Image Quality Assessment Using Spatial Frequency Component

Guangyao Cao[1], Luhong Liang[2], Siwei Ma[3], and Debin Zhao[1]

[1] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, P.R. China, 150001
[2] Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
[3] Peking University, Beijing, 100872, P.R. China
gycao@vilab.hit.edu.cn, lhliang@jdl.ac.cn, swma@jdl.ac.cn, dbzhao@vilab.hit.edu.cn

Abstract. Image quality assessment (IQA) is a crucial technique in perceptual image/video coding, because it is not only a ruler for performance evaluation of coding algorithms but also a metric for ratio-distortion optimization in coding. In this paper, inspired by the fact that distortions of both global and local information influence the perceptual image quality, we propose a novel IQA method that inspects these information in the spatial frequency components of the image. The distortion of the global information mostly existing in low spatial frequency is measured by a rectified mean absolute difference metric, and the distortion of the local information mostly existing in high spatial frequency is measured by SSIM. These two measurements are combined using a newly proposed abruptness weighting that describes the uniformity of the residual image. Experimental results on LIVE database show that the proposed metric outperforms the SSIM and achieves performance competitive with the state-of-the-art metrics.

Keywords: Image quality assessment, low spatial component, high spatial component, HVS.

1 Introduction

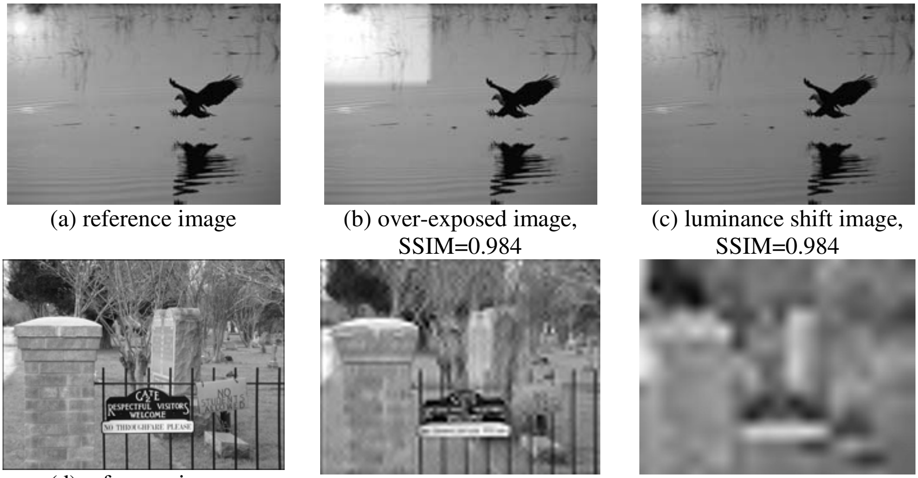
Compression techniques for visual information have been intensively researched for decades. Despite of the traditional algorithms that squeeze the statistical redundancy among pixels, one promising path to improve the coding efficiency is to incorporate the human visual system (HVS) model in the coding framework to remove the perceptual redundancy, namely perceptual image/video coding. In this field, one of the crucial techniques is the perceptual image quality assessment (IQA), which is not only a ruler to evaluate the performance of the perceptual image coding algorithm, but also a metric in optimizing the image coding algorithms in the loop. Since the objective of IQA is to give a score on image quality that can accord with human observers' judgment, the IQA also inherently relies on the characteristics of HVS. However, HVS is

P. Muneesawang et al. (Eds.): PCM 2009, LNCS 5879, pp. 201–211, 2009.

a rather complicated system and the current research on it is still very limited. Therefore, it is very challenging to give an accurate and robust prediction of the quality of an image that is distorted in compression.

The research on IQA has been conducted over 30 years [1]. In the early works, most of the IQA methods focus on the pixel distortion, e.g. the mean-squared-error (MSE), the peak signal to noise ratio (PSNR) and then the visual ability of the image distortions based on root-mean-squared (RMS) are proposed[2][3][4]. Among them, the PSNR metric dominates the image/video processing and coding fields. Recently, with the progress of the research on the HVS [5], IQA methods based on the distortions of local structures rather than pixels attract more and more attentions and a lot of new IQA methods are proposed. In [6], Structure Similarity (SSIM) is proposed to extract and evaluate the distortion of structure information in accord with the HVS. The SSIM computes the contrast, luminance and structure distortion of image, and then pools them to the final score. In most cases, the SSIM can achieve good correlation with subjective scores such as the Different Mean Opinion Score (DMOS), so SSIM becomes the most popular full-reference IQA metric. Lately, some new metrics based on the SSIM are also proposed, such as the SSIM based on different wavelet bands [7][8], the SSIM metric with more image information pooled [9][10], the multiscale modular similarity based method [11], etc. Following the similar strategy to the SSIM, researchers also develop some structure analysis based metrics [12][13][14]. A different strategy in IQA metric design is to focus on the spatial frequency component of the image. For example, in [15], a wavelet-based visual signal-to-noise ratio (VSNR) is proposed to capture the visual distortion threshold and quantize the distortion in an image in wavelet domain. In [16], the original and the distortion image are decomposed by Haar wavelet, and then the perceptual distortion at each pixel is measured by the weighted sum of the frequency mismatches in different bands, finally the whole image distortion is the sum of distortions at pixels. Besides the above two strategies, some IQA metric considers the different types of distortions. For example, in [17], visual information fidelity (*VIF*) method is proposed, where the signal source, the transmission channel and the HVS are simulated and the most salient distortions, i.e. blur and noise are considered in the transmission channel model. It is shown that VIF metric outperforms VSNR while has lower computational complexity.

As mentioned above, most of these metrics only focus on the local structure information [6-14] or the global spatial frequency component [15][16] of the image. However, neither single kind of information can completely reflect the perceptual distortion. Fig.1 (a)~(c) give a failure example of SSIM. It can be seen that both (b) and (c) in Fig.1 archive very high scores (SSIM=0.984), while the subjective quality of Fig.1 (b) is obviously worse than Fig.1(c) due to its over-exposed top left part. The reason is that the distortion caused by the overexposure in Fig.1 (b) mostly exists in low spatial frequency (LSF), however, the SSIM metric excessively focus on the local structure information that mostly exists in the high spatial frequency (HSF). On the other hand, Fig.1 (d)~(f) give a failure example of the VSNR, which focus on the global spatial frequency component. It can be seen that both (e) and (f) in Fig.1 archive very close VSNR scores (16.954 and 16.769 respectively), but their subjective qualities are rather different. The reason is opposite to that of SSIM.



(a) reference image (b) over-exposed image, SSIM=0.984 (c) luminance shift image, SSIM=0.984

(d) reference image (e) blurred image, VSNR=16.954 (f) blurred image, VSNR=16.769

Fig. 1. Failure examples of SSIM and VSNR

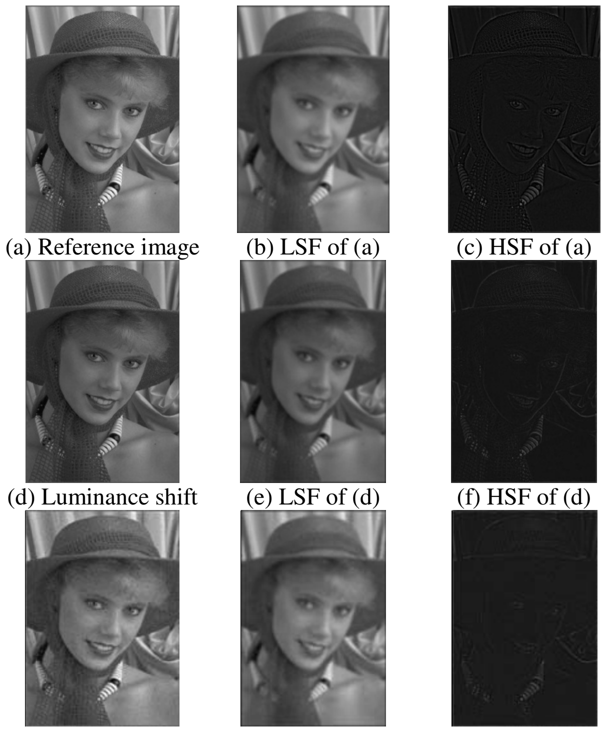
Actually, according to the global-to-local theory (GTL) [18-23], both HSF and LSF are very important in IQA, although they influence the visual effect of image in different ways. Inspired by the above fact, we propose a novel image quality metric (namely *Spatial Frequency Component with Global and Local information SFCGL*) that considers both the global and local distortions existing in different spatial frequency components of the image. The global distortion is measured by the mean absolute difference rectified by a scalar factor and a luminance shifting estimation in the decomposed LSF, and the local distortion is measured by the SSIM metric in the decomposed HSF. Moreover, we define *abruptness* measurement of the HSF and LSF to evaluate the visual influences of the local and global distortion. Finally, the image quality is predicted by the fusion of the global and local information using the weighting of the abruptness measure. Experimental results on the whole *LIVE Database Release 2* [24] show that the proposed metric outperforms the SSIM and achieves performance competitive with the state-of-the-art metrics.

The rest of the paper is organized as follows. Section 2 describes the details of the proposed image quality metric. Section 3 gives the experiment results. Conclusions are in the last section.

2 The Proposed IQA Metric

2.1 Overview

From the HVS perspective, the world is made up of different spatial frequencies [26]. The image quality can be quantified separately in different spatial frequencies, and then the results from different spatial frequencies are fused to the final score. Here we consider the decomposition into LSF and HSF [18]. As shown in Figure 2, the left column shows the reference image and the distortion images with luminance shift and add-in noise respectively. The other two columns are the LSFs and HSFs. From the comparisons of LSF and HSF of the luminance shifted image (d) and image with add-in noise (g) with that of reference image (a), we can see that on one hand, most global distortion such as luminance shift exists in LSF. On the other hand, most local distortion like add-in noise exists in HSF. Therefore the IQA method should take both global distortion and local distortion into account in order to give a reasonable judgment.



(a) Reference image (b) LSF of (a) (c) HSF of (a)
(d) Luminance shift (e) LSF of (d) (f) HSF of (d)
(g) Add-in noise (h) LSF of (g) (i) HSF of (g)

Fig. 2. Spatial frequency components of the image

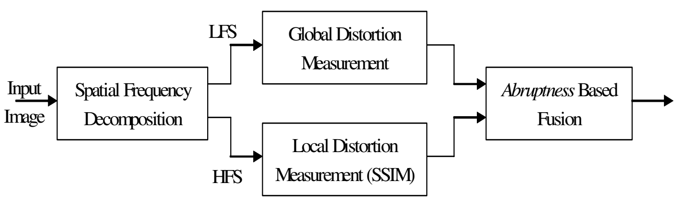


Fig. 3. Framework of SFCGL

Inspired by the above facts, we propose a novel IQA metric, i.e. the SFCGL, which considers both the global and local information distortion of the image. As shown in Fig.3, it consists of four steps: (1) spatial frequency decomposition of the input image into LFS and HFS [18]; (2) measurement of the global information distortion in LFS, where the rectified mean absolute difference is used; (3) measurement of the local information distortion in HFS, where the SSIM metric is adopted; and (4) fusion of the measurements using the *abruptness* measures the residual images of LFS and HFS to predict the observer's judgment.

2.2 Measurement on Global Information

In this section, we describe the rectified mean absolute difference to measure the distortion of the global information in the LSF.

As mentioned in Figure 1, the IQA metrics based on local distortion may fail in cases where there are some distortions globally or in a larger area. To solve this problem, we should measure the global distortion besides the local distortion in IQA. The LSF image reflects the global information of the image. Therefore we evaluate the global distortion in the LSF image

It is well known that some global distortion metrics such as mean absolute deviation (MAE), mean square error (MSE) and peak signal-to-noise ratio (PSNR) perform well in predicting the distortion of some images without many details, although they are heavily criticized for the bad performance in quantifying some images with local distortions such as add-in noise and structure distortion. Since the LSF nearly loses most of the details (local information) of image, we use the idea of MAE in the measurement on global information in LSF. Supposing I_{ij} and R_{ij} are the luminance of the reference image and the distortion image at location (i j), we measure the statistical global distortion RS as follows:

$$RS = MAE^{\lambda}\text{, where } MAE = \frac{1}{N}\sum_{(i,j)\in I} | I_{ij} - R_{ij} |. \tag{1}$$

Here we utilize the MAE to describe the global distortion and use a scale factor λ for normalization.

Since the luminance shift affects the observer's attention, i.e. if the slope of the luminance shift is relatively stable in a block, it draws less attention, and vice versa. Therefore, we rectify the global information measurement using the luminance shift. Considering a block B, we define the block luminance shift LS_{block} as follows:

$$LS_{block} = \sqrt{\sum_{(i,j)\in B} (\Delta_{ij} - \overline{\Delta})^2} \Big/ \overline{\Delta} \tag{2}$$

where $\Delta_{ij} = I_{ij} / R_{ij}$ and $\overline{\Delta}$ is the mean of Δ_{ij}. Then, the image luminance shift LS_I is defined as the sum of block luminance shift of the whole image:

$$LS_I = \sum LS_{block} \tag{3}$$

Before rectifying the statistical global distortion RS, we normalize the image luminance shift LS_I to [0, 1] as follows:

$$CP = 1 - Clip(LS)$$
$$\text{where } LS = \frac{1}{10}\ln(LS_I) \text{ and } Clip(x) = \begin{cases} 0 & x < 0 \\ x & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases} \quad (4)$$

The log function is introduced to reduce the effects of large LS and enlarge the effects of small LS .

Finally, the image quality metric based on global information CRS is described as follow:

$$CRS = (1 - RS) \cdot CP^{\alpha} \quad (5)$$

where parameter α is a consistent between 0 and 1 which is determined experimentally.

2.3 Measurement on Local Information

Since the HSF includes most of the detail information such as structure, edge and noise, we employ the HSF to predict the distortion of local information distortion in the image. Here, we adopt the most successful methods, the SSIM [6], as our measurement on local information.

Assume *x* and *y* are two image patches from the reference and distortion images, the SSIM are defined as follows [6]:

$$SSIM(x, y) = [l(x, y)]^{\beta}[c(x, y)]^{\gamma}[s(x, y)]^{\chi} \quad (6)$$

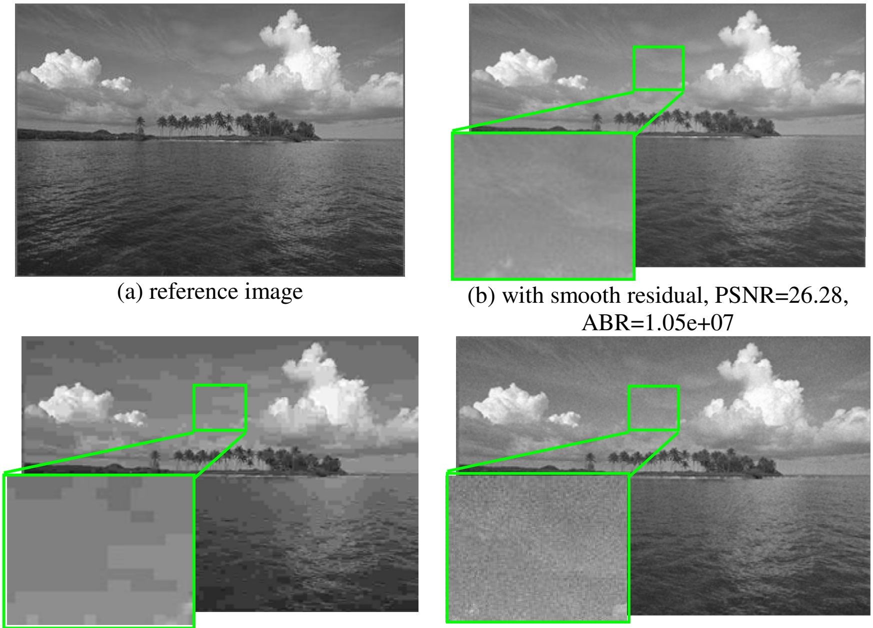
where $l(x, y)$ is the distortion of luminance, $c(x, y)$ is the distortion of contrast, and $s(x, y)$ is the distortion of structure. β, γ and χ are scale factors and all of them are set to 1 in practice.

2.4 Fusion of Global and Local Measurement

In this section, we propose a measurement '*abruptness*' and employ it in fusing the global and local measurements. The *abruptness* measurement is based on the fact that the distortion images with rough residual image usually draw more attention than the one with smooth residual image. Fig.4 shows an example, where the distortion images (b)~(d) have similar PSNR. It can be seen that (c) and (d) have rougher residual image than (b), and they have much worse subjective judgment than (b).

Inspired by the fact that eigevalue can reflect both the average value of matrix and the roughness in the form of proportional, we give the follow equation to quality the different visual attention drawn by different kinds of distortion.

Supposing E_{ij} is the luminance at location (*i,j*) in an residual image block and $\bar{E}$ is the mean luminance in this block, we define the block abruptness as follows:



(a) reference image

(b) with smooth residual, PSNR=26.28, ABR=1.05e+07

(c) with rough residual, PSNR=25.58, ABR=2.28e+07

(d) with rough residual, PSNR=26.18, ABR=2.59e+07

Fig. 4. The different residual image leads to quite different visual effects

$$ABR = \sqrt{S_{\max}(B^T B)}\text{, where } B = (b_{ij}) \text{ and } b_{ij} = E_{ij} - \bar{E}\,. \tag{7}$$

In equation (7), $S_{\max}(\bullet)$ denotes the calculation of the maximum eigenvalue of the matrix. Bigger ABR means the residual image is rougher and has more influence on the visual effect than the images with smaller ABR. In practice, the *abruptness* at each pixel is computed in an 11×11 sliding window, and then the total *abruptness* is the sum of abruptness in all pixels.

The proposed abruptness measurement can quantify the uniformity of the residual image well. As shown in the examples in Fig.4, when the change of residual image is more unevenly, the corresponding abruptness is larger.

Since the LSF or HSF with higher abruptness attracts more attention, we fuse the measurements in LSF and HSF, i.e. the CRS and the SSIM, using the sum weighted by the normalized abruptness measurements of the residual image of LSF and HSF . More specifically, the final image quality metric is defined as:

$$SFCGL = w_{abr} \cdot CRS + (1 - w_{abr}) \cdot SSIM$$

$$\text{where } w_{abr} = \frac{ABR_L}{ABR_L + ABR_H} \tag{8}$$

3 Experiments

In the experiments, we decompose the input image into LFS and HFS via Gauss 10/1.5 filter. This filter is set to extract more detail information and less global information from the reference image, according to the experiment results from different filter (Gauss, Garbor, Motion) with different parameters, we chose this one. The parameters λ (equation (1)) and α (equation (5)) are set to 1/3 and 0.9 respectively. The parameter λ effects the distribution of RS. We use 1/3 as the last parameter according the distribution (normalizing the distribution). The scale factor α in is 0.9, it weight the importance of luminance shift in the image distortion.

We verify the proposed metric using the *LIVE Database Release 2* [24], which consists of five subsets with different distortion types including JPEG compression, JPEG2000 compression, Gaussian Blur, Gaussian noise and Fast fading (Bit error).There are 982 distortion images in total. In experiments we set parameters in equation (of the proposed metric) τ_1=-10163, τ_2=63, τ_3=220 and τ_4=23. For every image quality metric, the logistic function (9) is fitted by Matlab with the function *nlinfit.*

$$f(x) = \frac{\tau_1 - \tau_2}{1 + \exp(-\frac{x - \tau_3}{\tau_4})} + \tau_2 \tag{9}$$

The initial value of the parameters $\tau_1, \tau_2, \tau_3, \tau_4$ are determined following the Video Quality Experts Group (VQEG) report [27].

We follow the guidance of Video Quality Expert Group (VQEG), in which the metric performance is evaluated through three indicators: prediction accuracy, prediction monotonicity and prediction consistency. They are quantified by *Pearson* linear correlation coefficient (CC), *Spearman* rank order correlation coefficient (SROCC) and Outlier Ratio of outlier-points to total points (OR) between subjective DMOS and objective ratings. We only use CC and SROCC in our experiments, because some subsets in the LIVE database do not provide DMOS (Difference Mean Opinion Score) scores of each observer, so that we can not calculate the OR. We also use the root mean square prediction error (RMSE) as an additional performance indicator to evaluate the prediction accuracy.

We compare the proposed SFCGL metric to the SSIM [6], VSNR [14], VIF [15] and PSNR. Table 1 shows the results on the whole LIVE database with 982 images. It can be seen that the proposed metric outperforms the SSIM, VSNR and PSNR according to all the three indicators, and achieves performance very close to the VIF.

Table 1. Performance comparison in the whole LIVE database

	CC	SROCC	RMSE
SSIM	0.9377	0.9256	8.0288
VSNR	0.8818	0.8103	10.9032
VIF	0.9629	0.9763	6.239
PSNR	0.9300	0.9100	8.4843
SFCGL	**0.9469**	**0.9412**	**7.4313**

Table 2. Performance comparison in each subsets of the LIVE database (CC indicator)

	JPEG2000	JPEG	White Noise	Gauss Blur	Fast Fading
SSIM	0.9084	0.9682	0.9527	0.9291	0.9538
VSNR	0.8884	0.8977	0.9273	0.8730	0.8636
VIF	0.9722	0.9628	0.9741	0.9779	0.9616
PSNR	0.9593	0.9516	0.9911	0.9116	0.9456
SFCGL	**0.9733**	**0.9691**	**0.9566**	**0.9533**	**0.9402**

Table 3. Performance comparison in each subsets of the LIVE database (SROCC indicator)

	JPEG2000	JPEG	White Noise	Gauss Blur	Fast Fading
SSIM	0.9718	0.9587	0.9785	0.9388	0.9659
VSNR	0.8103	0.8264	0.8947	0.7914	0.8153
VIF	0.9660	0.9631	0.9908	0.9844	0.9749
PSNR	0.9545	0.9325	0.9915	0.8736	0.9365
SFCGL	**0.9779**	**0.9615**	**0.9833**	**0.9675**	**0.9555**

Table 4. Performance comparison in each subsets of the LIVE database (RMSE indicator)

	JPEG2000	JPEG	White Noise	Gauss Blur	Fast Fading
SSIM	10.1989	6.0721	6.6783	8.0465	10.9946
VSNR	11.2005	10.6891	8.2259	10.6094	11.1379
VIF	5.7129	6.5512	4.9675	4.5425	6.0640
PSNR	6.8887	7.4571	2.9222	8.9389	7.5319
SFCGL	**5.5967**	**5.9983**	**6.4068**	**6.5706**	**7.5257**

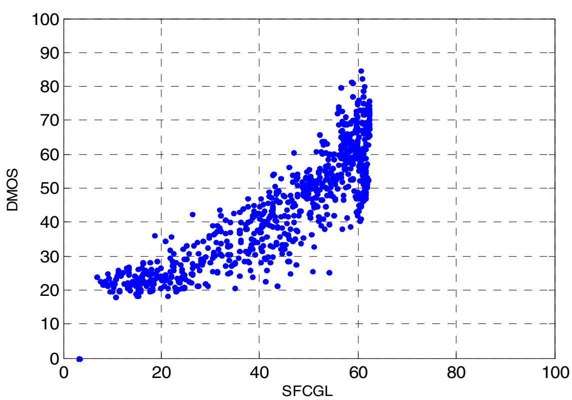


Fig. 5. Subjective ratings of perceived distortion for the 982 images of the LIVE database

Table 2~4 shows the performance comparison in each subset of the LIVE database. It can be seen that the proposed SFCGL metric achieve very good performance in all the subset, which means the SFCGL is rather robust to different kinds of distortions.

Fig.5 shows the scatter plots of the DMOS (scaled to the full range of 1~100) versus the objective prediction (after logistic) by the proposed SFCGL metric.

The experimental results above prove the effectiveness of the proposed SFCGL metric.

4 Conclusion

This paper presented a new perceptual IQA metric based on both global and local information of image. In the evaluation of global information distortion exists in LSF, we rectify the mean absolute difference and add a component to eliminate its poor prediction of luminance shift. In the evaluation of local information distortion exists in HSF, SSIM is adopted. Then the final score is fused by the novel abruptness that describes the uniformity of the residual image to weight the different influence of the global and local information distortion. Experimental results on the *LIVE Database Release 2* show that the proposed SFCGL metric outperforms both the SSIM and VSNR metrics, and achieves performance competitive with the recently proposed VIF metric. Moreover, it also shows that the proposed SFCGL metric is rather robust to different kinds of image distortions.

Acknowledgements

This work is supported by National Basic Research Program of China (973 Program) under contract 2009CB320902 and NSFC under contracts 60833013.

References

1. Wang, Z., Bovik, A.C.: Modern Image Quality Assessment. Synthesis Lectures on Image, Video, and Multimedia Processing (2006)
2. Moulden, B., Kingdom, F.A.A., Gatley, L.F.: The standard deviation of luminance as a metric for contrast in random-dot images. Perception 19, 79–101 (1990)
3. Kingdom, F.A.A., Hayes, A., Field, D.J.: Sensitivity to contrast histogram differences in synthetic wavelet-textures. Vis. Res. 41, 585–598 (1995)
4. Tiippana, K., Näsänen, R., Rovamo, J.: Contrast matching of two dimensional compound gratings. Vis. Res. 34, 1157–1163 (1994)
5. Sheikh, H.R., Bovik, A.C., de Veciana, G.: An information fidelity criterion for IQA using natural scene statistics. IEEE Trans. Image Process. 14(12), 2117–2128 (2005)
6. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: IQA: From error visibility to structural similarity. IEEE Trans. Image Process. 13(4), 600–612 (2004)
7. Yang, C.L., Gao, W.R., Po, L.M.: Discrete wavelet transform-based structural similarity for image quality assessment. In: IEEE international Conference on Image Processing, pp. 377–380 (2008)

8. Li, X.L., Lu, W., Tao, D.C., Gao, X.B.: Frequency structure analysis for IQA. In: IEEE international Conference on System, Man and Cybernetics, pp. 2246–2251 (2008)
9. Moorthy, A.K., Bovik, A.C.: Visual Importance Pooling for Image Quality Assessment. IEEE journal of selected topics in signal processing 3(2), 193–201 (2009)
10. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multi-scale structural similarity for image quality assessment. In: Proc. IEEE Asilomar Conference on Signals, System and Computers, pp. 1398–1402 (2003)
11. Zhai, G., Zhang, W.J., Yang, X.K., Xu, Y.: IQA Metrics Based on Multi-scale Edge Presentation. In: IEEE work shop on Signal Processing System Design and Implementation (2005)
12. Malpica, W.S., Bovik, A.C.: Range image quality assessment by Structural Similarity. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1149–1152 (2009)
13. Han, H.S., Kim, D.O., Park, R.H.: Structural information-based image quality assessment using LU factorization. In: Digest of Technical Papers International Conference on Consumer Electronics (2009)
14. Liao, B., Chen, Y.: An Image Quality Assessment Algorithm Based on Dual-scale Edge Structure Similarity. In: Second international Conference on Innovative Computing, Information and Control, p. 56 (2007)
15. Chandler, D.M., Hemami, S.S.: VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images. IEEE Trans. Image Process. 16(19) (September 2007)
16. Lai, Y.K., Li, J., Kuo, J.: A Wavelet Approach to Compressed Image Quality Measurement. Signals, Systems and Computers (2000)
17. Sheikh, H.R., Bovik, A.C.: Image information and visual quality. IEEE Trans. Image Process. 15(2), 430–444 (2006)
18. Loftus, G.R., Harley, E.M.: How Different Spatial-Frequency Components Contribute to Visual Information Acquisition. Journal of Experimental Psychology: Human Perception and Performance 30(1), 104–118 (2004)
19. Loftus, G.R., Nelson, W.W., Kallman, H.J.: Differential acquisition rates for different types of information from pictures. Quarterly Journal of Experimental Psychology: Human Experimental Psychology 35(A), 187–198 (1983)
20. Navon, D.: Forest before trees: The precedence of global features in visual perception. Cognitive Psychology 9, 353–383 (1977)
21. Parker, D.M., Costen, N.P.: One extreme or the other or perhaps the golden mean? Issues of spatial resolution in face perception. Current Psychology 19, 118–127 (1999)
22. Schyns, P.G., Oliva, A.: From blobs to boundary edges: Evidence for time and spatial scale dependent scene recognition. Psychological Science 5, 195–200 (1994)
23. Watt, R.J.: Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. Journal of the Optical Society of America 4, 2006–2021 (1987)
24. Sheikh, H.R., Wang, Z., Bovik, A.C., Cormack, L.K.: Image and Video Quality Assessment Research at LIVE: `http://live.ece.utexas.edu/research/quality/`
25. Eskicioglu, A.M., Fisher, P.S.: Image quality measures and their performance. IEEE Transactions on communications 43, 2959–2965 (1995)
26. Blakemore, C., Campbell, F.W.: On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images. Journal of Physiology 203, 237–260 (1969)
27. Rohaly, A.M., Libert, J., Corriveau, P., Webster, A.: Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment., Eds. (2000)