Correspondences

Reduced-Reference Image Quality Assessment Using Reorganized DCT-Based Image Representation

Lin Ma, *Student Member, IEEE*, Songnan Li, *Student Member, IEEE*, Fan Zhang, and King Ngi Ngan, *Fellow, IEEE*

***Abstract*—In this paper, a novel reduced-reference (RR) image quality assessment (IQA) is proposed by statistical modeling of the discrete cosine transform (DCT) coefficient distributions. In order to reduce the RR data rates and further exploit the identical nature of the coefficient distributions between adjacent DCT subbands, the DCT coefficients are reorganized into a three-level coefficient tree. Subsequently, generalized Gaussian density (GGD) is employed to model the coefficient distribution of each reorganized DCT subband. The city-block distance is employed to measure the difference between the two images. Experimental results demonstrate that only a small number of RR features is sufficient for representing the image perceptual quality. The proposed method outperforms the RR WNISM and even the full-reference (FR) quality metric PSNR.**

***Index Terms*—City-block distance, generalized Gaussian density (GGD), human visual system (HVS), image quality assessment (IQA), reduced-reference (RR).**

I. Introduction

The objective of the image quality assessment (IQA) is to provide computational models to measure the perceptual quality of an image, which plays a very important role in many image processing tasks, such as image coding and transmission. The subjective testing [1] is a straightforward way for assessing the image quality. However, it is time-consuming and expensive. Therefore, objective methods [2] that can automatically evaluate the image perceptual quality are desired.

Depending on the availability of a reference image, the IQA methods [3]–[12] can be divided into three categories [2]: full-reference (FR), no-reference (NR), and reduced-reference (RR). FR IQAs require the whole reference image for evaluating the perceptual quality of the distorted image, such as mean squared error (MSE), peak signal-to-noise ratio (PSNR), and structure similarity (SSIM) [3]. However, in practical scenarios where the reference image is not always available, NR IQA is required, which is an extremely difficult task. Therefore, most proposed NR IQAs [6]–[8] attempt to evaluate the perceptual qualities of the images inferred by a set of specific distortion types, which means that they can utilize the prior knowledge of the distortions. However, if the partial information of the reference image is available, then RR methods can be designed to evaluate the image quality based on the partial information. In [9], objective features are extracted from both the reference and distorted frames and a convolutional neural network is employed for video quality assessment. In [12], the multi-scale geometry analysis, contrast sensitivity function, and the Weber's law of just noticeable difference is incorporated to derive the IQA. Also Tao *et al.* proposed an RR IQA [27] by employing the city-block distance to measure the quantity differences of the visual sensitive coefficients in the contourlet domain. Furthermore, Wang [11] proposed a wavelet-domain natural image statistic metric (WNISM) based on the natural image statistic, which models the marginal probability distribution of the wavelet coefficients of a natural image by the generalized Gaussian density (GGD) function. WNISM employs the Kullback-Leibler distance (KLD) to depict the distorted image perceptual quality. WNISM has been developed to provide the so-called quality-aware images [10]. Although WNISM has been well recognized, our investigation has revealed some important limitations. Firstly, the relationships between the coefficients of neighboring wavelet subbands are not considered [12]. Secondly, although the method performs quite well on individual distortion types, its performance degrades significantly when images with different distortion types are tested together [28]. Thirdly, the computational cost for the KLD between the two GGDs is high [13] and its performance is not good enough for practical applications, as will be shown later in this paper. Moreover, the KLD is asymmetric [17], which means that $\mathrm{KLD}(a,b) \neq \mathrm{KLD}(b,a)$. Therefore, it is not suitable for image quality assessment, because the visual quality distance from one image to another should be identical no matter how it is measured.

In this paper, we develop a new RR IQA based on the mathematical analysis of DCT coefficient distributions [14]. Specifically, the coefficients of block-based DCT are firstly reorganized into several representative subbands [15], [16]. Subsequently, the GGD is employed to model the coefficient distribution of each reorganized DCT subband. Finally, the city-block distance, which is much simpler and more effective compared with KLD, is employed to measure the image perceptual quality.

II. Proposed Reduced-Reference Image Quality Assessment Method

The block diagram of the proposed RR IQA is illustrated in Fig. 1. In the sender side, the DCT reorganization is first applied to the reference image. Subsequently, the GGD is employed to model the coefficient distribution of each reorganized DCT subband. In the receiver side, the same DCT reorganization is applied to the distorted image. Then the histogram of each reorganized DCT subband is built. By referring to the extracted GGD features of the reference image, the city-block distance is employed to depict the histogram distance between the corresponding reorganized DCT subbands. Finally, the visual quality index of the image is obtained by pooling the distances together.

A. Reorganization Strategy of DCT Coefficients

As natural images can be viewed as smooth regions delimited by edge discontinuities, after block-based DCT, the image energy of smooth regions is compacted into the DC coefficients, and some high-frequency AC coefficients. For edges, only a small number of high-frequency AC coefficients contribute to its energy. The DCT coefficients of the image are reorganized [15], [16] according to Fig. 2. Firstly, the coefficients of each 8×8 DCT are decomposed into ten subbands. Then the coefficients of the same subband from different DCT blocks are grouped together and organized according to their

Manuscript received July 07, 2010; revised October 21, 2010 and January 13, 2011; accepted January 22, 2011. Date of publication January 31, 2011; date of current version July 20, 2011. This work was supported in part by a grant from the Chinese University of Hong Kong under the Focused Investment Scheme (Project 1903003). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Andrea Cavallaro.

The authors are with the Department of Electronic Engineering, the Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: lma@ee.cuhk.edu.hk; snli@ee.cuhk.edu.hk; fzhang@ee.cuhk.edu.hk; knngan@ee.cuhk.edu.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2011.2109701

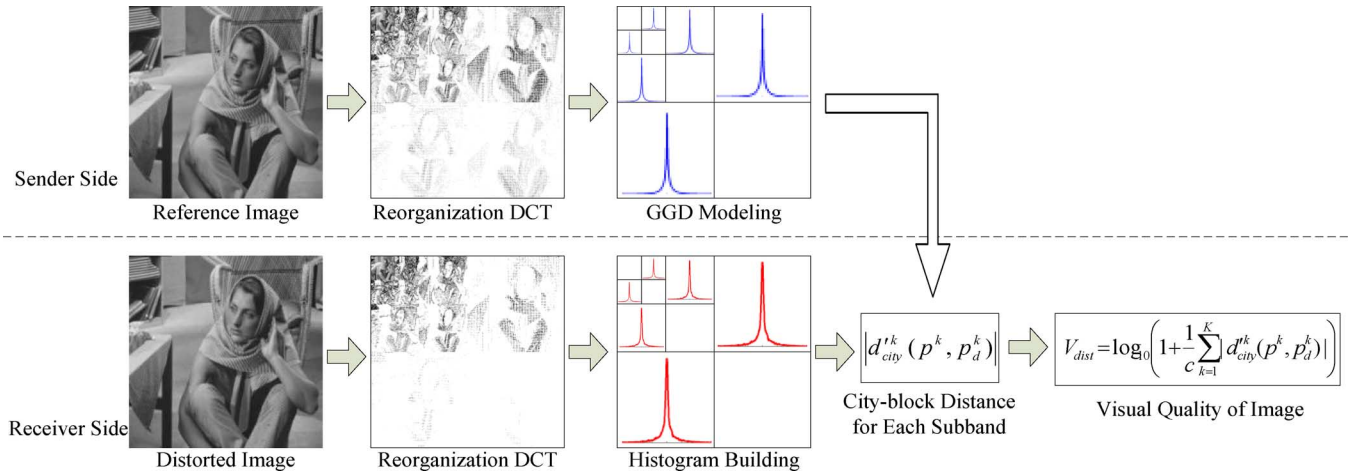

Fig. 1. Block diagram of the proposed RR IQA.

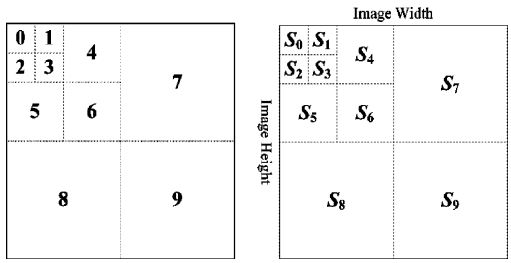

Fig. 2. Reorganization strategy of DCT coefficients (Left: one 8×8 DCT block with ten subband decomposition; right: the reorganized DCT image representation taken as three-level coefficient tree).

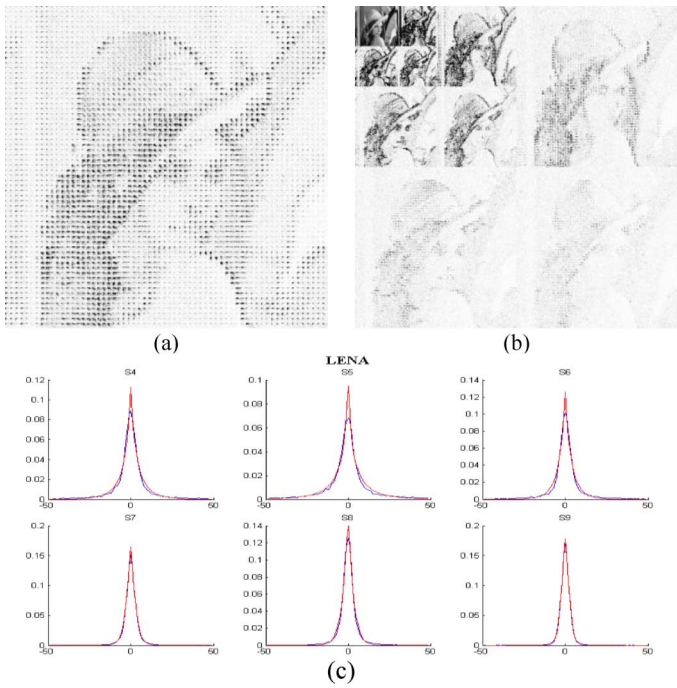

Fig. 3. Reorganization result of LENA image. (a) 8×8 DCT image representation. (b) Reorganized DCT image representation. (c) DCT coefficient distribution (blue line) and the fitted GGD curve (red line) of the reorganized subbands from S_4 to S_9.

corresponding positions. In Fig. 2, S_n denotes the grouped subband of all the DCT coefficients lying on the positions denoted by n.

Fig. 3 gives an example of the reorganization result of the LENA image. The 8×8 DCT image representation is shown in Fig. 3(a), which is obtained by applying the 8×8 DCT on the original LENA image. With the reorganization strategy, the reorganized DCT image representation is obtained, which is shown in Fig. 3(b). For better visualization, the DC components are rescaled to integers between 0 and 255, while the AC components are obtained by $255 - (5 \times |\mathrm{AC}|)$. It can be observed that the reorganized DCT image representation appears like a wavelet image representation, i.e., there are structural similarities between the subbands and the magnitude decays towards the high-frequency subbands.

B. GGD Modeling of DCT Coefficient Distributions

It has been claimed [10], [11] that the wavelet coefficient distributions of natural images are highly kurtotic (with a sharp peak at zero and a fat-tail distribution). Based on a strict mathematical analysis, Lam *et al.* [14] pointed out that the high-frequency DCT coefficients also follow the kurtotic distribution, which a GGD usually fits well. The probability density function of GGD is defined as

$$p_{\alpha,\beta}(x) = \frac{\alpha}{2\beta\,\Gamma\left(\frac{1}{\alpha}\right)} \exp\left\{-\left(\frac{|x|}{\beta}\right)^{\alpha}\right\} \tag{1}$$

where $\alpha > 0$ and β are two parameters of the GGD function. Γ is the Gamma function given by

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt. \tag{2}$$

Therefore, we can see that only two parameters are needed to completely define each GGD model. However, if the coefficient distributions of all the DCT subbands are to be modeled, too many parameters are needed for the RR IQA. Considering the 8×8 DCT as an example, if all the AC subbands are to be depicted, there are at least $63 \times 2 = 126$ parameters. It is too large and conflicts with the purpose of RR IQA, which requires less reference information for the quality assessment. In order to reduce the RR data rate and further utilize the identical nature of the coefficient distribution between adjacent DCT subbands, the aforementioned reorganization strategy is employed to group the DCT coefficients into fewer representative subbands.

After the reorganization process, the number of AC subbands is reduced to 9, which is more reasonable for the RR IQA. The GGD model is employed to model the coefficient distribution of each reorganized

subband. The DCT coefficient distribution (blue line) and the fitted GGD curve (red line) of the reorganized subbands S_4–S_9 from the LENA image are illustrated in Fig. 3(c). It can be observed that the two curves overlap with each other, which means that the GGD model can efficiently depict the coefficient distributions of the reorganized DCT subbands. Applying this process, we not only model the DCT coefficient distribution, but also exploit the identical nature of the coefficient distributions between adjacent DCT subbands, which will help improve the RR IQA performance.

C. Reduced-Reference Image Quality Assessment

It has been shown that the GGD model provides an efficient way to represent the coefficient histogram for each reorganized DCT subband of the reference image. Therefore, for each GGD model, two parameters $\{\alpha, \beta\}$ are needed for the RR IQA. In order to further improve the GGD modeling precision, another parameter denoted as the prediction error is introduced. As we have discussed before, the KLD is asymmetric, which is not suitable for measuring the visual quality distance between the two images. Therefore, in this paper, the city-block distance between two distributions r and q is proposed to depict their differences:

$$d_{city}(r,q) = \sum_{i=1}^{L} |r(i) - q(i)| \tag{3}$$

where L denotes the number of the histogram bins. Therefore, the city-block distance between the fitted GGD distribution $p_{\alpha,\beta}$ and the actual distribution p of each reorganized DCT subband can be obtained by $d_{city}(p_{\alpha,\beta}, p)$, which is introduced as the third parameter to denote the prediction error. From the definition, it can be observed that d_{city} is symmetric, which means that $d_{city}(r,q) = d_{city}(q,r)$. Therefore, the visual distance between two images is identical, which is reasonable for assessing the image quality. Now for each reorganized DCT subband, three parameters $\{\alpha, \beta, d_{city}(p_{\alpha,\beta}, p)\}$ need to be recorded and transmitted to the receiver side for the quality assessment. These parameters can be quantized into finite precision while maintaining a reasonable approximation. Same as [10], both α and $d_{city}(p_{\alpha,\beta}, p)$ are quantized into 8-bit precision, and β is represented using 11-bit floating point, with 8 bits for mantissa and 3 bits for exponent. The quantization steps are set uniformly to represent the GGD parameters in a limited number of bits. Therefore, for each subband, $8+8+8+3 = 27$ bits are required to represent the extracted three GGD parameters.

In the receiver side, for each distorted image, the aim is to compute the city-block distance between the coefficient distributions of the original image p and the distorted image p_d:

$$d_{city}(p, p_d) = \sum_{i=1}^{L} |p(i) - p_d(i)|. \tag{4}$$

However, the coefficient distributions of the original image are unavailable. Therefore, we employ the fitted GGD model and the prediction error to approximate the city-block distance between p and p_d. The inequality property

$$\begin{aligned} &\sum_{i=1}^{L} |p_{\alpha,\beta}(i) - p_d(i)| - \sum_{i=1}^{L} |p(i) - p_{\alpha,\beta}(i)| \\ &\quad \le \sum_{i=1}^{L} |p(i) - p_d(i)| \\ &\quad \le \sum_{i=1}^{L} |p_{\alpha,\beta}(i) - p_d(i)| + \sum_{i=1}^{L} |p(i) - p_{\alpha,\beta}(i)| \end{aligned} \tag{5}$$

implies that $d_{city}(p, p_d)$ is bounded by

$$\begin{aligned} &d_{city}(p_{\alpha,\beta}, p_d) - d_{city}(p_{\alpha,\beta}, p) \\ &\qquad \le d_{city}(p, p_d) \le d_{city}(p_{\alpha,\beta}, p_d) + d_{city}(p_{\alpha,\beta}, p). \end{aligned} \tag{6}$$

In this paper, we employ the lower bound denoted as $d'_{city}(p, p_d)$ to approximate the city-block distance between p and p_d:

$$d'_{city}(p, p_d) = d_{city}(p_{\alpha,\beta}, p_d) - d_{city}(p_{\alpha,\beta}, p). \tag{7}$$

For the distorted image, we need not fit p_d to a GGD model, which is not appropriate for depicting the distorted images. What we compute is the distance $d_{city}(p_{\alpha,\beta}, p_d)$ between the fitted GGD of the reference image and the coefficient distribution of the distorted image according to (3). By considering the prediction error of GGD modeling, we could obtain the approximated distance according to (7).

As we have approximated the city-block distance for each reorganized DCT subband, finally by pooling over all the reorganized DCT subbands, the visual distance between the reference and distorted image is obtained:

$$\begin{aligned} V_{dist} &= \log_{10}\left(1 + \frac{1}{c}\sum_{k=1}^{K} |d'^{k}_{city}(p^k, p^k_d)|\right) \\ &= \log_{10}\left(c + \sum_{k=1}^{K} |d'^{k}_{city}(p^k, p^k_d)|\right) - \log_{10}(c) \end{aligned} \tag{8}$$

where k denotes the reorganized DCT subband index, which ranges from S_1 to S_9 as shown in Fig. 2. The parameter c is utilized for scaling the distortion measure to avoid the variation of V_{dist} being too small. As shown in (8), if the summation of city-block distances is not significantly smaller than c, the variation of V_{dist} will be amplified. It just helps depict the visual distance clearly. Therefore, as a scaling factor, the parameter c will not influence the performance of the proposed RR IQA, which is demonstrated by the experimental results in the following section. In this paper, c is set as 0.0001 for simplicity.

III. Experimental Results

In this section, we firstly show the efficiency of the reorganized DCT strategy for the proposed method. Subsequently, the performances of different IQAs will be compared to demonstrate the efficiency of the proposed RR IQA for evaluating the image perceptual quality.

A. Efficiency of the DCT Reorganization Strategy

All the reference images from the LIVE image database [18] are employed to demonstrate the efficiency of the DCT reorganization strategy, compared with the steerable pyramid [19], [20], which has been employed in the FR IQAs, such as visual information fidelity (VIF) [21]. As illustrated in [10] and [11], after the 3-scale, 3-orientation steerable pyramid decomposition, the high-frequency subbands correspond to the reorganized DCT subbands from S_1 to S_9. As described in Section II, the average prediction error, specifically the city-block distance, between the fitted GGD function and the actual coefficient distribution of the 6 subbands (from S_4 to S_9), is employed as the criterion to evaluate the performances of different transforms. According to the definition in (3), the smaller the prediction error, the better fitting is the GGD function, which means that the GGD can more accurately describe the coefficient distribution. The prediction error of each reference image in the LIVE image database is illustrated in Fig. 4. It can be observed that for most images, the prediction errors using the reorganized DCT are smaller than those using the steerable pyramid. The average prediction error using the reorganized DCT of

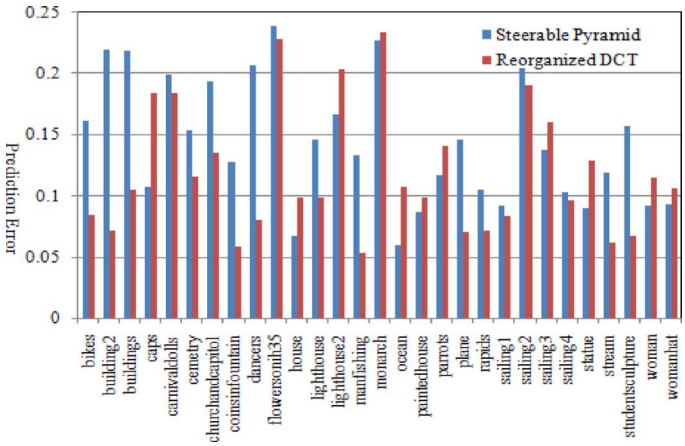

Fig. 4. Prediction error of each reference image in the LIVE image database.

all the images is only 0.1183, compared with 0.1664 using the steerable pyramid. This result means that the coefficient distributions of the reorganized DCT subbands are more suitable for GGD modeling, which will further help improve the RR IQA performance.

B. Performance of the Proposed RR IQA

We compare the performance of our proposed RR IQA method with the representative RR image quality metric WNISM [10], [11], and the FR metrics: PSNR, and SSIM [3]. The IQA methods are evaluated on the LIVE [18] image database, which comprises the most prevailing distortion types, including JPEG, JPEG 2000 (J2K), blur, white Gaussian noise (WGN), and fast fading.

We follow the performance evaluation procedure employed in the Video Quality Experts Group (VQEG) HDTV test [22] and that in [23]. Let x_j represent the visual quality index of the jth distorted image obtained from the corresponding IQA. A five parameter $\{\delta_1, \delta_2, \delta_3, \delta_4, \delta_5\}$ monotonic logistic function is employed to map x_j to V_j:

$$V_j = \delta_1 logistic(\delta_2, (x_j - \delta_3)) + \delta_4 x_j + \delta_5$$
$$logistic(\tau, \upsilon) = \frac{1}{2} - \frac{1}{1 + \exp(\tau\upsilon)}. \quad (9)$$

After the nonlinear mapping process, the correlation coefficients (CC) between the subjective and the nonlinearly mapped objective scores, which provides an evaluation of the prediction accuracy, and the Spearman rank-order correlation coefficients (SROCC), which measures the prediction monotonicity, are employed to evaluate the different IQA performances. Larger CC and SROCC values mean that the objective and subjective scores correlate better, that is to say, a better performance of the IQA. Furthermore, the root mean square prediction error (RMSE) of the fitting procedure is also utilized to measure the IQAs' efficiencies. Here, smaller RMSE values indicate smaller errors between the two scores, hence a better performance.

The performance comparisons between different IQAs are listed in Table I. It can be observed that our proposed RR IQA method, with larger CC/SROCC and smaller RMSE values, outperforms the well-recognized RR WNISM [11], and the FR PSNR, although it is slightly inferior to the FR SSIM [3]. However, SSIM requires the whole reference image for assessing the image perceptual quality. For the proposed RR IQA, if we employ all the 9 reorganized DCT subbands (from S_1 to S_9), only $9 \times 3 = 27$ parameters are needed to represent the reference image. We provide experimental results based on all the 63 DCT subbands without the reorganization strategy, whose performance is illustrated in the 5th column of Table I. As 3 parameters

TABLE I
PERFORMANCES OF DIFFERENT IQAS

	PSNR	SSIM	WNISM	63 DCT subbands	The proposed RRIQA (with all 9 subbands)
CC	0.871	0.904	0.738	0.833	0.883
SROCC	0.876	0.910	0.779	0.836	0.880
RMSE	13.40	11.67	18.43	15.12	12.83

are needed for depicting each GGD, $63 \times 3 = 189$ parameters are required for modeling coefficient distributions of all the 63 DCT subbands, which is much larger than WNISM and the proposed method. However, it performs worse than our proposed method even with such a larger number of parameters. The reason is that it does not consider the identical nature of the coefficient distributions between neighboring DCT subbands. In our proposed method, the adjacent DCT subbands are reorganized as a whole subband. In this simple way, the relationship between neighboring DCT subbands has been exploited.

Furthermore, we are interested in whether all the reorganized DCT subbands are necessary for designing an efficient RR IQA. Therefore, we evaluate the performances of each reorganized subband individually, as well as some subband combinations, which are illustrated in Table II. Firstly, it can be observed that the horizontal subbands (S_1, S_4, and S_7) and vertical subbands (S_2, S_5, and S_8) outperform the diagonal subbands (S_3, S_6, and S_9) in the same scale. The reason is that the HVS exhibits the orientation selectivity and preference [24], [25]. Usually, the eye is more sensitive to the horizontal and vertical frequency components, compared to the diagonal components, known as the oblique effect [26]. Secondly, by referring to the performance of each individual subband and the oblique effect, the combination of S_4, S_5, S_7, and S_8 is evaluated. The performance is better than any individual subband, which means that the subband combinations can help improve the IQA performance. Thirdly, the 6 high-frequency subbands (from S_4 to S_9) are employed to evaluate the performance. However, the performance degrades, which is even worse than the performance of S_5. Thus, we believe that the diagonal subbands cannot efficiently help improve the RR IQA performance, which only introduce additional RR data rates. Finally, the horizontal subbands (S_1, S_4, and S_7) and vertical subbands (S_2, S_5, and S_8) are employed for evaluation. The performance is significantly improved, approaching the performance of all 9 subbands. One reason for the improvement in model performance is the aforementioned oblique effect, where the HVS is more sensitive to the horizontal and vertical frequency components. The diagonal subbands are proved to be unnecessary for the proposed RR IQA. The other is that some distortions are also introduced into the low-frequency components, which need to be accounted for when evaluating image quality. As to the well-recognized RR WNISM, 6 of 12 oriented subbands generated from the 3-scale, 4-orientation steerable pyramid are employed for the feature extraction. For a fair comparison, 6 reorganized DCT subbands (only the horizontal and vertical ones) are employed for designing the RR IQA. As introduced in Section II-C, $8+8+8+3 = 27$ bits are required for representing the GGD parameters of each subband. $27 \times 6 = 162$ bits are needed for representing the 6 subbands of the reference image. Moreover, if the transmitted bits are needed to be further reduced, we can only employ the S_5 subband for designing the RR IQA, which according to Table II provides a good performance and requires only 27 bits to represent the reference image.

Compared with WNISM, the performance of the proposed RR IQA has been significantly improved based on the same amount of extracted features. The improvement may be attributed to two reasons. Firstly, the employed reorganized DCT can efficiently provide better energy compaction, compared with the steerable pyramid. It has been demonstrated that it is more suitable for the GGD modeling, which generates smaller prediction errors. Secondly, the symmetric city-block distance

TABLE II
PERFORMANCES OF THE INDIVIDUAL SUBBAND AND SOME SUBBAND COMBINATIONS

Subband	CC	SROCC	RMSE	Subband	CC	SROCC	RMSE
S_1	0.636	0.557	21.09	S_7	0.779	0.786	17.13
S_2	0.638	0.558	21.03	S_8	0.794	0.803	16.62
S_3	0.498	0.450	23.69	S_9	0.590	0.603	22.06
S_4	0.839	0.817	14.88	$S_4+S_5+S_7+S_8$	0.865	0.863	13.70
S_5	0.861	0.845	13.91	$S_4+S_5+S_6+S_7+S_8+S_9$	0.856	0.857	14.12
S_6	0.812	0.814	15.94	$S_1+S_2+S_4+S_5+S_7+S_8$	0.883	0.879	12.84

TABLE III
PERFORMANCE COMPARISONS BETWEEN DIFFERENT RR IQAS

	WNISM (steerable pyramid + KLD)	SPC (steerable pyramid + city-block distance)	RDK (reorganized DCT + KLD)	Proposed RRIQA (reorganized DCT + city-block distance)
CC	0.738	0.751	0.854	0.883
SROCC	0.779	0.760	0.848	0.879
RMSE	18.43	18.06	14.22	12.84

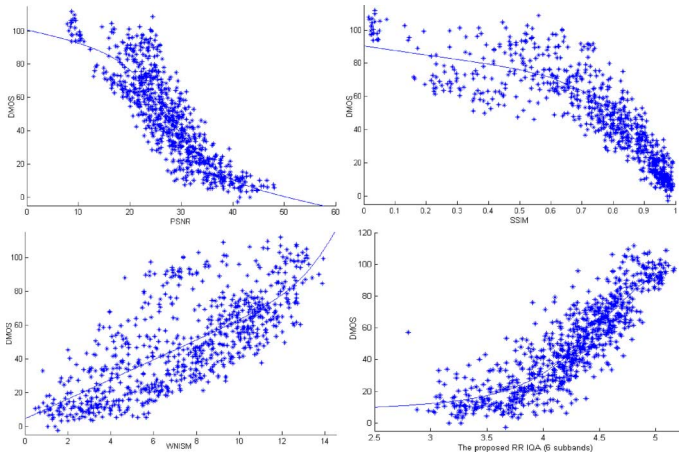

Fig. 5. Scatter plots of the DMOS values versus model predictions on the LIVE image database. Each sample point represents one test image. [Top left: PSNR; top right: SSIM; bottom left: WNISM; bottom right: the proposed method (6 subbands)].

is more reasonable for measuring the image visual quality distance, compared with the asymmetric KLD distance. Therefore, the efficiencies of the DCT reorganization strategy and the city-block distance are evaluated individually. The performances of WNISM (steerable pyramid + KLD), the proposed RR IQA (reorganized DCT + city-block distance), SPC (steerable pyramid + city-block distance), and RDK (reorganized DCT + KLD) are compared in Table III. It can be observed that SPC is comparable with WNISM, with larger CC and smaller RMSE/SROCC. Therefore, the city-block distance performs comparably with KLD for depicting the histogram distances of the wavelet subbands. Moreover, RDK and the proposed RR IQA significantly outperform WNISM and SPC. It means that the reorganized DCT is more efficient than the steerable pyramid for evaluating the image quality, which has been proved to be more suitable for the GGD modeling. Furthermore, the proposed RR IQA outperforms RDK, which implies that the city-block distance is more efficient to describe the histogram distances of reorganized DCT subbands.

Furthermore, the performances of the proposed RR IQA (6 subbands) over different values of c are illustrated in Table IV. It can be observed that the performance is not influenced by the variation of c. Therefore, in this paper, c is simply set as 0.0001. The scatter-plots of different IQAs on the whole LIVE image database are shown in Fig. 5. It can be observed that the results of our proposed method scatter closely around the fitted line, which indicates a good performance.

Table V illustrates the performances of different IQAs over individual distortion types from the LIVE image database. It can be observed that our proposed method outperforms WNISM except for the J2K distortion. The reason is that J2K images are coded based on the wavelet transform, which can be more accurately depicted by the steerable pyramid transform than DCT. As the proposed RR IQA method employs the block-based DCT and reorganizes the coefficients into a three-level tree, it is much more efficient for depicting images processed by DCT. Therefore, the proposed RR IQA performs very well on the JPEG distortion type. Moreover, although WNISM performs better than PSNR except for the WGN distortion, its performance degrades significantly when images with different types of distortions are tested together, which is shown in Table I. This is also the main drawback of WNISM, which has been revealed in the previous literature [28]. However, the proposed RR method is more robust, which can not only perform very well on individual distortion types, but also demonstrate a good performance over all the distortion types, which is even better than the FR IQA PSNR.

TABLE IV
PERFORMANCES OF THE PROPOSED RR IQA OVER DIFFERENT VALUES OF c

	c=0.00001	c=0.0001	c=0.001	c=0.01	c=0.05	c=0.1	c=1	c=2
CC	0.8825	0.8827	0.8801	0.8803	0.8804	0.8801	0.8796	0.8732
SROCC	0.8782	0.8786	0.8756	0.8762	0.8764	0.8762	0.8753	0.8750
RMSE	12.852	12.840	12.972	12.964	12.958	12.972	12.997	13.314

TABLE V
PERFORMANCES OF DIFFERENT IQAS OVER INDIVIDUAL DISTORTION TYPES

		J2K	JPEG	WGN	blur	fast fading
PSNR	CC	0.896	0.887	0.978	0.785	0.891
	SROCC	0.895	0.881	0.985	0.782	0.890
	RMSE	11.22	17.72	8.74	11.45	12.94
WNISM	CC	0.928	0.862	0.869	0.926	0.918
	SROCC	0.924	0.854	0.848	0.930	0.907
	RMSE	9.37	16.17	13.87	6.96	11.31
The proposed RR IQA (with 6 subbands)	CC	0.845	0.931	0.891	0.930	0.919
	SROCC	0.838	0.924	0.880	0.930	0.910
	RMSE	13.50	11.66	12.68	6.79	11.24

IV. CONCLUSION

In this letter, an RR IQA is proposed by GGD modeling of the coefficient distributions of the reorganized DCT subbands. The symmetric city-block distance is employed to measure the image visual quality distance. Experimental results have demonstrated that only a small number of parameters is sufficient to represent the image quality. The proposed RR IQA outperforms the WNISM and even PSNR, which means that the proposed metric correlates well with the human perception of the image quality.

REFERENCES

[1] G. Zhai, J. Cai, W. Lin, X. Yang, W. Zhang, and M. Etoh, "Cross-dimensional perceptual quality assessment for low bit-rate videos," *IEEE Trans. Multimedia*, vol. 10, no. 7, pp. 1316–1324, Nov. 2008.

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

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

[4] K. Yang, C. C. Guest, K. E. Maleh, and P. K. Das, "Perceptual temporal quality metric for compressed video," *IEEE Trans. Multimedia*, vol. 9, no. 7, pp. 1528–1535, Nov. 2007.

[5] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video quality assessment," *IEEE Trans. Multimedia*, vol. 4, no. 1, pp. 129–132, Mar. 2002.

[6] Z. Yu, H. R. Wu, S. Winkler, and T. Chen, "Vision-model-based impairment metric to evaluate blocking artifact in digital video," *Proc. IEEE*, vol. 90, no. 1, pp. 154–169, Jan. 2002.

[7] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using nature scene statistics: JPEG 2000," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1918–1927, Nov. 2005.

[8] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 717–728, Apr. 2009.
[9] P. L. Callet, C. V. Gaudin, and D. Barba, "A convolutional neural network approach for objective video quality assessment," *IEEE Trans. Neural Netw.*, vol. 17, no. 5, pp. 1316–1327, May 2006.
[10] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E. Yang, and A. C. Bovik, "Quality-aware images," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1680–1689, Jun. 2006.
[11] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," in *Proc. SPIE, Human Vision and Electronic Imaging*, Jan. 2005.
[12] X. Gao, W. Lu, D. Tao, and X. Li, "Image quality assessment based on multiscale geometric analysis," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1409–1423, Jul. 2009.
[13] L. Li, C. S. Tong, and S. K. Choy, "Texture classification using refined histogram," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1371–1378, May 2010.
[14] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1661–1666, Oct. 2000.
[15] Z. Xiong, K. Ramchandran, M. T. Orchard, and Y. Q. Zhang, "A comparative study of DCT- and wavelet-based image coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 5, pp. 692–695, Aug. 1999.
[16] D. Zhao, W. Gao, and Y. K. Chan, "Morphological representation of DCT coefficients for image compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 9, pp. 819–823, Sep. 2002.
[17] T. M. Cover and J. A. Thomas, *Element of Information Theory*. New York: Wiley, 1991.
[18] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, LIVE Image Quality Assessment Database. [Online]. Available: http://live.ece.utexas.edu/research/quality.
[19] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multi-scale transforms," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 587–607, Mar. 1992.
[20] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. ICIP*, pp. 444–447, Oct. 1995.
[21] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
[22] VQEG, Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, 2000. [Online]. Available: http://www.vqeg.org.
[23] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.
[24] A. Ahumada and H. Peterson, "Luminance model based DCT quantization for color image compression," in *Proc. SPIE Human Vision Visual Process., Digit. Display III*, vol. 1666, pp. 365–374, 1992.
[25] H. Peterson, A. Ahumada, and A. B. Watson, "Improved detection model for DCT coefficient quantization," in *Proc. SPIE Human Vision Visual Process., Digit. Display IV*, vol. 1913, pp. 191–201, 1993.
[26] B. Li, M. R. Peterson, and R. D. Freeman, "Oblique effect: A neural basis in the visual cortex," *J. Neurophysiol.*, pp. 204–217, 2003.
[27] D. Tao, X. Li, W. Lu, and X. Gao, "Reduced-reference IQA in contourlet domain," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 6, pp. 1623–1627, Dec. 2009.
[28] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE J. Select. Topics Signal Process.*, vol. 3, no. 2, pp. 202–211, Apr. 2009.

Perceptually Guided Fast Compression of 3-D Motion Capture Data

A. Firouzmanesh, I. Cheng, and A. Basu

***Abstract*—A time efficient compression technique, incorporating attention stimulating factors, for motion capture data is proposed. Compression ratios of 25:1 to 30:1 can be achieved with very little noticeable degradation in perceptual quality of animation. Experimental analysis shows that the proposed algorithm is much faster than comparable approaches using wavelets, thereby making our approach feasible for motion capture, transmission, and real-time synthesis on mobile devices, where processing power and memory capacity are limited.**

***Index Terms*—Compression, motion capture, online 3-D environments, perception of animation.**

I. Introduction

Using motion capture data is an effective way to produce skeletal animations. In online applications, efficient compression of motion capture data can contribute to optimal use of available bandwidth while preserving the transmission of higher quality animation. In recent years, different approaches have been proposed to address the motion data compression problem. However, none of them directly discusses the possibility of achieving further reduction in data size with little noticeable perceptual degradation, considering the human visual system.

Even though perceptual factors in image and video have been widely studied [7], there is no existing qualitative metric [5] that considers different aspects of perception in animation. Thus, in this paper we propose a technique for perceptually guided compression of motion data, considering some of the most important factors, which can affect the perception of animation.

Motion capture refers to "the process of recording movement and translating that movement on to a digital model" [15]. Motion can be recorded using optical, mechanical, or magnetic devices by tracking the movements of key points (such as joints) on an object (Fig. 1). Motion capture is useful in many applications including military, entertainment, sports, medicine, computer vision, and robotics. In computer animation, motion capture data can be used to create realistic 2-D and 3-D animated characters.

Motion data are usually sampled at frequencies between 60 to 240 Hz. A hierarchical biped structure (skeleton) is typically used to coordinate the relative position and movement of each child key point with respect to its parent key point. The movement of a key point can be precisely described using 9 degrees of freedom (3 translations, 3 rotations, and 3 scaling factor) in the x-, y-, and z- coordinates. We use the term channel to associate with one degree of freedom (DOF), tracing

Manuscript received August 10, 2010; revised December 21, 2010 and March 04, 2011; accepted March 07, 2011. Date of publication March 17, 2011; date of current version July 20, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Fernando M. B. Pereira.

A. Firouzmanesh is with the Multimedia Research Center, Department of Computing Science, University of Alberta, Edmonton, AB T8N 3N1, Canada (e-mail: firouzma@ualberta.ca).

I. Cheng and A. Basu are with the Department of Computing Science, University of Alberta, Edmonton, AB T8N 3N1, Canada (e-mail: locheng@ualberta.ca; basu@ualberta.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2011.2129497