

Image Quality Assessment Based on Multi-scale Geometric Analysis

Mingna Liu¹, Xin Yang¹, and Yanfeng Shang²

¹ Institute of Image Processing and Pattern Recognition,
Shanghai Jiaotong University, P.R. China

² Department of Electronics and Informatics, Vrije Universiteit Brussel,
IBBT, 1050 Brussel, Belgium
{mingnal,yangxin}@sjtu.edu.cn, yshang@etro.vub.ac.be

Abstract. A novel objective full-reference image quality assessment metric based on Multi-scale Geometric Analysis (MGA) of contourlet transform is proposed. Contourlet transform has excellent properties for image representation, such as multiresolution, localization and directionality, which are the key characteristics of human vision system. Utilizing multiresolution and directionality of MGA, we extract the distortion of structural information from different vision scale and edge direction. The degradation of image quality is evaluated based on the defined energy of structural distortion. Performance experiments are made on professional image quality database with five different distortion types. Compared with some state-of-the-art measures, the results demonstrate that the proposed method improves accuracy and robustness of image quality prediction.

Keywords: Image quality assessment, contourlet transform, image structure.

1 Introduction

Image Quality Assessment (QA) is of fundamental importance in image processing systems, such as image acquisition, compression, enhancement, de-noising, reproduction, and etc.. Generally the approaches for image QA can be categorized into two main classes: subjective and objective ways. The former surveys the opinion of human observers through subjective tests, where a group of persons evaluate several corrupted images, with appropriate criteria, methodologies and hardware. However, it is difficult to implement the time-consuming subjective evaluation in many cases, for example in real-time system. Thus, in this paper, we are focusing on the objective image QA which intends to automatically quantify image degradation in a perceptual manner.

Current approaches to quantifying the visual fidelity of a distorted image can roughly be divided into the following paradigms: 1) metrics based on the statistics of pixels: they assess pixel-based brightness difference between distorted and reference images, such as Mean-Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR). They are attractive image QA measures due to the simplicity and

mathematical convenience. However, they present poor correlation with subjective ratings for ignoring the inter-dependence among pixels[6]. 2) metrics based on Human Vision System (HVS): these approaches are constructed on the mathematical models of HVS[1]. However, although they are mostly accepted, the limited understanding of HVS and the complexity of simulation models keep them from going much further. 3) metrics based on overarching principles: this kind of metrics capture structural feature or information loss from distorted images, which are believed to be the HVS preferred top-down characteristics[2,3,4,5]. Among them, the Structural SIMilarity (SSIM) index attracts a lot of attention for its first puts forward the assumption that *human visual perception is highly adaptive for extracting structural features from natural scene*[2]. Furthermore, Considering the dependence of image quality on viewing conditions (such as viewing distance and display resolution), in [7] the single-scale SSIM index is developed into a multi-scale method. In light of this, we intend to implement a multi-scale image quality measure to deal with structural information change and provide a better approximation to perceived image distortion.

Recently, a theory for high dimensional signals called Multi-scale Geometric Analysis (MGA) has been developed. Several MGA tools were proposed, such as ridgelet[8], curvelet[10], and contourlet[9,11]. Those methods present better directional sensitivity than wavelet transform, i.e. they possess *anisotropy*, so that they could describe the edge feature better, then adaptive to HVS. So in our scheme for structural information extraction, except multiresolution, the direction information is another consideration.

In this paper, we propose a Multi-Scale Directional Difference (MSDD) metric utilizing contourlet transform, capturing the directional structure features in different scales. Firstly the reference and distorted images are decomposed into three level and up to twenty subbands. Every subband represents the image structure of different scale and direction. We evaluate the image perception information loss by calculating the difference between every corresponding subbands of distorted and reference images. The sum up of all these difference are defined as the Energy of Structural Distortion(ESD). Finally, the image quality score is obtained by applying Weber-Fechner law to ESD. Compared with other image QA metric, the validation experiment displays that the proposed metric is a novel and successful scheme for objective image quality predicting.

The organization of this paper is given as follow. The contourlet transform is introduced in section 2. The definition and characteristics of MSDD metric for image QA is presented in section 3. In section 4, the performance of the proposed image quality metric is validated through subjective evaluation experiment with five distortion types. Section 5 concludes the study.

2 Contourlet Transform

In 2001, contourlet was pioneered by Do and Vetterli as the latest MGA tool[11]. Contourlet is a true 2-D sparse representation for 2-D signals like images. The overall result of contourlet transform is a sparse image expansion expressed by

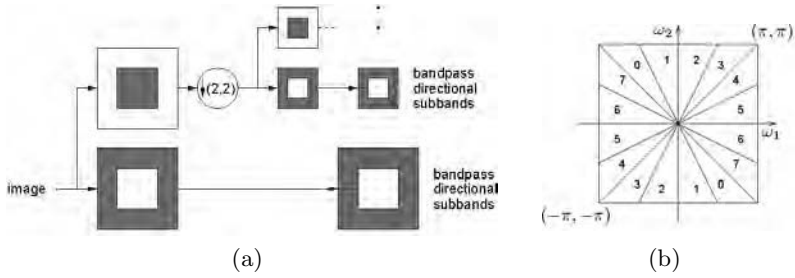


Fig. 1. The scheme of contourlet transform. (a) The block diagram; (b) The frequency partition.

contour segments, so it can capture 2-D geometrical structures in visual information much more effectively than traditional multiresolution analysis methods.

Contourlet can be described as Pyramidal Directional Filter Bank (PDFB), which include the Laplacian Pyramid (LP) and the Directional Filter Bank (DFB). In each scale decomposition, the Laplacian pyramid is used to capture the point discontinuities, and then followed by the DFB to link point discontinuities into linear structures, so that directional information can be captured with a rich set of basis oriented at various directions and scales. This allows contourlets to efficiently approximate a smooth contour at multipleresolutions and multidirections. Contourlet can effectively capture the intrinsic contours and edges in natural images whose traditional multiresolution analysis methods are difficult to handle. Fig. 1 depicts the block diagram of contourlet decomposition and frequency partition of contourlet transform. Here 8-direction directional decomposition is applied in the finest scale, subbands 0-3 correspond to the mostly horizontal directions, while subbands 4-7 correspond to the mostly vertical directions. Contourlet offers a much richer subband set of different directions and shapes, which helps to capture geometric structures in images much more efficiently. In the frequency domain, the contourlet transform provides a multiscale and directional decomposition.

Specifically, let $a_0[\mathbf{n}]$ be the input image. The output after the LP stage is J bandpass images $b_j[\mathbf{n}]$, $j = 1, 2, \dots, J$ (in the fine-to-coarse order) and a lowpass image $a_J[\mathbf{n}]$. That means, the j -th level of the LP decomposes the image $a_{j-1}[\mathbf{n}]$ into a coarser image $a_j[\mathbf{n}]$ and a detail image $b_j[\mathbf{n}]$. Each bandpass image is further decomposed by an l_j -level DFB into 2^{l_j} bandpass directional images $c_{j,k}^{l_j}[\mathbf{n}]$, $k = 0, 1, \dots, 2^{l_j} - 1$. Thus the whole decomposition can be represented as

$$\begin{aligned} a_0[\mathbf{n}] &\rightarrow \{a_J[\mathbf{n}], b_1[\mathbf{n}], b_2[\mathbf{n}], \dots, b_J[\mathbf{n}]\} \\ b_j[\mathbf{n}] &= \{c_{j,0}^{l_j}[\mathbf{n}], c_{j,1}^{l_j}[\mathbf{n}], \dots, c_{j,2^{l_j}-1}^{l_j}[\mathbf{n}]\} \end{aligned} \quad (1)$$

3 Multi-scale Directional Difference (MSDD) Metric

Natural image signals are highly structured. They contain intrinsic geometrical structures that are key features in visual information[11]. Thus, a good image QA

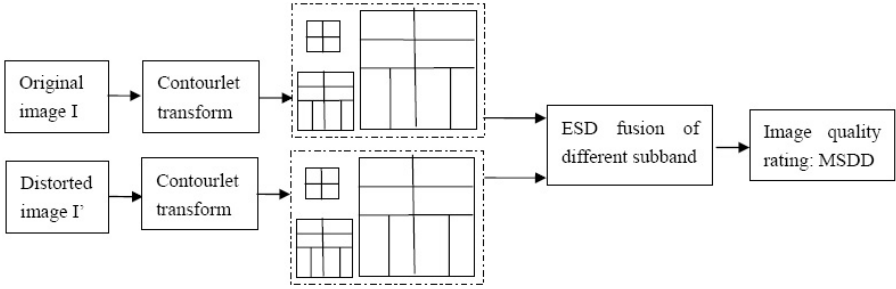


Fig. 2. The block diagram of MSDD

metric should capture the key structural features and present image fidelity ratings well coherent with subjective scores. Contourlet can offer better anisotropy, multiresolution, directionality and localization properties for 2-D signals than existing image representation methods, so better image QA performance can be expected. In this paper, considering the direction and texture information, we develop a new set of characteristics of the structural information by contourlet transform, extracting the directional structure feature in different scales, named Multi-Scale Directional Difference (MSDD) Metric. The block diagram is illustrated in Fig 2.

Firstly, taking the original and distorted images as the inputs, the system iteratively applies a LP decomposition on the images and generated lowpass images and bandpass images. The bandpass images are decomposed by DFB to achieve the multidirectionality. This step finish the extraction of multi-scale structural feature as well as direction and texture information. Secondly, the difference between the subbands for distorted and original images is calculated individually, quantifying the Energy of Structural Distortion (ESD) for the distorted image. Thirdly, the ESD of lowpass and bandpass images is fused weighted. Finally, employing the Weber-Fechner law[12], objective image quality score of our measure is obtained.

Suppose the original and distorted images are I and I' , with size $P \times Q$. They are decomposed by contourlet transform and represented as:

$$\begin{aligned} I[\mathbf{n}] &\rightarrow \{a_J[\mathbf{n}], b_1[\mathbf{n}], b_2[\mathbf{n}], \dots, b_J[\mathbf{n}]\} \\ b_j[\mathbf{n}] &= \{c_{j,0}[\mathbf{n}], c_{j,1}[\mathbf{n}], \dots, c_{j,2^{t_j}-1}[\mathbf{n}]\} \end{aligned} \tag{2}$$

$$\begin{aligned} I'[\mathbf{n}] &\rightarrow \{a'_J[\mathbf{n}], b'_1[\mathbf{n}], b'_2[\mathbf{n}], \dots, b'_J[\mathbf{n}]\} \\ b'_j[\mathbf{n}] &= \{c'_{j,0}[\mathbf{n}], c'_{j,1}[\mathbf{n}], \dots, c'_{j,2^{t_j}-1}[\mathbf{n}]\} \end{aligned} \tag{3}$$

ESD of the lowpass and bandpass subbands for distorted image are:

$$ESD_J = \frac{1}{M \times N} \left(\sum_{m=1}^M \sum_{n=1}^N (a_J(m, n) - a'_J(m, n))^2 \right)^{\frac{1}{2}} \tag{4}$$

$$\text{ESD}_{j,k} = \frac{1}{W_{j,k} \times H_{j,k}} \left(\sum_{w=1}^{W_{j,k}} \sum_{h=1}^{H_{j,k}} (c_{j,k}(w, h) - c'_{j,k}(w, h))^2 \right)^{\frac{1}{2}} \quad (5)$$

where $M \times N$ is the size of decomposed lowpass subband, $W_{j,k} \times H_{j,k}$ is the size of decomposed bandpass subband in j -th level and k -th direction ($j = 1, 2, \dots, J$; $k = 0, 1, \dots, 2^{l_j} - 1$).

In the ESD fusion step, the weights of lowpass and bandpass subbands for ESD fusion deserve some consideration. In our method, to balance the contribution of different frequency bands, we utilize the subband size to define the weights:

$$\begin{aligned} D &= \langle \mathbf{w} \cdot \mathbf{ESD} \rangle \\ &= w_J * \text{ESD}_J + \sum_{j=1}^J \sum_{k=0}^{2^{l_j}-1} (w_{j,k} * \text{ESD}_{j,k}) \end{aligned} \quad (6)$$

where w_J and $w_{j,k}$ are the weights for the distortion of lowpass and bandpass subbands in j -th level and k -th direction:

$$w_J = \frac{P \times Q}{M \times N}, w_{j,k} = \frac{P \times Q}{W_{j,k} \times H_{j,k}} \quad (7)$$

Based on the Weber-Fechner law [12], the relationship between the physical magnitude of stimuli and the perceived intensity of the stimuli is logarithm, the quality measurement is defined in terms of "visual decibels", coincident with the traditional error metrics[13]:

$$MSDD = \log_{10}(1 + D), \quad D > 0 \quad (8)$$

where the constant 1 is included to avoid negative value of quality score.

MSDD models any distortion as the difference of ESD and measure the difference between the original and distorted images by the decomposing processing. Here the structural vector a_j and b_j are desired to describe the shape, edge, texture and others by themselves. The difference of structural information, which are gained by computing the ESD, are desired to represent the variation of the images' degradation along different scales and directions. Then we compare the difference of MSDD values to measure the distortion magnitude. The typical MSDD values range between 0 and 1.3. The actual value is meaningless, but the comparison between two values for different test images gives one measure of quality. The lower the predicted score of MSDD is, the better the image quality is. When $a_J = a'_J$ and $b_J = b'_J$, the distorted and the original images are identical, so MSDD = 0. The metric satisfies the following conditions:

1. symmetry: $\text{MSDD}(x,y) = \text{MSDD}(y,x)$;
2. boundedness: $\text{MSDD}(x,y) \geq 0$;
3. unique minimum: $\text{MSDD}(x,y) = 0$ if and only if $x = y$.

4 Experimental Results and Discussion

4.1 Performance in Predicting Visual Fidelity

In this section, the performance of the MSDD metric is analyzed in terms of its ability to predict image quality fidelity in a manner that agrees with subjective



Fig. 3. Some original images in LIVE database

ratings. Specifically, the professional LIVE image database is utilized[15]. The LIVE image database consists of 29 original 24-bits/pixel color images, and 779 distorted images. Five subdatabases with different types of distortions are tested: 1) JPEG2000 compression(JP2K), 2) JPEG compression(JPEG), 3) Gaussian white noise(WM), 4) Gaussian blurring(GBlur), and 5) Rayleigh-distributed bit-stream errors of a JPEG2000 compressed stream(FF). The distortion varies at a broad range of quality, from imperceptible levels to high levels of impairment. The subjective ratings are obtained from about 25000 individual human quality judgments. Fig 3 shows some original images in LIVE database.

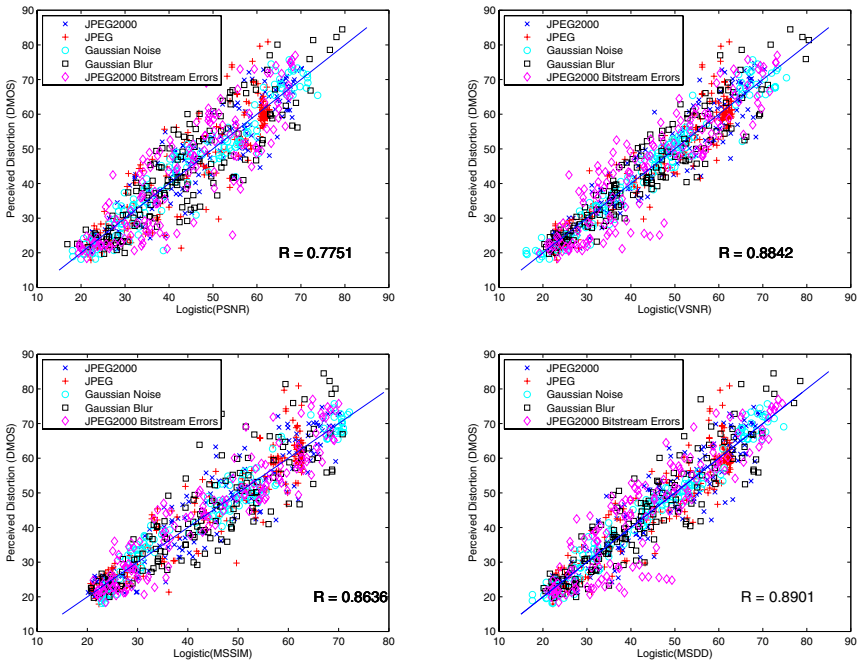


Fig. 4. Subjective ratings of perceived distortion for the 779 images of the LIVE database plotted against predicted values from each of the four metrics. In all graphs, the vertical axis denotes perceived distortion (Difference Mean Opinion Score) as reported by subjects[6]. The horizontal axes correspond to the normalized four metrics' outputs.

For comparison, these same sets of images from LIVE are analyzed by PSNR, Mean SSIM (MSSIM)[2] and Visual Signal-to-Noise Ratio (VSNR) metric[5]. In MSDD, considering the computation cost, we choose the decomposing scale $J = 3$, the decomposing direction $l_j = 3$, $for j = 1, 2$. MSSIM is computed on filtered and down sampled versions of the images which the downsampling factor is chosen based on the height of each image as described at [14]. VSNR are applied using their default implementations provide at [5]. After a logistic transformation for the outputs normalization[6], Some validation criteria exploited by Video Quality Experts Group (VQEG) for objective quality metric validation are utilized for our experiments[16], such as linear Correlation Coefficient, Root-Mean-Square-Error(RMSE) and Spearman Rank Order Correlation Coefficient (SROCC). And for LIVE database, the four metrics are applied to grayscale versions of the images which are obtained via a pixel-wise transformation of $I = 0.2989R + 0.5870G + 0.1140B$, where R , G and B denote the 8-bit grayscale, red, green and blue intensities, respectively.

The results are listed in Table 1 and Fig 4. For considering the combination of low-level visual property of perceived contrast and the mid-level visual property of global precedence, VSNR shows the best prediction performance on all subdatabases except FF. And MSDD is a strong competitor. It acquires close performance with VSNR across the five distortion types, even better in WN. MSSIM does much better on FF subdatabase than the other three metrics, while much worse in GBlur. PSNR acts well on WN distortion type when errors are statistically independent. While for the other four distortion types, it is not adaptable well. In the cross-distortion validation (all 779 distorted images taken as a whole, much more complicated than subdatabases), PSNR degrades worst for the loss of structural information consideration. MSSIM and VSNR performs better than PSNR while still worsen by a sizable margin. MSDD displays the

Table 1. Correlation Coefficient, RMSE and Rank-Order Correlation Coefficient between subjective ratings and normalized four metrics' outputs

Measure	Metric	JP2K	JPEG	WN	GBlur	FF	All
Correlation Coefficient	PSNR	0.9338	0.8918	0.9578	0.8757	0.8598	0.7751
	MSSIM	0.9367	0.9282	0.9747	0.8739	0.9448	0.8633
	VSNR	0.9517	0.9420	0.9741	0.9371	0.8947	0.8842
	MSDD	0.9369	0.9320	0.9832	0.9119	0.8852	0.8901
RMSE	PSNR	5.7976	7.2352	4.5900	7.5932	8.3995	10.1760
	MSSIM	5.6703	5.9482	3.5668	7.6445	5.3877	8.1264
	VSNR	4.9751	5.3654	3.6135	5.4867	7.3477	7.5230
	MSDD	5.6618	5.7957	2.9125	6.4534	7.6513	7.3397
Rank Order Correlation Coefficient	PSNR	0.9282	0.8703	0.9562	0.8711	0.8661	0.7692
	MSSIM	0.9317	0.9025	0.9630	0.8943	0.9409	0.8510
	VSNR	0.9447	0.9077	0.9745	0.9449	0.8994	0.8834
	MSDD	0.9324	0.8967	0.9813	0.9195	0.8852	0.8928

best capability dealing with the complicated situations. Finally, we can draw the conclusion that MSDD exhibits the better general performance than the other three state-of-the-art measures, showing the robustness and feasibility for image QA.

4.2 Discussion

In fact, for the appealing characteristics of contourlet transform, MSDD avoids the usual limitation of most the other image QA metrics. The decomposing coefficients have the corresponding locations as the spatial locations. So the MSDD can be expanded to be a distortion map to describe the spatially localized fidelity. A primary shortcoming about MSDD is that the metric is limited to grayscale images. An extension of the MSDD metric might include an additional stage which accounts for perceived distortion due to degradations in color. The MSDD metric has relatively high computational complexity. Using a MATLAB implementation, the metric requires for a 512×512 image approximately 1s on a 1.80-GHz Intel Pentium machine.

5 Conclusion

This paper presented a multi-scale directional structural difference metric for quantifying the visual fidelity of natural images. Via multiscale geometric analysis of contourlet transform, the proposed MSDD metric operates based on structure distortion in different scales and directions. Multiresolution, localization and multidirectionality are modeled as subband decomposition of contourlet transform. Finally MSDD is determined based on the weighted sum of the ESD. This metric is well adaptable for five typical distortion types, especially outstanding in cross-distortion validation. MSDD could be extended for colorful image quality evaluation. There should be a lot of work to investigate potential improvements in the future.

References

1. Watson, A.B.: DCT quantization matrices visually optimized for individual images. *Human Vision, Visual Processing, and Digital Display IV* (1993)
2. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error measurement to structural similarity. *IEEE Trans. Image Process* 13(4), 600–612 (2004)
3. Sheikh, H.R., Bovik, A.C., de Veciana, G.: An Information Fidelity Criterion for Image Quality Assessment Using Natural Scene Statistics. *IEEE Trans. Image Process* 14(12), 2117–2128 (2005)
4. Sheikh, H.R., Bovik, A.C.: Image information and visual quality. *IEEE Trans. Image Process* 15(2), 430–444 (2006)
5. Chandler, D.M., Hemami, S.S.: VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images. *IEEE Trans. Image Process* 16(9), 2284–2298 (2007)

6. Sheikh, H.R., Sabir, M.F., Bovik, A.C.: A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms. *IEEE Trans. Image Process* 15(11), 3440–3451 (2006)
7. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multi-scale structural similarity for image quality assessment. In: *Proceedings of the 37th IEEE asilomar conference on signals, systems and computers*, November 9–12 (2003)
8. Cands, E.J.: *Ridgelets: Theory and Applications*. Department of statistics, Stanford University, USA (1998)
9. Do, M.N.: Contourlets and Sparse Image Expansions. *Proceedings of SPIE - The International Society for Optical Engineering* 5207(2), 560–570 (2003)
10. Donoho, D.L., Duncan, M.R.: Digital curvelet transform: strategy, implementation and experiments. In: *Proc. Aerosense 2000, Wavelet Applications VII*, SPIE, vol.4056, pp. 12–29 (2000)
11. Do, M.N., Vetterli, M.: The Contourlet Transform: An Efficient Directional Multiresolution Image Representation. *IEEE Trans. Image Process* 14(12), 2091–2106 (2005)
12. Levine, M.W.: *Fundamentals of sensation and perception*, 3rd edn. Oxford University Press, New York (2000)
13. van den branden Lambrecht, C.J., Verscheure, O.: Perceptual quality measure suing a spatio-temporal model of the huamn visual system. In: Bhaskaran, V., Sijstermans, F., Panchanathan, S. (eds.) *Digital Video Compression: Algorithms and Technologies*, proc. SPIE, San Jose, CA, vol. 2668, pp. 450–461 (1996)
14. SSIM website, <http://www.ece.uwaterloo.ca/~z70wang/research/ssim/>
15. Sheikh, H.R., Wang, Z., Cormack, L., Bovik, A.C.: Image and Video Quality Assessment Research at LIVE, <http://live.ece.utexas.edu/research/quality>
16. VQEG, Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II (August 2003), <http://www.vqeg.org/>