Reduced-Reference Image Quality Assessment Using A Wavelet-Domain Natural Image Statistic Model

Zhou Wang and Eero P. Simoncelli

Howard Hughes Medical Institute, Center for Neural Science and Courant Institute of Mathematical Sciences, New York University, New York, NY 10003, USA
zhouwang@ieee.org, eero.simoncelli@nyu.edu

ABSTRACT

Reduced-reference (RR) image quality measures aim to predict the visual quality of distorted images with only partial information about the reference images. In this paper, we propose an RR image quality assessment method based on a natural image statistic model in the wavelet transform domain. We use the Kullback-Leibler distance between the marginal probability distributions of wavelet coefficients of the reference and distorted images as a measure of image distortion. A generalized Gaussian model is employed to summarize the marginal distribution of wavelet coefficients of the reference image, so that only a relatively small number of RR features are needed for the evaluation of image quality. The proposed method is easy to implement and computationally efficient. In addition, we find that many well-known types of image distortions lead to significant changes in wavelet coefficient histograms, and thus are readily detectable by our measure. A Matlab implementation of the method has been made available online at `http://www.cns.nyu.edu/~lcv/rriqa/`.

1. INTRODUCTION

Objective image quality assessment models typically require the access to a reference image that is assumed to have perfect quality. In practice, such full-reference (FR) methods may not be applicable because the reference image is often not available. On the other hand, no-reference (NR) or "blind" image quality assessment is an extremely difficult task. Most proposed NR quality metrics are designed for one or a set of predefined specific distortion types[1–7] and are unlikely to generalize for evaluating images degraded with other types of distortions. In real-world applications, they are useful only when the types of distortions between the reference and distorted images are fixed and known. Although there is a strong need of NR quality assessment methods that are applicable to a wide variety of distortions, no such method, to the best of our knowledge, has been proposed and extensively tested.

Reduced-reference (RR) image quality metrics provide a solution that lies between FR and NR models. They are designed to predict the perceptual quality of distorted images with only partial information about the reference images. Recently, the video quality experts group[8] has included RR image/video quality assessment as one of its directions for future development. RR methods are useful in a number of applications. For example, in real-time visual communication systems, they can be used to track image quality degradations and control the streaming resources. Figure 1 shows how an RR quality analysis system may be deployed. The system includes a feature extraction process at the sender side and a feature extraction/quality analysis process at the receiver side. The extracted RR features usually have a much lower data rate than the image data and are typically transmitted to the receiver through an ancillary channel*.[9] Although it is often assumed that the ancillary channel is error-free, this is not an absolutely necessary requirement since even partly decoded RR features may still be helpful in evaluating the quality of the distorted image, though the accuracy may be severely affected.

A successful RR quality assessment method must achieve a good balance between the data rate of RR features and the accuracy of image quality prediction. On the one hand, with a high data rate, one can include a large amount of information about the reference image, leading to more accurate estimation of distorted image quality, but it becomes a heavy burden to transmit the RR features to the receiver. On the other hand, a lower data rate makes it easier to transmit the RR information, but harder for accurate quality estimation.

*Another choice is to send the RR features in the same channel as the images being transmitted. In that case, a better protection (e.g., through error control coding) of the RR features than the image data is usually needed.

Human Vision and Electronic Imaging X, edited by Bernice E. Rogowitz, Thrasyvoulos N. Pappas, Scott J. Daly, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 5666 © 2005 SPIE and IS&T · 0277-786X/05/$15

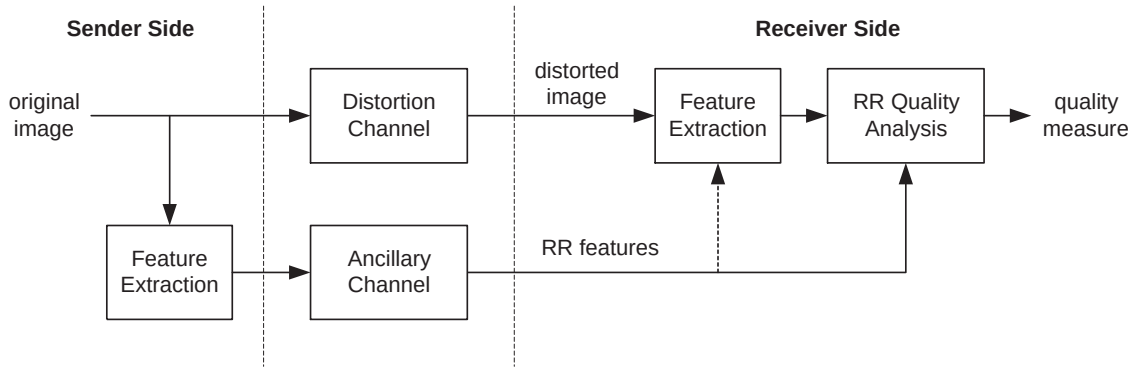

Figure 1. Deployment of reduced-reference image quality assessment system.

In this paper, we propose an RR quality assessment method based on a *natural image statistic* model in the *wavelet* transform domain. The basic assumption behind natural image statistics-based approaches is that most real-world image distortions disturb image statistics and make the distorted image "unnatural". The unnaturalness measured based on models of natural image statistics can then be used to quantify image quality degradation. In particular, here we observe that the marginal distribution of the wavelet coefficients within a given subband changes in different ways for different types of image distortions. We then use an information distance measure between probability distributions to quantify such changes and examine if this provides a useful quality measurement of images through comparisons with subjective image quality evaluations. Since no specific distortion model is assumed, the proposed method is potentially useful for a wide range of distortion types.

2. METHOD

2.1. Motivation

Wavelet transforms provide a convenient framework for localized representation of signals simultaneously in space and frequency. They have been widely used to model the processing in the early stages of biological visual systems and have also become the preferred form of representations for many image processing and computer vision algorithms. In recent years, natural image statistics have played an important role in the understanding of sensory neural behaviors of the human visual system.[10] In the image processing literature, statistical prior models of natural images have been employed as fundamental ingredients in many image coding and estimation algorithms.[11] They have also been used for image quality assessment purposes.[6,12]

Figure 2 shows the histograms of the coefficients computed from one of the wavelet subbands in a steerable pyramid decomposition[13] (a type of redundant wavelet transform that avoids aliasing in subbands). It has been pointed out that the marginal distributions of such oriented bandpass filter responses of natural images are highly kurtotic (with sharp peaks at zero and much longer tails than Gaussian density, as demonstrated in Fig. 2(a)) and have a number of important implications to sensory neural coding of natural visual scene.[14] It was demonstrated that many natural looking texture images can be synthesized by matching the histograms of the filter responses of a set of well-selected bandpass filters.[15,16] Psychophysical visual sensitivity to histogram changes of wavelet-textures had also been studied.[17] In Fig. 2, it can be seen that the marginal distribution of the wavelet coefficients changes in different ways for different types of image distortions. Such histogram changes in images contaminated with white Gaussian noise have been observed previously and used for image denoising.[18–21]

2.2. Distortion Measure

Let $p(x)$ and $q(x)$ denote the probability density functions of the wavelet coefficients in the same subband of two images, respectively. Let $\mathbf{x} = \{x_1, ..., x_N\}$ be a set of N randomly and independently selected coefficients. The

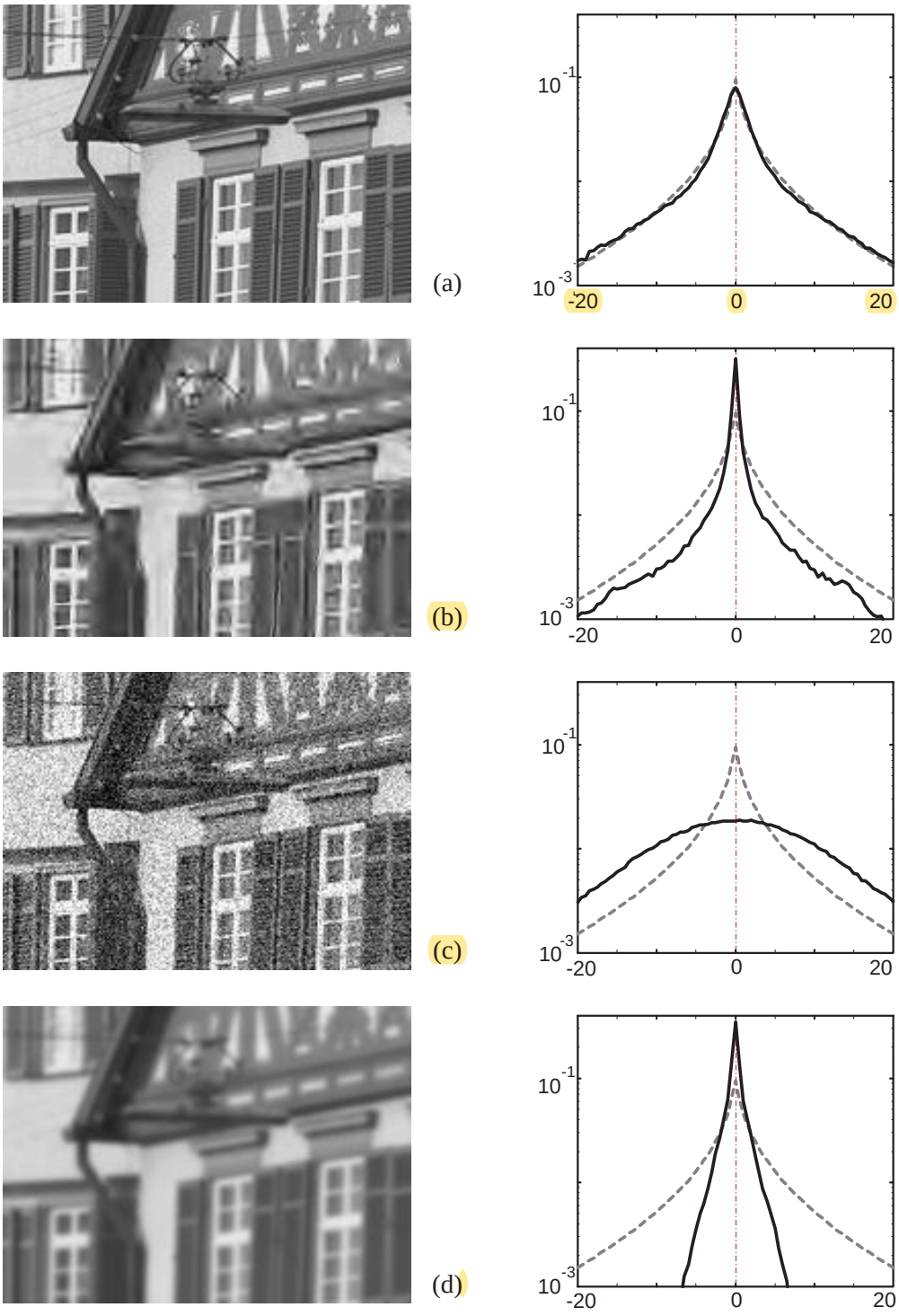

Figure 2. Comparisons of wavelet coefficient histograms (solid curves) calculated from the same horizontal subband in the steerable pyramid decomposition.[13] (a) original (reference) "buildings" image (cropped for visibility); (b) JPEG2000 compressed image; (c) white Gaussian noise contaminated image; (d) Gaussian blurred image. The histogram of the original image coefficients can be well fitted with a generalized Gaussian density model (dashed curves).

log-likelihoods of $\mathbf{x}$ being drawn from $p(x)$ and $q(x)$ are

$$l(p|\mathbf{x}) = \frac{1}{N}\sum_{n=1}^{N} \log\, p(x_n) \quad \text{and} \quad l(q|\mathbf{x}) = \frac{1}{N}\sum_{n=1}^{N} \log\, q(x_n)\,, \tag{1}$$

respectively. The difference of the log-likelihood (or equivalently, the log-likelihood-ratio) between $p(x)$ and $q(x)$ is

$$l(p|\mathbf{x}) - l(q|\mathbf{x}) = \frac{1}{N}\sum_{n=1}^{N} \log \frac{p(x_n)}{q(x_n)}\,. \tag{2}$$

Now assume that $p(x)$ is the true probability density distribution of the coefficients. Based on the law of large numbers, when N is large, the difference of the log-likelihoods (or equivalently, the log-likelihood-ratio) between $p(x)$ and $q(x)$ asymptotically approaches the Kullback-Leibler distance[22] (KLD) between $p(x)$ and $q(x)$:

$$l(p|\mathbf{x}) - l(q|\mathbf{x}) \longrightarrow d(p\|q) = \int p(x) \log \frac{p(x)}{q(x)} dx\,. \tag{3}$$

In previous work, a number of authors have pointed out the relationship between KLD and log-likelihood function and used KLD to compare images, mainly for classification and retrieval purposes.[23–26] KLD has also been used to quantify the distributions of image pixel intensity values for the evaluation of compressed image quality.[27, 28] In this paper, we use KLD to quantify the difference of wavelet coefficient distributions between a distorted image and a perfect quality reference image. We then examine how this quantity correlates with perceptual image quality for a wide range of distortion types.

Assume that $p(x)$ and $q(x)$ are associated with the reference and distorted images, respectively. To estimate the KLD between them, the coefficient histograms of both the reference and distorted images must be available. The latter can be easily computed from the received distorted image. The difficulty is in obtaining the coefficient histogram of the reference image at the receiver side. On the one hand, if the histogram bin size is small, the bandwidth required to transmit the RR features is very demanding. On the other hand, if the histogram bin size is large, the estimation accuracy of KLD would be significantly affected. In any case, depending on the application environment, transmitting all the histogram bins as RR features may not be a realistic choice.

One important discovery in the literature of natural image statistics is that the marginal distribution of the coefficients in individual wavelet subbands can be well-fitted with a 2-parameter generalized Gaussian density (GGD) model[18, 20, 29, 30]:

$$p_m(x) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x|/\alpha)^{\beta}}, \tag{4}$$

where $\Gamma(a) = \int_0^{\infty} t^{a-1}e^{-t}dt$ (for $a > 0$) is the Gamma function. This model provides a very efficient means to summarize the coefficient histogram of the reference image, so that only two model parameters $\{\alpha,\ \beta\}$ need to be transmitted to the receiver. Such an efficient representation of wavelet coefficient histograms has been used in previous work for image compression[30] and texture image retrieval.[26] In addition to the two model parameters, here we also include the prediction error as a third RR feature parameter, which is defined as the KLD between $p_m(x)$ and $p(x)$:

$$d(p_m\|p) = \int p_m(x) \log \frac{p_m(x)}{p(x)} dx. \tag{5}$$

In practice, this quantity has to be evaluated numerically using histograms:

$$d(p_m\|p) = \sum_{i=1}^{L} P_m(i) \log \frac{P_m(i)}{P(i)}\,, \tag{6}$$

where $P(i)$ and $P_m(i)$ are the normalized heights of the i-th histogram bins, and L is the number of bins in the histograms.

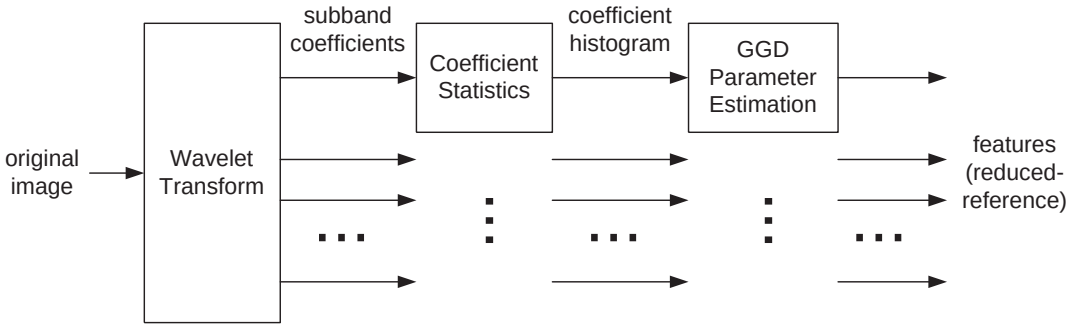

Figure 3. Feature extraction system at the sender side.

At the receiver side, we first compute the KLD between $p_m(x)$ and $q(x)$:

$$d(p_m\|q) = \int p_m(x) \log \frac{p_m(x)}{q(x)} dx. \tag{7}$$

Different from the sender side, here we do not fit $q(x)$ with a GGD model (If we do, then the KLD between the two GGD distributions can be calculated analytically[26]). The main reason is that a distorted image may not be a "natural" image anymore, and thus may not be well-fitted with a GGD model. Therefore, we extract the histogram bins of the wavelet coefficients as the features at the receiver side and use them to numerically (similar to Eq. (6)) evaluate $d(p_m\|q)$. The KLD between $p(x)$ and $q(x)$ are then estimated as

$$\hat{d}(p\|q) = d(p_m\|q) - d(p_m\|p)\,. \tag{8}$$

It can be easily shown that

$$\hat{d}(p\|q) = \int p_m(x) \log \frac{p_(x)}{q(x)} dx\,. \tag{9}$$

The estimation error is:

$$\begin{aligned} & d(p\|q) - \hat{d}(p\|q) \\ = \; & d(p\|q) - [\, d(p_m\|q) - d(p_m\|p)\,] \\ = \; & \int [p(x) - p_m(x)] \log \frac{p_(x)}{q(x)} dx\,. \end{aligned} \tag{10}$$

This error is small when $p_m(x)$ and $p(x)$ are close, which is true for typical natural images, as demonstrated in Fig. 2(a). With the additional cost of sending one more parameter $d(p_m\|p)$, Eq. (9) not only gives a more accurate estimator of $d(p\|q)$ than Eq. (7), but also provides a useful feature that when there is no distortion between the original and received images (which implies $p(x) = q(x)$ for all x), both the targeted distortion measure $d(p\|q)$ and estimated distortion measure $\hat{d}(p\|q)$ are exactly zero. Note that this is in general not true for Eq. (7).

Finally, the overall distortion between the distorted and reference images is defined as:

$$D = \log_2 \left(1 + \frac{1}{D_0} \sum_{k=1}^{K} |\hat{d}^k(p^k\|q^k)| \right), \tag{11}$$

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

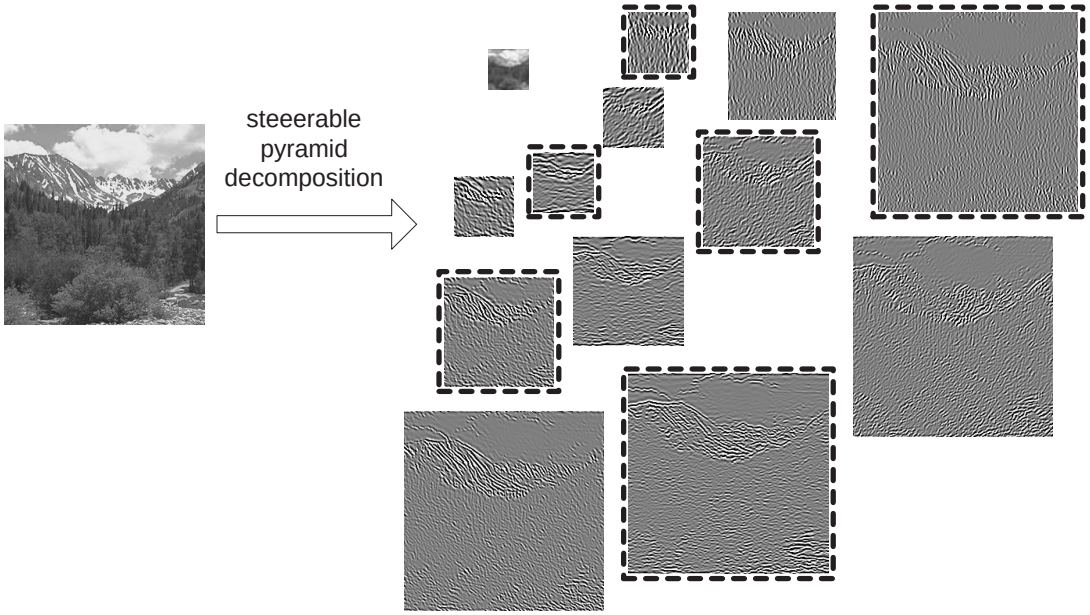

Figure 4. Steerable pyramid decomposition[13] of image (highpass residual band not shown). A set of selected subbands (marked with dashed boxes) are used for GGD feature extraction.

2.3. Implementation

The feature extraction system for the reference image at the sender side is illustrated in Fig. 3. We first apply a 3-scale, 4-orientation steerable pyramid wavelet transform[13] to decompose the image into 12 oriented subbands (4 for each scale) and a highpass and a lowpass residual subbands. Six (2 for each scale) of the 12 oriented subbands are selected for feature extraction, as demonstrated in Fig. 4. The major purpose of selecting only a subset of all the subbands is to reduce the data rate of RR features†. For each selected subband, the histogram of the coefficients is computed and its feature parameters $\{\alpha,\ \beta,\ d(p_m\|p)\}$ are then estimated using a gradient descent algorithm to minimize the KLD between $p(x)$ and $p_m(x)$. This results in a total of 18 extracted scalar features for each reference image that can be used for RR image quality assessment. These RR features are further quantized to finite precision. Specifically, both β and $d(p_m\|p)$ are quantized into 8-bit precision, and α is represented using 11-bit floating point, with 8 bits for mantissa and 3 bits for exponent. In summary, a total of $(8+8+8+3)\times 6 = 162$ bits are used to represent the RR features.

Figure 5 shows the reduced-reference quality analysis system at the receiver side. The same wavelet transform as the sender side is first applied to the distorted image, and the coefficient histograms of the corresponding subbands are computed. To numerically evaluate $d(p_m\|q)$ at each subband, the subband histogram is compared with the histogram calculated from the corresponding RR features $\{\alpha,\ \beta\}$ about the reference image. The third RR feature, $d(p_m\|p)$, is then subtracted from this quantity (as of Eq. 8) to estimate $d(p\|q)$. Finally, at the distortion pooling stage, the KLDs evaluated at all subbands are combined using Eq. (11) to provide a single distortion measure.

3. TEST

We use the LIVE database[31] to evaluate the performance of the proposed method. The database contains 29 high-resolution (typically 768×512) original images altered with five types of distortions at different distortion levels. The distorted images were divided into seven datasets. Datasets 1 (87 images) and 2 (82 images) are JPEG2000 compressed images; Datasets 3 (87 images) and 4 (88 images) are JPEG compressed images; and

†In our current test, selecting the other six oriented subbands or all the 12 oriented subbands gives similar overall performance for image quality prediction.

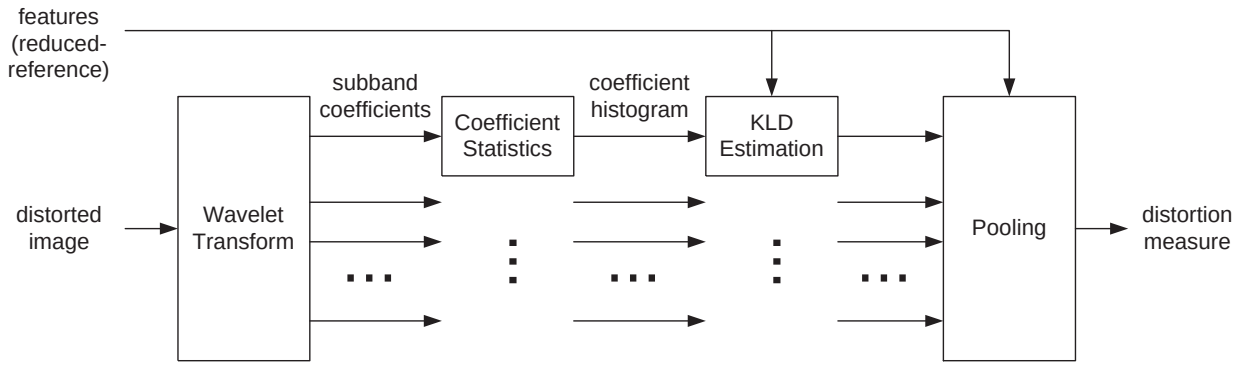

Figure 5. Quality analysis system at the receiver side.

Datasets 5, 6 and 7 (each containing 145 images) are distorted with white Gaussian noise, Gaussian blur, and transmission errors in the JPEG2000 bitstream using a fast-fading Rayleigh channel model, respectively. Subjects were asked to provide their perception of quality on a continuous linear scale that was divided into five equal regions marked with "Bad", "Poor", "Fair", "Good" and "Excellent", respectively. Each image was rated by 20-25 subjects. The raw scores for each subject were converted into Z-scores and rescaled within each dataset to fill the range from 1 to 100. Mean opinion score (MOS) and the standard deviation between subjective scores were then computed for each image.

Three measures are computed to quantify the performance of the proposed quality assessment method. First, following the suggestions given in the video quality experts group Phase I FR-TV test, we use a 4-parameter logistic function to provide a nonlinear mapping between the objective and subjective scores.[32] Figure 6 shows the scatter plots (in which each data point represents one test image) of true mean opinion score (MOS) versus the predicted score by the proposed method after the nonlinear mapping. After such a nonlinear mapping, the correlation coefficient between the predicted and true subjective scores is calculated to evaluate *prediction accuracy*. Second, the Spearman rank-order correlation coefficient is computed to evaluate *prediction monotonicity*. Finally, to evaluate *prediction consistency*, the outlier ratio is used, which is defined as the percentage of predictions outside the range of ± 2 standard deviations between subjective scores.

To the best of our knowledge, no other RR quality assessment method has been proposed that (1) aims for general-purpose image quality assessment (as opposed to distortion- or application-specific), and (2) uses such small amount of information about the reference image as compared to the proposed method. Therefore, we choose a set of full-reference quality assessment models for comparison, which include peak signal-to-noise-ratio (PSNR), Lubin's Sarnoff model,[33–35] and the structural similarity (SSIM) index.[36, 37] Although such comparison is highly unfair to the proposed method (because all other methods require full access to the reference images, while the proposed method uses only 18 scalar features), it provides a useful indication about the relative performance of the proposed method. Specifically, Lubin's Sarnoff model and the SSIM index give roughly the performance of state-of-the-art full-reference quality assessment methods.

The performance evaluation results are shown in Table 1. It can be seen that the proposed method performs quite well for a wide range of distortion types. Specifically, for five of the seven datasets, the proposed method provides better prediction accuracy (higher correlation coefficients), better prediction monotonicity (higher Spearman rank-order correlation coefficients) and better prediction consistency (lower outlier ratios) than PSNR, which is the most widely used full-reference image quality measure in the image processing literature. Therefore, we believe the proposed method is a reasonable and useful choice in practical RR quality assessment systems. It needs to be emphasized that none of the other methods being compared can be used in this scenario.

4. DISCUSSION

We propose a wavelet domain information distance measure for RR image quality assessment. Experiments with a large image database demonstrate the effectiveness of the proposed method. Several properties of the proposed

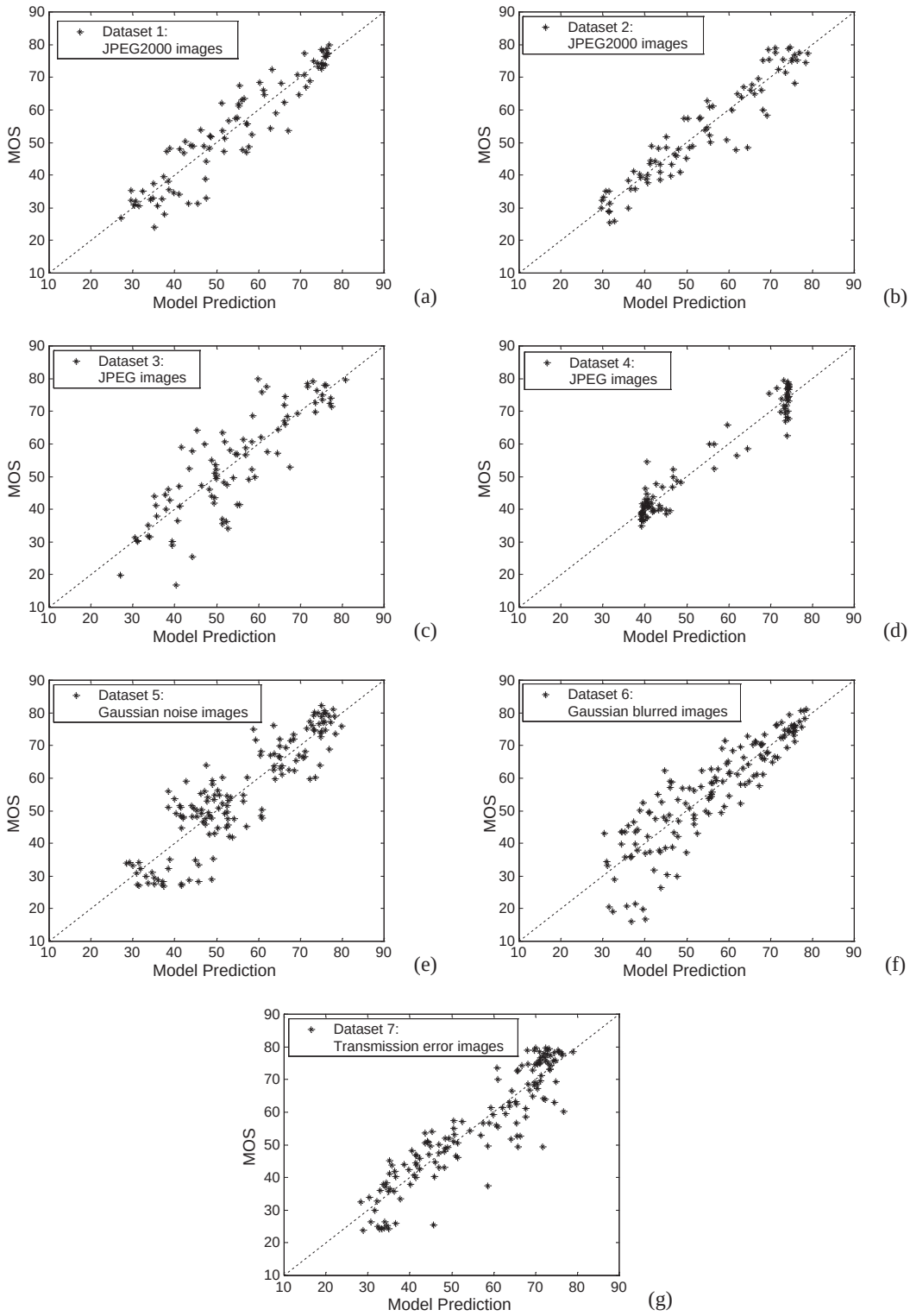

Figure 6. Scatter plots of mean opinion score (MOS) versus model prediction for the seven image datasets in the LIVE database. Each data point represents one test image in the database.

Table 1. Performance evaluation of image quality measures using the LIVE database. JP2: JPEG2000 dataset; JPG: JPEG dataset; Noise: white Gaussian noise dataset; Blur: Gaussian blur dataset; Error: transmission error dataset.

Dataset	JP2(1)	JP2(2)	JPG(1)	JPG(2)	Noise	Blur	Error
# of images	87	82	87	88	145	145	145
	Correlation Coefficient (prediction accuracy)						
Proposed	0.9353	0.9490	0.8452	0.9695	0.8889	0.8872	0.9175
PSNR	0.9337	0.8948	0.9015	0.9136	0.9866	0.7742	0.8811
Sarnoff	0.9706	0.9650	0.9589	0.9837	0.9631	0.9480	0.9144
MSSIM	0.9675	0.9668	0.9646	0.9856	0.9706	0.9361	0.9439
	Rank-Order Correlation Coefficient (prediction monotonicity)						
Proposed	0.9298	0.9470	0.8332	0.8908	0.8639	0.9145	0.9162
PSNR	0.9231	0.8816	0.8907	0.8077	0.9855	0.7729	0.8785
Sarnoff	0.9668	0.9565	0.9528	0.8904	0.9411	0.9381	0.9048
MSSIM	0.9566	0.9677	0.9572	0.9441	0.9718	0.9421	0.9497
	Outlier Ratio (prediction consistency)						
Proposed	0.0690	0.0366	0.1839	0.0341	0.1793	0.1172	0.0621
PSNR	0.0805	0.0976	0.0920	0.1818	0.0000	0.2069	0.1517
Sarnoff	0.0000	0.0366	0.0115	0.0000	0.0345	0.0276	0.0552
MSSIM	0.0000	0.0000	0.0000	0.0114	0.0000	0.0414	0.0345

method may be of interest for real-world users. First, it is a general-purpose method that performs well for a wide range of distortion types, as opposed to application-specific methods that are designed or trained for some specific types of distortions (e.g., block-DCT and wavelet-based image/video compression). Second, it has a relatively low data rate for representing RR features (currently only 18 features/image, or 162 bits/image after quantization). Third, the method is easy to implement, computationally efficient, and only uses a few parameters. Fourth, since the measurement is based on marginal distributions of wavelet coefficients, the method is insensitive to small geometric distortions such as spatial translation, rotation and scaling.

Although it has been widely noted that the marginal distributions of oriented bandpass filter responses of natural images have important connections with human visual perception of images,[14–16] their relationship with subjective image quality has not been tested. The experimental results presented in this paper represent a simple initial step toward confirmation of such a connection. In the future, the method may be further improved by incorporating joint statistics of wavelet coefficients (e.g.[11, 21, 38–41]), which are much more powerful in characterizing the statistical structures of natural images.

REFERENCES

1. H. R. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," *IEEE Signal Processing Letters* **4**, pp. 317–320, Nov. 1997.
2. Z. Wang, A. C. Bovik, and B. L. Evans, "Blind measurement of blocking artifacts in images," in *Proc. IEEE Int. Conf. Image Proc.*, **3**, pp. 981–984, Sept. 2000.
3. Z. Yu, H. R. Wu, S. Winkler, and T. Chen, "Vision-model-based impairment metric to evaluate blocking artifact in digital video," *Proceedings of the IEEE* **90**, pp. 154–169, Jan. 2002.
4. Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proc. IEEE Int. Conf. Image Proc.*, (Rochester), Sept. 2002.
5. X. Li, "Blind image quality assessment," in *Proc. IEEE Int. Conf. Image Proc.*, **1**, pp. 449–452, Sept. 2002.
6. H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Processing* , 2004. to appear.
7. P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Application to JPEG2000," *Signal Processing: Image Communication* **19**, pp. 163–172, Feb. 2004.

8. VQEG: The Video Quality Experts Group, `http://www.vqeg.org/`.
9. Z. Wang, H. R. Sheikh, and A. C. Bovik, "Objective video quality assessment," in *The Handbook of Video Databases: Design and Applications*, B. Furht and O. Marques, eds., pp. 1041–1078, CRC Press, Sept. 2003.
10. E. P. Simoncelli and B. Olshausen, "Natural image statistics and neural representation," *Annual Review of Neuroscience* **24**, pp. 1193–1216, May 2001.
11. E. P. Simoncelli, "Statistical models for images: Compression, restoration and synthesis," in *Proc 31st Asilomar Conf on Signals, Systems and Computers*, pp. 673–678, IEEE Computer Society, (Pacific Grove, CA), November 1997. Available from http://www.cns.nyu.edu/∼eero/publications.html.
12. H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Processing* , 2004. to appear.
13. E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multi-scale transforms," *IEEE Trans. Information Theory* **38**, pp. 587–607, 1992.
14. D. J. Field, "What is the goal of sensory coding?," *Neural Computation* **6**(4), pp. 559–601, 1994.
15. D. Heeger and J. Bergen, "Pyramid-based texture analysis/synthesis," in *Proc. ACM SIGGRAPH*, pp. 229–238, Association for Computing Machinery, August 1995.
16. S. C. Zhu, Y. N. Wu, and D. Mumford, "FRAME: Filters, random fields and maximum entropy – towards a unified theory for texture modeling," *Intl. J. Comp. Vis.* **27**(2), pp. 1–20, 1998.
17. F. A. A. Kingdom, A. Hayes, and D. J. Field, "Sensitivity to contrast histogram differences in synthetic wavelet-textures," *Vision Research* **41**(5), pp. 585–598, 2001.
18. E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *Third Int'l Conf on Image Proc*, **I**, pp. 379–382, IEEE Sig Proc Society, (Lausanne), September 1996.
19. E. P. Simoncelli, "Bayesian denoising of visual images in the wavelet domain," in *Bayesian Inference in Wavelet Based Models*, P. Müller and B. Vidakovic, eds., ch. 18, pp. 291–308, Springer-Verlag, New York, 1999. Lecture Notes in Statistics, vol. 141.
20. P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using a generalized Gaussian and complexity priors," *IEEE Trans. Info. Theory* **45**, pp. 909–919, 1999.
21. A. Hyvärinen, P. O. Hoyer, and E. Oja, "Sparse code shrinkage: Denoising by nonlinear maximum likelihood estimation," in *Advances in Neural Information Processing Systems 11 (NIPS'98)*, pp. 473–479, MIT Press, May 1999.
22. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, New York, 1991.
23. J.-Y. Chen, C. A. Bouman, and J. P. Allebach, "Multiscale branch and bound image database search," in *SPIE/IS&T Conf. Storage and Retrieval for Image and Video Databases V, Proc. SPIE*, **3022**, pp. 133–144, Feb. 1997.
24. J. De Bonet and P. Viola, "Texture recognition using a non parametric multi-scale statistical model," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, pp. 641–647, June 1998.
25. N. Vasconcelos and A. Lippman, "A probabilistic architecture for content-based image retrieval," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, pp. 216–221, June 2000.
26. M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized gaussian density and Kullback-Leibler distance," *IEEE Trans. Image Proc.* **11**, pp. 146–158, Feb. 2002.
27. İ. Avcibaş, B. Sankur, and K. Sayood, "Statistical evaluation of image quality measures," *Journal of Electronic Imaging* **11**, pp. 206–223, Apr. 2002.
28. J. A. Garcia, J. Fdez-Valdivia, R. Rodriguez-Sánchez, and X. R. Fdez-Vidal, "Performance of the Kullback-Leibler information gain for predicting image fidelity," in *Proc. IEEE Int. Conf. Pattern Recognition*, **III**, pp. 843–848, 2002.
29. S. G. Mallat, "Multifrequency channel decomposition of images and wavelet models," *IEEE Trans. Acoustics, Speech, and Signal Processing* **37**, pp. 2091–2110, Dec. 1989.
30. R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Trans Image Proc* **8**, pp. 1688–1701, December 1999.
31. H. R. Sheikh, Z. Wang, A. C. Bovik, and L. K. Cormack, "Image and video quality assessment research at LIVE," http://live.ece.utexas.edu/research/quality/.

32. VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," Apr. 2000. http://www.vqeg.org/.
33. J. Lubin, "The use of psychophysical data and models in the analysis of display system performance," in *Digital images and human vision*, A. B. Watson, ed., pp. 163–178, The MIT Press, Cambridge, Massachusetts, 1993.
34. J. Lubin, "A visual discrimination mode for image system design and evaluation," in *Visual Models for Target Detection and Recognition*, E. Peli, ed., pp. 207–220, World Scientific Publishers, Singapore, 1995.
35. Sarnoff Corporation, "JNDmetrix Technology," http://www.sarnoff.com/products_services/video_vision/jndmetrix/.
36. 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* **13**, pp. 600–612, Apr. 2004.
37. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "The SSIM index for image quality assessment," http://www.cns.nyu.edu/∼lcv/ssim/.
38. M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Proc.* **46**, pp. 886–902, April 1998.
39. A. Srivastava, X. Liu, and U. Grenander, "Universal analytical forms for modeling image probability," *IEEE Pat. Anal. Mach. Intell.* **28**(9), 2002.
40. J. Portilla, V. Strela, M. Wainwright, and E. P. Simoncelli, "Image denoising using a scale mixture of Gaussians in the wavelet domain," *IEEE Trans Image Processing* **12**, November 2003.
41. Z. Wang and E. P. Simoncelli, "Local phase coherence and the perception of blur," in *Adv. Neural Information Processing Systems (NIPS03)*, **16**, MIT Press, (Cambridge, MA), May 2004.