Reduced-Reference IQA in Contourlet Domain

Dacheng Tao, *Member, IEEE*, Xuelong Li, *Senior Member, IEEE*, Wen Lu, and Xinbo Gao, *Senior Member, IEEE*

***Abstract*—The human visual system (HVS) provides a suitable cue for image quality assessment (IQA). In this paper, we develop a novel reduced-reference (RR) IQA scheme by incorporating the merits from the contourlet transform, contrast sensitivity function (CSF), and Weber's law of just noticeable difference (JND). In this scheme, the contourlet transform is utilized to decompose images and then extract features to mimic the multichannel structure of HVS. CSF is applied to weight coefficients obtained by the contourlet transform to simulate the appearance of images to observers by taking into account many of the nonlinearities inherent in HVS. JND is finally introduced to produce a noticeable variation in sensory experience. Thorough empirical studies are carried out upon the Laboratory for Image and Video Engineering database against the subjective mean opinion score and demonstrate that the proposed framework has good consistency with subjective perception values and the objective assessment results can well reflect the visual quality of images.**

***Index Terms*—Contourlet transform, human visual system (HVS), image quality assessment (IQA), reduced reference (RR).**

I. Introduction

The objective of image quality assessment (IQA) [2], [4] is very important for image retrieval and content analysis, multimedia information organization [18], watermarking [19], face image analysis [20], [21], palmprint recognition [3], human motion analysis [1], and motion-image-based gender recognition [22]. It provides computational models to measure the perceptual quality of an image. In recent years, many methods have been designed to evaluate the quality of an image, which may be distorted during acquisition, transmission, compression, restoration, and processing (e.g., watermark embedding). According to the availability of a reference image, existing objective IQA methods can be categorized into three categories: full-reference (FR), no-reference (NR) [16], and reduced-reference (RR) methods. The focus of this paper is RR IQA because it is a compromise between FR and NR and it is designed for IQA by employing partial information of the corresponding reference.

Based on results in natural image statistics, Wang and Bovik [2] proposed the wavelet-domain natural image statistic metric (WNISM), which achieves promising performance for image visual perception quality evaluation. The underlying factor in WNISM is that the marginal distribution of wavelet coefficients of a natural image conforms to the generalized Gaussian distribution. Based on this fact, WNISM measures the quality of a distorted image by the fitting error between the wavelet coefficients of the distorted image and the Gaussian distribution of the reference.

Although WNISM has been recognized as the standard method for RR IQA, it fails to consider the statistical correlations of wavelet coefficients in different subbands and the visual response characteristics of the mammalian cortical simple cells. Moreover, wavelet transforms cannot explicitly extract the image geometric information, e.g., lines and curves, and wavelet coefficients are dense for smooth image edge contours. Therefore, there is still a big room to further improve the performance of RR IQA.

In this paper, to target the aforementioned problems in WNISM, to further improve the performance of RR IQA, and to broaden RR-IQA-related applications, a novel human visual system (HVS)-driven scheme is proposed. This framework is constructed by pooling *the contourlet transform* [7], *contrast sensitivity function* (CSF) [10], and Weber's law of *just noticeable difference* (JND) [10] together. In 2002, Do and Vetterli proposed the contourlet transform [7], which is also termed the pyramidal directional filter bank, which has the following two key characteristics: 1) the underlying multiresolution mechanism can represent images in continuous resolution values, which is normally called bandpass, and 2) the basis of the contourlet transform are directional and local in time and frequency domains. The new framework is consistent with HVS: the contourlet transform decomposes images for feature extraction to mimic the multichannel structure of HVS, CSF reweights the contourlet-decomposed coefficients to mimic the nonlinearities inherent in HVS, and JND produces a noticeable variation in sensory experience. Extensive experiments based on the Laboratory for Image and Video Engineering (LIVE) database [14] against the subjective *mean opinion score* (MOS) have been conducted to demonstrate the effectiveness of the new framework.

II. IQA in the Contourlet Domain

In this paper, we develop a novel scheme for IQA by applying the contourlet transform to decompose images and extract effective features. This scheme quantifies the errors between the distorted and the reference images by mimicking the error sensitivity function [2] in HVS. The objective of this scheme is to provide IQA results, which have good consistency with subjective perception values. This framework incorporates merits from three components, i.e., the contourlet transform, CSF, and Weber's law of JND, to model the process of image perception.

The scheme works with the following stages: 1) The contourlet transform is utilized to decompose both the reference image at the sender side and the distorted image at the receiver side. 2) CSF masking is utilized to balance subband coefficients in different scales obtained by the contourlet transform. With this stage, we can simulate the appearance of images to observers by taking into account many of the nonlinearities inherent in HVS. 3) JND produces a noticeable variation in sensory experience. 4) A histogram is constructed for image representation, each bin of the histogram corresponds to the amount of visually sensitive coefficients of a selected subband, and finally, the normalization step is applied to the histogram. 5) The IQA result is the transformed city-block distance between the normalized histograms of the reference and distorted images. In this section, these five stages are detailed as follows.

Manuscript received May 30, 2008; revised January 27, 2009. First published May 19, 2009; current version published November 18, 2009. This work was supported in part by the National Science Foundation of China under Grant 60771068, Grant 60702061, and Grant 60832005, by the Open-End Fund of National Laboratory of Pattern Recognition in China and National Laboratory of Automatic Target Recognition, Shenzhen University, China, by the Program for Changjiang Scholars and Innovative Research Team in the University of China (IRT0645), and by Nanyang Technological University under Start-Up Grant M58020010. This paper was recommended by Associate Editor H. Zhang.

D. Tao is with the School of Computer Engineering, Nanyang Technological University, Singapore 639798 (e-mail: dctao@ntu.edu.sg).

X. Li is with the School of Computer Science and Information Systems, Birkbeck College, University of London, WC1E 7HX London, U.K. (e-mail: xuelong@dcs.bbk.ac.uk).

W. Lu and X. Gao are with the School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: xbgao@mail.xidian.edu.cn; terry.luwen@gmail.com).

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/TSMCB.2009.2021951

1083-4419/$25.00 © 2009 IEEE

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on November 8, 2009 at 03:47 from IEEE Xplore. Restrictions apply.

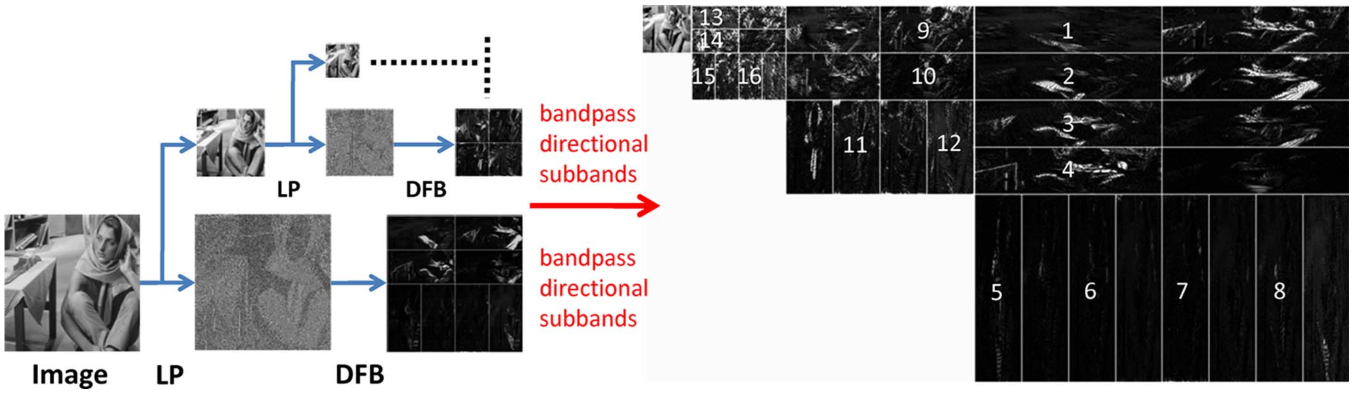

Fig. 1. Illustration of the contourlet transform.

A. Contourlet Transform

Wavelet transform has successfully been applied in a wide variety of signal processing tasks, e.g., speech signal compression and voice-based person identification, because it is an optimal greedy approximation to extract a singularity structure for 1-D piecewise smooth signals. Although the 2-D extension, i.e., 2-D wavelet transform, can also be applied to image-processing-relevant applications, e.g., compression, denoising, restoration, segmentation, and structure detection, it can only deal with the singularity problem of a point [5]. To our knowledge, images contain rich and varied information, e.g., texture and edges, and wavelet transform is not effective in dealing with directional information. Therefore, it is essential to develop a directional representation framework for precisely detecting orientations of singularities like edges in a 2-D image while providing near-optimal sparse representations.

The contourlet transform [7] is such a method for optimally representing a high-dimensional function. It can detect, organize, represent, and manipulate data, e.g., edges, which nominally span a high-dimensional space but contain important features approximately concentrated on lower dimensional subsets, e.g., curves. The contourlet transform is constructed via filter banks and can be viewed as an extension of wavelets with directionality. For implementation, it utilizes the Laplacian pyramid [8] to capture point discontinuities and directional filter banks [9] to link point discontinuities into linear structures, as shown in Fig. 1. Based on these two steps, it captures the intrinsic geometrical structure of an image. In the proposed framework, image is decomposed into three pyramidal levels. Based on the characteristics of directional filter banks ("9-7" biorthogonal filters [9]) for decomposition, coefficients in half of the directional subbands are selected for further processing. As shown in Fig. 1, contourlet transform is utilized to decompose an image, and a set of selected subbands are marked with white-dashed boxes and numerals.

B. CSF Masking

The contourlet transform is introduced to decompose images and then extract features to mimic the multichannel structure of HVS, i.e., HVS [10], [11] works similar to a filter bank (containing filters with various frequencies). CSF [10], [13] measures how sensitive we are to the various frequencies of visual stimuli, i.e., we are unable to recognize a stimuli pattern if its frequency of visual stimuli is too high. For example, given an image consisting of horizontal black and white stripes, we will perceive it as a gray image if stripes are very thin; otherwise, we can distinguish these stripes. Because coefficients in different frequency subbands have different perceptual importance, it is essential to balance the contourlet decomposed coefficients via a weighting scheme, CSF masking. In this framework, the CSF masking coefficients are obtained by the *modulation transfer function* [17], i.e.,

$$H(f) = a(b + cf)\exp(-cf)^d \tag{1}$$

where $f = f_n \cdot f_s$, the center frequency of the band, is the radial frequency in cycles per degree of the visual angle subtended, f_n is the normalized spatial frequency in cycles per pixel, and f_s is the sampling frequency with in pixels per degree. According to [12], a, b, c, and d are 2.6, 0.192, 0.114, and 1.1, respectively.

The sampling frequency f_s is defined as

$$f_s = \frac{2 \cdot v \cdot \tan(0.5^\circ) \cdot r}{0.0254} \tag{2}$$

where v is the viewing distance in meters, and r is the resolution power of the display in pixels per inch. In this framework, v is 0.8 m (about 2–2.5 times the height of the display), the display is 21 in with a resolution of 1024 × 768, and $r = \sqrt{1024^2 + 768^2}/21 = 61$ pixels/in. According to the *Nyquist sampling theorem*, f changes from 0 to $f_s/2$, so f_n changes from 0 to 0.5. Because the contoured transform is utilized to decompose an image into three scales from coarse to fine, we have three normalized spatial frequencies, i.e., $f_{n1} = 32/3$, $f_{n2} = 16/3$, and $f_{n3} = 8/3$. Weighting factors are identical for coefficients in an identical scale.

In detail, when the contourlet transform is utilized to decompose an image, we obtain a series of contourlet coefficients $c_{i,j}^k$, where k denotes the level index (the scale sequence number) of contourlet transform, i stands for the serial number of the directional subband index at the kth level, and j represents the coefficient index. By using CSF masking, the coefficient $c_{i,j}^k$ is scaled to $x_{i,j}^k = H(f_k) \cdot c_{i,j}^k$.

C. JND Threshold

Because HVS is sensitive to coefficients with a larger magnitude, it is valuable to preserve visually sensitive coefficients. JND, a research result in psychophysics, is a suitable way for this function. It measures the minimum amount by which stimulus intensity must be changed to produce a noticeable variation in the sensory experience. In our framework, the contourlet transform is introduced to decompose an image, high-pass subbands contain the primary contour and texture information of the image, CSF masking makes coefficients have similar perceptual importance in different frequency subbands, and then JND is calculated to obtain a threshold to remove visually insensitive coefficients. The amount of visually sensitive coefficients reflects the visual quality of the reconstructed images. The lower the JND threshold is, the more coefficients are utilized for image reconstruction and the better visual quality of the reconstructed image is. Therefore,

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on November 8, 2009 at 03:47 from IEEE Xplore. Restrictions apply.

TABLE I
COMPARISON OF CONSISTENCY BETWEEN SUBJECTIVE AND OBJECTIVE ASSESSMENT

MODEL	JPEG					JPEG2000				
	CC	ROCC	OR	MAE	RMS	CC	ROCC	OR	MAE	RMS
PSNR	0.9229	0.8905	0.1886	7.1180	9.1540	0.9330	0.9041	0.0947	6.4070	8.3130
WNISM	0.9291	0.9069	01486	6.0236	8.2446	0.9261	0.9135	0.1183	6.1350	7.9127
Proposed	**0.9493**	**0.9309**	**0.1029**	**5.3177**	**6.6941**	**0.9435**	**0.9258**	**0.0888**	**5.6616**	**7.1299**

the normalized histogram reflects the visual quality of an image. Here, the JND threshold is defined as

$$T = \frac{\alpha}{M}\sum_{i=1}^{M}\sqrt{\frac{1}{N_i-1}\sum_{j=1}^{N_i}(x_{i,j}-\overline{x}_i)^2} \tag{3}$$

where $x_{i,j}$ is the jth coefficient of the ith subband in the finest scale, $\overline{x}_i$ is the mean value of the ith subband coefficients, M is the amount of selected subbands in the finest scale, N_i is the amount of coefficients of the ith subband, and α is a tuning parameter corresponding to different types of distortion.

D. Normalized Histogram for Image Representation

By using JND threshold T, we can count the number of visually sensitive coefficients in the nth selected subband and define the value as $C_T(n)$, which means the number of coefficients in the nth selected subband that are larger than T obtained from (3). The number of coefficients in the nth selected subband is $C(n)$. Therefore, for a given image, we can obtain the normalized histogram with L bins ($L-1$ subbands are selected, and the other bin consists of the threshold value) for representation, and the nth entry is given by

$$P(n) = \frac{C_T(n)}{C(n)} \tag{4}$$

where L is 17. Fig. 1 shows that the contourlet transform decomposes an image into 4 levels with 33 subbands. The right-hand side of the figures shows that 16 particular subbands are selected to form $P(n)$.

E. Sensitive Errors Pooling

Based on (4), we can obtain the normalized histograms for both the reference and the distorted images as $P_R(n)$ and $P_{\mathrm{D}}(n)$, respectively. In this framework, we define the metrics of the distorted image quality as

$$Q = \frac{1}{1+\log_2\left(\frac{S}{Q_0}+1\right)} \tag{5}$$

where $S=\sum_{n=1}^{L}|P_R(n)-P_D(n)|$ is the city-block distance between $P_R(n)$ and $P_{\mathrm{D}}(n)$, and Q_0 is a constant used to control the scale of the distortion measure. In this framework, we set Q_0 as 0.1. The log function is introduced here to reduce the effects of a large S and enlarge the effects of a small S, so that we can conveniently analyze a large scope of S. There is no particular reason for choosing the city-block distance, which can be replaced by others, e.g., Euclidean norm. The same goes for the base 2 for the logarithm. The entire function preserves the monotonic property of S.

III. EXPERIMENTS

We design two sets of experiments to demonstrate the effectiveness of the proposed RR IQA scheme. The first set of experiments aims to examine the scheme by a single distortion (e.g., JPEG and JPEG2000), while the second set focuses on mixed distortions in different images. Note that most of the existing metrics do not work well under a mixed-distortion situation.

All reported experiments were carried out upon the LIVE databases constructed at the University of Texas [14]. The LIVE database includes 29 high-resolution 24-b/pixel red–green–blue images (typically 768 × 512) and a series of corresponding distorted images after JPEG and/or JPEG2000 compression. As a result, the LIVE database contains 175 JPEG images (the bit rates in the range of 0.150–3.336 b/pixel) and 169 JPEG2000 images (the bit rates in the range of 0.028–3.150 b/pixel). The database also provides subjective evaluation results, e.g., MOS, for evaluating the consistency of objective IQA metrics against the human perception. In detail, each JPEG compressed image was viewed by 13–20 subjects, and each JPEG2000 image was viewed by 25 subjects. Each subject was asked to mark "bad," "poor," "fair," "good," or "excellent" to a compressed image. The raw scores of each subject were normalized and rescaled from 1 to 100. Afterward, MOS was obtained.

After calculating IQA scores of distorted images, we choose three measurements to test the consistency of the proposed method and subjective perception [15]: 1) the *correlation coefficient* (CC), which expresses the accuracy of objective metrics; 2) the *rank order CC* (ROCC), which expresses the monotonic property of objective metrics; and 3) the *outlier ratio* (OR), which expresses the stability of objective metrics. In addition, to further evaluate the proposed metric, we also report the RMS and *mean absolute error* (MAE) values between objective and subjective sensitivity.

A. Test 1: Different Images With an Identical Distortion

We first utilize consistency experiments between subjective and objective assessments for performance evaluation in comparing the proposed scheme with peak signal-to-noise ratio (PSNR) [4] and WNISM [2]. Experiments independently test JPEG and JPEG2000 images in LIVE IQA databases. The comparison results are shown in Table I and Fig. 2. The proposed scheme contains only one free parameter, i.e., α, which corresponds to different distortions, for calculating the JND threshold. In an identical distortion, α is identical to different images. Empirical studies show that by tuning α to different distortions, we can achieve better performances for IQA. In addition, if we set α as a constant for different distortions, the IQA performance is still acceptable. In our experiments, we tune α from $\{1,2,3,\ldots,9\}$ and choose the one corresponding to the best performance, as shown in experiments. We set α as 3 for the JPEG distortion and as 2 for JPEG2000 distortion.

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on November 8, 2009 at 03:47 from IEEE Xplore. Restrictions apply.

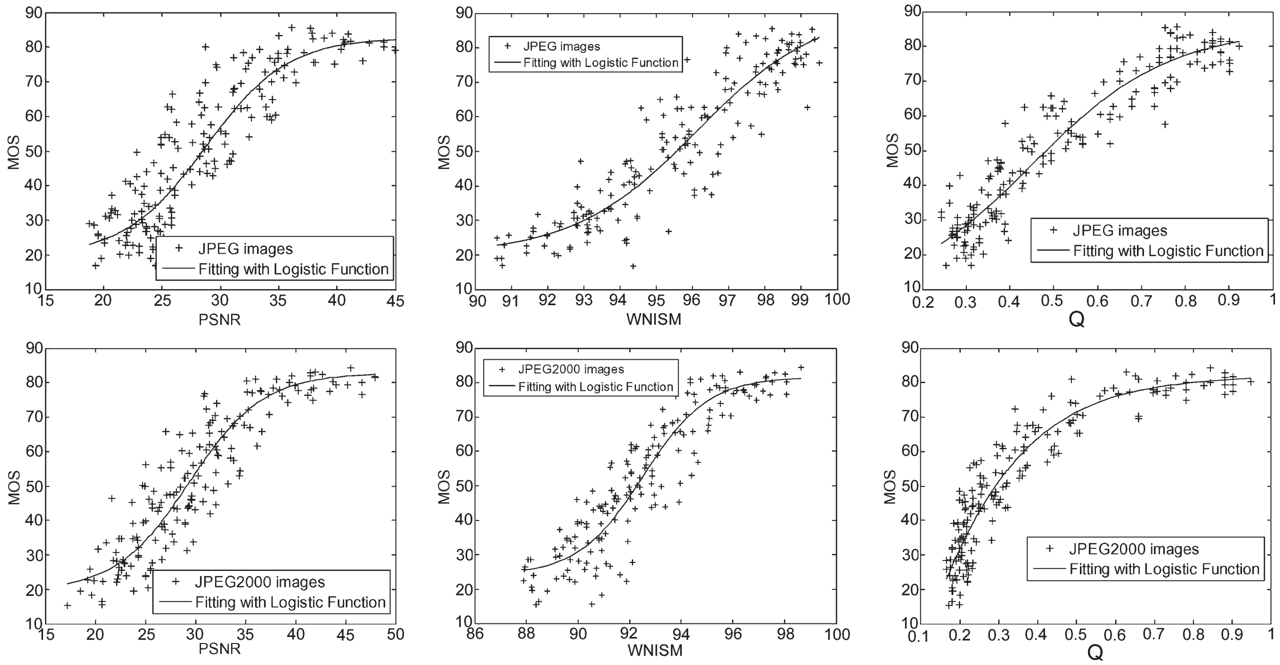

Fig. 2. Prediction of MOS by IQA metrics with an identical distortion.

TABLE II
RESULTS OF PSNR, WNISM, AND THE PROPOSED SCHEME

MODEL	TYPE	CC	ROCC	OR	MAE	RMSE
PSNR	FR	0.8766	0.8977	0.1366	6.846	8.878
WNISM	RR	0.7716	0.7651	0.3140	10.6148	13.0967
Proposed	**RR**	**0.9164**	**0.9032**	**0.1134**	**6.2184**	**7.9580**

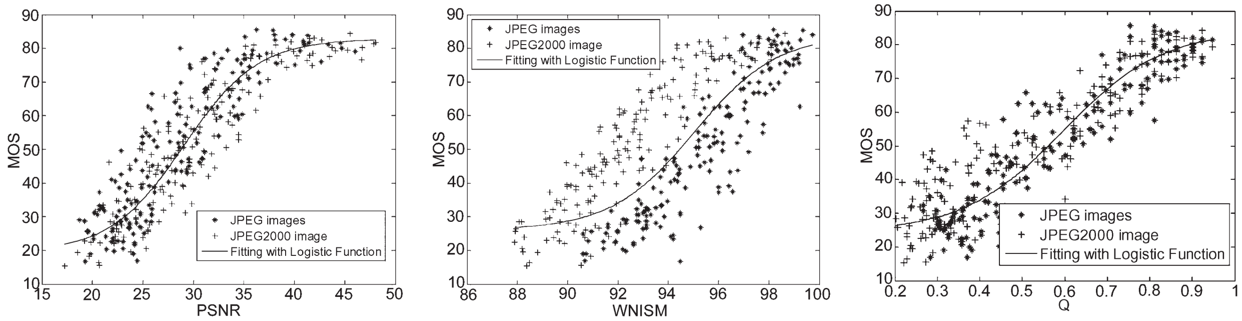

Fig. 3. Prediction of MOS by different IQA metrics with mixed distortions.

B. Test 2: Different Images With Mixed Distortions

Apart from the *identical distortion* in the aforementioned experiments, we also design a set of experiments for different images with mixed distortions. Similar to test 1, we use the LIVE IQA database and consider CC, ROCC, OR, MAE, and RMS as measurements for evaluation. The difference is that images are changed to JPEG and JPEG2000 mixed images. The experimental results are showed in Table II and Fig. 3.

As shown in both sets of experiments, the proposed RR IQA scheme performs better than PSNR and WNISM not only on the identical-distortion test but also on the mixed-distortion test in terms of: 1) accuracy (CC); 2) monotonic property (ROCC); and 3) stability (OR).

IV. CONCLUSION

In this paper, an RR IQA scheme has been proposed by incorporating merits of the contourlet transform, CSF, and Weber's law of JND. In comparing with existing IQA approaches, the proposed one has strong links with HVS: the contourlet transform, which aims to achieve an optimal approximation rate of piecewise smooth functions with discontinuities along twice continuously differentiable curves, is utilized to mimic the multichannel structure of HVS, CSF is utilized to balance magnitude of coefficients obtained by the contourlet transform to mimic the nonlinearities of HVS, and JND is utilized to produce a noticeable variation in sensory experience. In this scheme, images are represented by normalized histograms, which correspond to visually sensitive coefficients. The quality of a distorted image is measured

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on November 8, 2009 at 03:47 from IEEE Xplore. Restrictions apply.

by comparing the normalized histogram of the distorted image and that of the reference image. Thorough empirical studies show that the novel scheme performs better than the conventional standard RR method, i.e., WNISM. Since RR methods require limited information of the reference image, which could be a serious impediment for many applications, it is desirable to extend the proposed method to the NR [16] circumstance in the future.

Acknowledgment

The authors would like to thank the handling editor and all three anonymous reviewers for the constructive comments.

References

[1] Z. Liu and S. Sarkar, "Effect of silhouette quality on hard problems in gait recognition," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 35, no. 2, pp. 170–183, Apr. 2005.

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

[3] T. Zhang, X. Li, J. Yang, and D. Tao, "Multimodal biometrics using geometry preserving projections," *Pattern Recognit.*, vol. 41, no. 3, pp. 805–813, Mar. 2008.

[4] I. Avcibas, B. Sankur, and K. Sayood, "Statistical evaluation of image quality measures," *J. Electron. Imaging*, vol. 11, no. 2, pp. 206–213, Apr. 2002.

[5] J. K. Romberg, "Multiscale geometric image processing," Ph.D. dissertation, Rice Univ., Houston, TX, 2003.

[6] E. J. Candès and D. L. Donoho, "Curvelets—A surprising effective nonadaptive representation for objects with edges," in *Curves and Surfaces*. Nashville, TN: Vanderbilt Univ. Press, 2000, pp. 105–120.

[7] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.

[8] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. COM-31, no. 4, pp. 532–540, Apr. 1983.

[9] R. H. Bamberger and M. J. T. Smith, "A filter bank for the directional decomposition of images: Theory and design," *IEEE Trans. Signal Process.*, vol. 40, no. 4, pp. 882–893, Apr. 1992.

[10] B. A. Wandell, *Foundations of Vision*. Sunderland, MA: Sinauer Associates, Inc., 1995.

[11] I. Lee, J. Kim, Y. Kim, S. Kim, G. Park, and K. T. Park, "Wavelet transform image coding using human visual system," in *IEEE Asia-Pacific Conf. Circuits Syst.*, 1994, pp. 619–623.

[12] M. Miloslavski and Y.-S. Ho, "Zerotree wavelet image coding based on the human visual system model," in *IEEE Asia-Pacific Conf. Circuits Syst.*, 1998, pp. 57–60.

[13] M. J. Nadenau, J. Reichel, and M. Kunt, "Wavelet-based color image compression: exploiting the contrast sensitivity," *IEEE Trans. Image Process.*, vol. 12, no. 1, pp. 58–70, Jan. 2003.

[14] H. R. Sheikh, Z. Wang, A. C. Bovik, and L. K. Cormack, *Image and Video Quality Assessment Research*, 2003. [Online]. Available: http://live.ece.utexas.edu/rese-arch/quality/

[15] 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/

[16] R. Barland and A. Saadane, "Reference free quality metric for JPEG-2000 compressed images," in *8th Int. Symp. Signal Process. Appl.*, 2005, pp. 351–354.

[17] B. Chitprasert and K. R. Rao, "Human visual weighted progressive image transmission," *IEEE Trans. Commun.*, vol. 38, no. 7, pp. 1040–1044, Jul. 1990.

[18] D. Tao, X. Li, W. Hu, and S. J. Maybank, "Stable third-order tensor representation for color image classification," in *Proc. IEEE Int. Conf. Web Intell.*, 2005, pp. 641–644.

[19] C. Deng, X. Gao, X. Li, and D. Tao, "A local Tchebichef moments-based robust image watermarking," *Signal Process.*, vol. 89, no. 8, pp. 1531–1539, Aug. 2009.

[20] X. Gao, B. Xiao, D. Tao, and X. Li, "Image categorization: Graph edit distance + edge direction histogram," *Pattern Recognit.*, vol. 41, no. 10, pp. 3179–3191, Oct. 2008.

[21] D. Cai, X. He, J. Han, and H.-J. Zhang, "Orthogonal Laplacian faces for face recognition," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3608–3614, Nov. 2006.

[22] X. Li, S. J. Maybank, S. Yan, D. Tao, and D. Xu, "Gait components and their application to gender recognition," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 38, no. 2, pp. 145–155, Feb. 2008.

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on November 8, 2009 at 03:47 from IEEE Xplore. Restrictions apply.