tubaro, "A reduced-reference video structural similarity metric based on no-reference estimation of channel-induced distortion," in *IEEE Inter. Conf. Acoustics, Speech and Signal Processing*, 2009, pp. 1857–1860.

[6] Hamid R. Sheikh, Zhou Wang, Alan C. Bovik, and L. K. Cormack, "Image and video quality assessment research at LIVE," http://live.ece.utexas.edu/research/quality/.

[7] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of gaussians and the statistics of natural images," in *in Adv. Neural Info. Processing Systems*. 2000, pp. 855–861, MIT Press.

[8] S. M. Weiss and N. Indurkhya, "Rule-based machine learning methods for functional prediction," *Journal of Artificial Intelligence Research*, vol. 3, pp. 383–403, 1995.

[9] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Trans. Information Theory*, vol. 38, no. 2 pt II, pp. 587–607, 1992.

[10] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, 2009.