ISSN : 0974 - 7435 Volume 7 Issue 10

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 7(10), 2013 [372-378]

Reduced-reference image quality assessment using energy change in reorganized DCT domain

Sheng Ding[1], Mei Yu[1,2]*, Xin Jin[1], Yang Song[1], Kaihui Zheng[1], Gangyi Jiang[1,2]

[1]Faculty of Information Science and Engineering, Ningbo University, Ningbo, (CHINA)

[2]National Key Lab of Software New Technology, Nanjing University, Nanjing, (CHINA)

ABSTRACT

Reduced-reference (RR) image quality assessment (IQA) intends to utilize less information of the reference image and yield higher evaluation accuracy. In this paper, a novel RR-IQA metric that measures differences between the energy in reorganized discrete cosine (RDCT) domain of reference and distorted images as perceived by Human is presented. Firstly, we decompose an image into ten sub-bands in this new frequency domain. Since RDCT representation exhibits structural similarities between sub-bands, and can mimic the function of Human Visual System (HVS). Secondly, we extract features from the ten sub-bands by analyzing what are the key elements that influence subjective quality. Finally, we fuse these extracted features based on the principal that we exert different importance in accordance with the different impact each individual sub-band plays on the perceptual quality. Experimental results demonstrate that the proposed metric outperforms the state-of-the-art (RR) IQA metrics and even the full-reference (FR) IQA metrics SVD and SSIM. What is more, compared with many existing RR IQAs, the proposed metric earns obvious superiority in terms of the amount of its required information from reference image and computational complexity.

KEYWORDS

Image quality assessment;
Reduced reference;
Reorganized discrete cosine transform;
Energy change.

INTORDUCTION

It is a prevailing tendency that a wide range of information exchange, particularly in the form of image and video transmission will be a significant and changing task in the days to come. For the reason that signal will be distorted as a consequence of the limited transmittal condition, So we have to accurately evaluate the quality of the distorted images, to what extent it is distorted in comparison with the reference image. As we know, human eyes are the ultimate receivers, hence the most accurate method of measuring the quality of image is subjective image quality assessment (IQA). In practice, however, subjective IQA is too inconvenient cumbrous and time-consuming. As a result, a metric enabling automatically to measure image's perceptual quality should be developed and ameliorated.

Objective image quality metrics can be classified into three categories in accordance with the amount of information used for predicting subjective quality. Most prevalent approaches are known as full-reference (FR) IQA, indicating that a complete reference image is known. In practical applications, however, all the information of reference image is not accessible. Hence, reduced-reference (RR) IQA have been developed which merely require a fraction of image information. No-reference (NR) IQA should also have to be developed at the worst condition that none of the reference image information is available.

The simplest FR metrics are the mean square error (MSE) and the peak signal-to-noise (PSNR). MSE and PSNR are prevalently adopted for its simplicity and clear meaning. However, these two IQAs merely take the difference in terms of pixel intensity into consideration, disregarding the perception properties. As a consequence, their performances will be negative. The structure similarity (SSIM) index[8] was developed which considers that eyes are sensitive to the structural distortion and proved to earn superiority to some FR IQAs. Eskicioglu *et al.*[12] utilized the singular value decomposition theorem to measure the quality of images, and yielded satisfactory results.

As a compromise between FR and NR, RR IQA indices are formulated to evaluate the perceptual quality by means of using partial information of the reference images. Figure 1 shows the fundamental structure of the RR model. An image x is transmitted to the receiver side via a transmission channel which probably evokes x to be a distorted one y, the feature which denote as X of image x extracted at the sender side is sent to the receiver side without distortion hypothetically, then extract the feature of the distorted image y denoted Y at the receiver side, the rigorous mathematical analysis of the difference between X and Y indicates the discrepancy in perception quality. Finally we get the predicted quality score s . In general, a good RR IQA should achieve a good tradeoff between the amount of information and prediction accuracy. To some degree, the more the amount of reference image information is required, the higher the accuracy will be.

A RR IQA metric[17] based on the average directional information, which is obtained from complex wavelet coefficient which is derived from shift invariance and multi-directional selectivity, then the inter-coefficient product is computed to indicate the feature orientation and feature strength. Another metric[1] advocated that image understanding mainly caused by the distortions on primary visual information; hence they computed the quantities of the decisive visual information and residual uncertainty to predict the subjective quality, and Zhang presented a RR IQA by calculating the discrepancy of Statistics in edge region. A metric based on the RDCT extracted the feature by calculating city-block distance, mutual information and the ratio of different frequency information shows excellent performance. In 2011, a metric named RR-PCA[10] was developed, it transformed pixel information into a new domain which represents the original data solely in terms of the eigenvectors. The mean gradient values were calculated from the transformed data using edge detection methods based on the Sobel-operator and utilized this feature to measure the degree of image distortion.

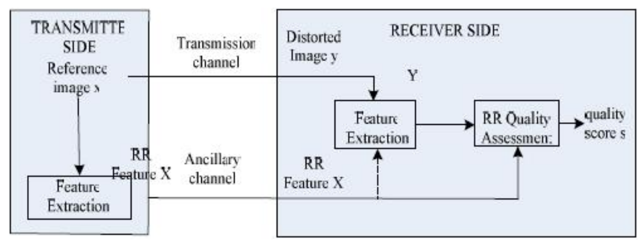


Figure 1 : The framework for the deployment of RR IQA system

The remainder of this paper is organized as follows. In part two, we give the detailed explanation of the proposed metric, and then extensive experimental invalidation and related analysis is followed in part three. Ultimately, conclusions are drawn in part four.

PROPOSED RR IQA INDEX

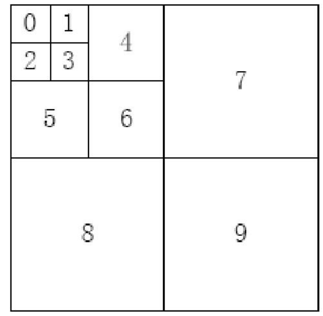
Reorganization of DCT

As far as the representation of an image is concerned, the luminance of image is more decisive compared with chrominance. Consequently, we convert images into luminance space. Spatial information of signal can embody the intensity and some other important properties, but it can not reflect the frequency information being essential for signal analysis. As a result, we decompose an image into a three-level coefficient tree, and the specific steps are shown as below

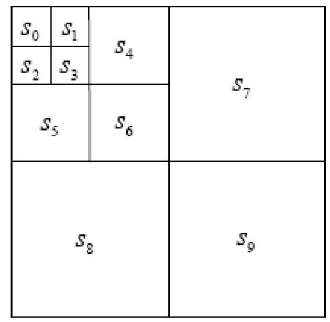
An image with the size of $M \times N$ (M stands for its height, N stands for its width) can be divided into $g \times h$ non-overlapping blocks whose size is 8×8, g is equal to $floor(\frac{M}{8})$, h is equal to $floor(\frac{N}{8})$. The rear columns and rows whose number is less than 8 can be eliminated for the purpose of calculation convenience, because, to some degree, this loss do not impact the overall image understanding at all. An individual DCT coefficient block is decomposed into ten sub-bands manually, as shown in Figure 2(a). For the sub-bands 0, 1, 2 and 3, each sub-band only contains one DCT coefficient. For the sub-bands 4, 5 and 6, each sub-band contains a 2×2 DCT coefficient matrix. And for the sub-bands 7, 8 and 9, each sub-band contains a 4×4 DCT coefficient matrix. After the decomposition, the same sub-bands of all the 8×8 blocks are grouped and organized together according to their corresponding coordinate. Ultimately, the block-based DCT coefficients are reorganized into three-level coefficient tree. In Figure 2(b), S_n indicates the grouped sub-band whose unit is separated from an individual 8×8 coefficient matrix. Figure 2(c) shows the original luminance picture of 'womanhat', and the RDCT representation of this picture is illustrated in Figure 2(d). It can be clearly observed that the RDCT representation appears like a wavelet representation, which shows the analogous structure that the coefficients on the top left corner imply image's low frequency, when the distance adjacent to the row right corner gets closer, which indicates the information of image high frequency is ascending. Moreover, The RDCT representation is more efficient for the RR metrics design than wavelet analysis.

Feature extraction and fusion

From the analysis presented above, we decompose an image into RDCT domain, and assume sub-bands S_0, S_1, S_2 and S_3 as a collective which mainly embodies the low frequency information of an image. Here S_0 stands for the image's direct current (DC) information and the most basic structure of the image is presented by this coefficient matrix, coefficient matrixes S_1, S_2, S_3 stand for low frequency information. Similarly, we assume sub-bands S_4, S_5, S_6 and sub-bands S_7, S_8, S_9 as two separated collectives, they embody the image's median frequency and high frequency information respectively. From another perspective, the sub-

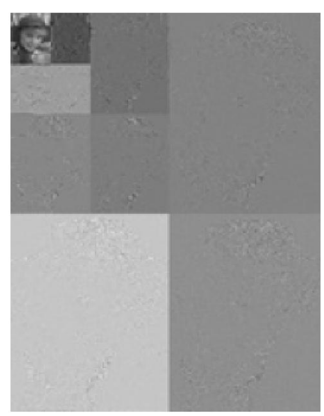


(a)



(b)

(c)



(d)

Figure 2 : (a) 8×8 block-based DCT coefficient matrix; (b) Reorganized DCT decomposition sub-bands; (c) Original luminance image of 'womanhat'; (d) RDCT representation of original image

bands S_1, S_4, S_7 stand for the image's horizontal information and sub-bands S_2, S_5, S_8 stand for the image's vertical information. Psychological studies provide solid evidence in favor of the equal importance of these two orientation information, hence horizontal and vertical information can be combined appropriately. Sub-bands S_3, S_6, S_9 stand for the image's diagonal information and also play an important role in the internal representation of image. We define the energy for an individual sub-band as

$$e_k = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} \log_2(|c|+1) \quad (1)$$

where k denotes the reorganized sub-band index ranging from 0 to 9, e_k presents energy information of the k^{th} sub-bands respectively; c denotes the coefficient in the k^{th} sub-band, and m,n denotes the height and width of the corresponding sub-band respectively. In other words, the energy information of an image can be interpreted by the sub-bands in the RDCT domain indirectly. The distance of two corresponding sub-bands can be defined as

$$\begin{aligned} d_k &= \left|e_k^{org} - e_k^{dis}\right| \\ &= \left|\frac{1}{m \times n}\sum_{i=1}^{m}\sum_{j=1}^{n}\log_2(|c^{org}|+1) - \frac{1}{m \times n}\sum_{i=1}^{m}\sum_{j=1}^{n}\log_2(|c^{dis}|+1)\right| \end{aligned} \quad (2)$$

where e_k^{org} and e_k^{dis} denote the energy of reference and distorted images in the k^{th} sub-band respectively. Consequently, we can extract ten features that represent the change of image energy information.

The problem that lies ahead is how best to fuse those ten features to a single feature as the ultimate evaluation criterion. It is a truth that the sub-band S_0 embodies the most fundamental information of image's structure. What's more, if a reference image is not distorted to the extent that we can hardly distinguish, we can assume these distorted images still earn high similarity in the sub-band S_0. Based on this fact, we can eliminate the miner effect of this sub-band. As mentioned above, human eyes present similar sensitivity to the implication of the diagonal and vertical information in an image, what is more, S_1,S_2 are in the same collective representing image's low frequency information. We define m_1 as the fusion of the energy information of sub-band s_1 and s_2.

$$m_1 = \frac{e_1 + e_2}{2} \quad (3)$$

Here we define the first effective feature as

$$f_1 = \left|m_1^{org} - m_1^{dis}\right| = \left|\frac{e_1^{org} + e_2^{org}}{2} - \frac{e_1^{dis} + e_2^{dis}}{2}\right| \quad (4)$$

Then we define f_2 as the second feature by calculating the distance of the fourth sub-band's energy of a reference and its distorted edition.

$$f_2 = | e_3^{org} - e_3^{dis} | \quad (5)$$

Continuously, we analyze the distance of the next three sub-bands energy for different images, these three features is demonstrated that the efficiency to predict subjective quality is low in comparison with that of the others. For the sack of feature's transmission efficiency, we eliminate this energy information of these three sub-bands. We fuse the rest three sub-bands in high frequency domain into two features. The distance of average energy of sub-band S_7 and S_8 denoted as f_3 acts as the fusion of these two features

$$f_3 = | m_2^{org} - m_2^{dis} | = | \frac{e_7^{org} + e_8^{org}}{2} - \frac{e_7^{dis} + e_8^{dis}}{2} | \quad (6)$$

We define f_4 which measures the distance of sub-band S_9 for a reference and its distorted one as the last feature.

$$f_4 = | e_9^{org} - e_9^{dis} | \quad (7)$$

Thus, we get a feature vector $\vec{F}$ consisting of f_1,f_2,f_3, f_4, and it can be denoted as $\vec{F} = \{f_1,f_2,f_3,f_4\}$.

Ultimately, we define Q as the overall objective quality, and it's given by

$$Q = \sum_{i=1}^{4} \lambda_i f_i \quad (8)$$

where λ_1 is the weight factor, which can be denoted as $\lambda = \{\lambda_i \mid \sum_{i=1}^{4} \lambda_i = 1\}$. Obviously, the higher λ_1 is, the more importantly the i^{th} feature f_i plays on perceptual quality. For white Gaussian noise distorted image, the addition of noise obeying gauss distribution evokes the increase of image's high frequency, consequently, the value of λ_4 will be comparatively bigger. And for images losing low frequency information, the value of λ_1 and λ_2 will be bigger.

FULL PAPER

EXPERIMENTAL RESULTS

Database and performance assessment criteria

We conducted experiments on the largest and very authoritative image quality database, namely Laboratory for Image and Video Engineering (LIVE) image database. LIVE database contains five different distortion types including JPEG, JPEG2000, additive white Gaussian noise (WN), Gaussian blur (GBLUR) and bit errors due to the transmission of JPEG2000 image in a fast fading channel (FF). There are 779 distorted images all across all the distortion categories in total. Each distorted image has a subjective index, called DMOS, which is derived from subjective experiments. It can be considered that these subjective scores are absolutely correct. As a result, the measurement of objective assessment performance is to measure the consistency between objective quality and subjective quality. We follow the performance evaluation procedure introduced in the Video Quality Experts Group (VQEG) HDTV test. Here we take v_i as the representative of perceptual quality index of the i^{th} distorted image. The four parameter $\{b_1,b_2,b_3,b_4\}$ monotonic logistic function is employed to map x_i into v_i

$$v_i = \frac{b_2 - b_2}{1+e^{\left(-\frac{x_i - b_3}{|b_4|}\right)}} + b_2 \tag{9}$$

where x_i denotes the objective quality, v_i is the predicted subjective value. The corresponding four parameters are confirmed by minimizing the sum squared differences between v_i and the original subjective DMOS. After the nonlinear mapping, we use the following three evaluation criteria: (1) Pearson linear correlation coefficient (CC), which evaluates the accuracy of predicted quality; (2) Spearman rank-order correlation coefficient (SROCC), which evaluates the prediction monotonicity; (3) root mean square error (RMSE), which measures the difference between the realistic subjective value and the predicted subjective quality after nonlinear regression.

As mentioned in the previous part, different types of distortion is equipped with its own weighting parameters, the specific allocation of these parameters is that the weights for JP2K form λ_1 to λ_4 are 0.6,0.1,0.3,0; weights for JPEG images are 0,0.8,0.2,0; weights for WN are 0,0,0.1,0.9; weights for GBLUR images are 0,1,0,0 and as to FF distortion, the weights are 0.2,0.7,0.1,0.

TABLE 1 : Performance of IQA indices on live database

Distortion Type	Metrics	RR					FR	
	Criteria	Proposed	RR-PCA	RR-VIF	RR-RDCT	RR-Edge	SSIM	SVD
JPEG2000	CC	0.9433	0.9340	0.9320	0.8983	0.9300	0.9481	0.9110
	SROCC	0.9381	0.9121	0.9500	0.8912	0.9230	0.9432	0.8949
	RMSE	5.3758	9.3802	5.8110	11.0511	9.2800	5.1509	10.72
JPEG	CC	0.9458	0.9040	0.8951	0.9528	0. 8760	0.9346	0.9088
	SROCC	0.9089	0.8471	0.8851	0.9520	0.8510	0.9075	0.8221
	RMSE	5.1917	17.6201	7.1482	9.7660	7.7110	5.6882	19.07
White Noise	CC	0.9702	0.9802	0.9570	0.9275	0.9260	0.9684	0.9751
	SROCC	0.9656	0.9810	0.9461	0.9093	0.9160	0.9565	0.9738
	RMSE	3.8663	5.3301	4.6582	10.4711	10.580	3.9844	6.5600
Gaussian Blur	CC	0.9634	0.7773	0.9501	0.9459	0.9530	0.8969	0.4809
	SROCC	0.9643	0.6694	0.9614	0.9525	0.9530	0.9144	0.4492
	RMSE	4.2147	13.8420	4.6601	5.8801	5.6220	6.9818	17.15
Fast Fading	CC	0.9360	0.9230	0.9442	0.9437	0.9220	0.9527	0.8230
	SROCC	0.9347	0.9260	0.9411	0.9204	0.9160	0.9525	0.8240
	RMSE	5.7861	14.7321	5.4241	9.2770	11.030	5.0014	10.72

Performance comparison

We make a comparison between the proposed RR IQA over individual distortion with some excellent metrics in LIVE image database. On the whole, TABLE 1 presents that the proposed method perform better (in three aspects) than five RR schemes and two FR schemes. We conclude our proposed metric correlates well with the subjective quality. The change of energy in RDCT domain has been greatly proved to have high efficiency in this study. For the two FR schemes, we can observe that the metric[12] based on Singular Value Decomposition (SVD) perform very well on the white noise distortion, and comparatively inferior in other distortion types, and SSIM (Wang, Z.and A. C. Bovik, 2004) excels on the white noise distorted images as SVD does, the reason for this cause is that the addition of white noise to reference will evoke more obvious change of image statistical property in a local region and the core idea of SSIM is to measure the change of image structure realized by basic local statistical property. Obviously, the proposed metric is superior to the two *FR-IQA*s which are not practical in the context of our largest application of image streaming over wireless networks. *RR-PCA*[10] is observed that its performance shows outstanding results for the evaluation of JPEG2000 and FF distortion, especially for white noise. *RR-RDCT* earns an extremely satisfactory performance on JPEG that is constantly difficult to be evaluated accurately in most circumstances. *RR-PCA*[1] and *RR-Edge* metrics also fail to outperform the proposed algorithm.

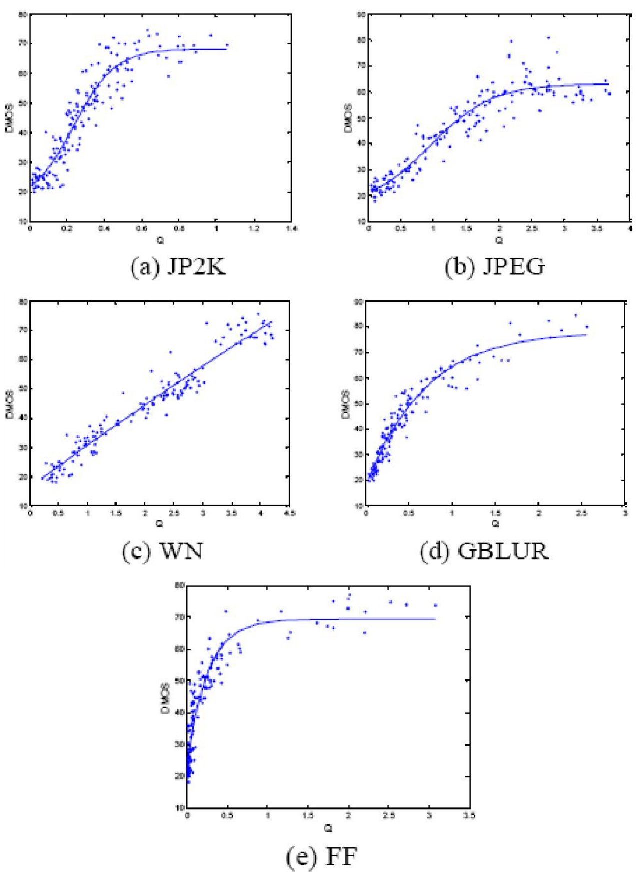


Figure 3 : Scatter plots of DMOS versus objective quality over the five categories of distortions

Scatter plots shown in Figure 3 demonstrate our proposed metric is highly consistent with perception quality over the five distortion types, and it solidly demonstrates that the proposed methods exhibit highly competitive performance in most cases concerning the specific distortion types.

CONCLUSION

In this paper, the change of energy in RDCT domain is introduced to develop a novel RR IQA method, Experimental result on LIVE database shows the proposed metric exhibits satisfactory correlation with subjective evaluation over a wide range of distortion types. Above all, the complexity of the proposed metric and the amount of feature information system needs is superior to most of the existing RR metrics However, its function is still deserved to be studies further in terms of the specific impact it plays on Human Visual System. The change of energy in RDCT domain will be further applied to the research of NR IQA and video quality assessment.

ACKNOWLEDGEMENTS

The research work was supported by National Natural Science Foundation of China under Grant Nos. 61271270, 61071120, 61111140392.

REFERENCES

[1] J.Wu, W.Lin, G.Shi; Reduced-reference Image Quality Assessment with Visual Information Fidelity. IEEE Transactions on Multimedia, online, 1-14 **(2013)**.

[2] M.Zhang, H.Fujita; Local Statistics Feature for

Reduced Reference Image Quality Assessment, SPIE-IS&T/Vol.8660 86600L-1, May, **(2013)**.

[3] L.Ma, S.Li, K.N.Ngan; Reduced-reference image quality assessment in reorganized DCT domain. Signal processing: image communication, 1-19 **(2012)**.

[4] Q.Li, Z.Wang; Reduced-Reference image quality assessment based on multiscale geometric analysis. IEEE Transactions on Image Processing, **18**, 1409-1432 **(2009)**.

[5] A.Liu, W.Lin, M.Narwaria; Image quality assessment based on gradient similarity. IEEE Transactions on Image Processing, **21**, 1500-1512 **(2012)**.

[6] R.Soundararajan, A.C.Bovik; RRED Indices: Reduced Reference Entropic Differencing for Image Quality Assessment. IEEE Transactions Image on Processing, **21**, 517-526 **(2012)**.

[7] Final report from the video quality experts group on the validation of objective models for video quality assessment II, Video Quality Expert Group (VQEG). [Online].Available: http://www.vqeg.org.

[8] Z.Wang, A.C.Bovik; Image quality assessment: From error visibility to structure similarity. IEEE Transactions on Image Processing, **13,** 600-612 **(2004)**.

[9] M.Zhang, W.Xue, X.Mou; Reduced reference image quality assessment based on statistics of edge. Proceedings of SPIE, Digital Photography VII, San Francisco, California, USA, **7876**, **(2011)**.

[10] M.Uzair, D.Fayek; Reduced Reference Image Quality Assessment using Principal Component Analysis. IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, July, 8-10, 2011, 1-6 **(2011)**.

[11] I.P.Gunawan, M.Ghanbari; Reduced-reference video quality assessment using discriminative local harmonic strength with motion consideration. IEEE Transactions on Circuits and Systems for Video Technology, **18**, 71-83 **(2010)**.

[12] A.Shnayderman, A.Gusev, A.M.Eskicioglu; An SVD-based Grayscale Image Quality Measure for Local and Global Assessment. IEEE Transactions on Image Processing, **15**, 442-429 **(2006)**.

[13] H.R.Sheikh, A.C.Bovik, G.D.Veciana; An information fidelity criterion for image quality assessment using natural scene statistics. IEEE Transactions on Image Processing, **14**, 2117-2128 **(2005)**.

[14] Z.Wang, E.P.Simoncelli; Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In Proc. SPIE Human Vision and Electronic Imaging, **5666**, 149-159 **(2005)**.

[15] H.R.Sheikh, Z.Wang, L.Cormack; LIVE Image Quality Assessment Database. [Online]. Available: http://live.ecce.utexax.edu/research/quality, **(2003)**.

[16] Q.Li, Z.Wang; RR image Quality assessment using divisive normalization-based image representation. IEEE Transactions on Image processing, **3**, 202-211 **(2009)**.

[17] Z.Lin, J.Tao, Z.Zheng; Reduced-reference image quality assessment based on average directional information. 2012 IEEE International conference Signal Processing (ICSP), 787-791 **(2012)**.