2012 IEEE International Conference on Systems, Man, and Cybernetics
October 14-17, 2012, COEX, Seoul, Korea

Regularization of The Structural Similarity Index Based on Preservation of Edge Direction

Peng Peng and Ze-Nian Li
School of Computing Science
Simon Fraser University
Burnaby, BC, Canada

***Abstract*—The goal of this paper is to improve the performance of the SSIM indexes while retaining their computational efficiency. To this end, we first design an edge-quality term based on the preservation of edge direction, and then adaptively combine it with the SSIM indexes, yielding the regularized SSIM indexes. The proposed method is based on two assumptions: (1) as the quality of a distorted image declines, the human vision system (HVS) is more likely to judge its quality based on the difficulty of recognizing its content; and (2) the preservation of edge direction is a good measurement of this difficulty.**

Extensive evaluation shows that the regularized SSIM indexes achieve comparable performance to the state-of-the-art method while requiring much less computation time.

***Index Terms*—Image quality assessment, structural similarity (SSIM) index, multi-scale SSIM, edge preservation**

I. Introduction

Image quality assessment plays an important role in acquiring, processing, transmitting and storing multimedia data. A good image quality metric should correlate well with the human perception of image quality. The well-cited Structural SIMilarity (SSIM) index introduced by Wang et al. [1], [2] is a significant step towards this goal. Based on the assumption that the human vision system (HVS) is highly adapted for extracting structural information from a scene, the SSIM index uses a simple model to conduct local comparison of luminance, contrast and structure between a distorted image and a reference image. When compared with the traditional mean square error (MSE) based metrics (e.g. PSNR), it achieves much better correlation with the human perception of image quality.

With the development of modern image quality metrics [3], [4], some other candidates for perceptual image quality assessment stand out. Sheikh and Bovik introduced a Visual Information Fidelity (VIF) method [5] that measures the amount of information about the reference image that can be extracted from the distorted one based on a sophisticated vector Gaussian Scale Mixture (GSM) model. Chandler and Hemami introduced a Visual-Signal-to-Noise (VSNR) [6] method that operates in two stages. The first stage determines whether the distortion in a distorted image is visible via wavelet-based models of visual masking and visual summation. If the distortion is above a detection threshold, a second stage is applied, where the amount of distortion is computed in the distortion-contrast space of multi-scale wavelet decomposition. Another image quality metric with good theoretical ground is the Most Apparent Distortion (MAD) proposed by Larson and Chandler [7]. They advocate that the HVS uses multiple strategies to determine image quality. For images with near-threshold distortion, the HVS tends to look past the image and look for the distortions (a detection-based strategy). For images containing clearly visible distortions, the HVS tends to look past the distortions and look for the image's subject matter (an appearance-based strategy). These assumptions are to some extent in accordance with the recent subjective experiments in [8], where the eye fixation pattern of a viewer is recorded under a task of image quality assessment. The two strategies are then adaptively combined to give the final quality prediction. With its prominent ability in predicting image quality, the MAD suffers from the long computation time and high memory footprint due to its model complexity.

In this paper, we propose an effective and yet computationally efficient model to regularize the SSIM indexes. Based on the fact that the HVS heavily relies on edges and contours to understand a natural scene [14], [15], an edge-quality term based on the preservation of edge direction is used to measure the difficulty for the HVS to recognize the content in a distorted image. After that, we combine this edge-quality term with the SSIM indexes by adopting the adaptive combination scheme of the MAD metric [7], where the contribution of the edge quality to the overall quality prediction is dependent on the image quality. Specifically, the lower the image quality, the greater the contribution from the edge quality. Extensive evaluation on multiple publicly available image quality databases shows that the regularized SSIM indexes stand out with their remarkable performance of quality prediction and low computational complexity, when compared with a set of competitive image quality metrics.

II. Structural SIMilarity Index and Edges

The basic SSIM [1] includes separated comparisons of local luminance, contrast and structure between a reference image and a distorted image:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \tag{1}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \tag{2}$$

$$s(x,y) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \tag{3}$$

where x and y are two local image blocks under comparison, μ_x and μ_y are the means of the intensity values of x and y, σ_x and σ_y are the standard deviations, σ_{xy} is the covariance between x and y, and C_1, C_2 and C_3 are small constants. The general form of the SSIM index between x and y is defined as

$$SSIM(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma}, \quad (4)$$

where α, β, and γ are parameters determining the relative importance of the three components. In [1], $\alpha = \beta = \gamma = 1$ and $C_3 = C_2/2$. This results in a specific form of the SSIM index:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \quad (5)$$

This index is calculated within a local 11×11 window at each pixel, yielding a quality map. In most implementations, the mean value of the quality map is used as the overall image quality:

$$Q_{ssim} = \frac{1}{N} \sum_{n=1}^{N} SSIM(x_n, y_n) \quad (6)$$

where N is the number of image blocks in the reference or distorted image.

Since the perception of image details depends on the scale-related factors, e.g. the distance from the image plane to the viewer, a multi-scale SSIM approach is proposed in [2]. The multiple scales of the reference image and distorted image are obtained by iteratively low-pass filtering and down sampling by a factor of 2. Let $x_{i,j,r}$ and $y_{i,j,r}$ be the local image patches centered at (i,j) at the r-th scale. The original scale is indexed as 1. Let M be the scale obtained after $M-1$ iterations, N_r be the number of image patches at the r-th scale, and $SSIM_r$ be the r-th scale SSIM. For $r = 1, \ldots, M-1$

$$SSIM_{i,j,r} = c(x_{i,j,r}, y_{i,j,r})s(x_{i,j,r}, y_{i,j,r}), \quad (7)$$

and for $r = M$

$$SSIM_{i,j,r} = l(x_{i,j,r}, y_{i,j,r})c(x_{i,j,r}, y_{i,j,r})s(x_{i,j,r}, y_{i,j,r}) \quad (8)$$

The overall multi-scale SSIM quality score is computed as

$$Q_{msssim} = \prod_{r=1}^{M} (\frac{1}{N_r} \sum_{i,j} SSIM_{i,j,r})^{\beta_r} \quad (9)$$

Obviously, the SSIM indexes do not measure the edge quality explicitly, though, as we mentioned before, edges and contours play an important role in image understanding. There is some recent work that attempts to improve the SSIM index by incorporating the edge quality. Chen et al. propose a gradient-based structural similarity (GSSIM) method for image quality assessment, where the contrast and structure comparison are done on the gradient maps of the reference and distorted images [11]. This method misses a lot of quality cues in a color image by counting mainly on the gradient maps. The ESSIM [12] and EDHSSIM [13] methods calculate the structural similarity index based on the Histogram of Gradients (HoG) descriptors. These two methods are computationally very expensive due to the extensive calculation of the HoG descriptors for a vast number of overlapping image blocks. Besides, Li and Bovik propose a content-partitioned SSIM index which assigns twice more weights to the changed-edge pixels when averaging the SSIM quality map [10]. Though improvement of performance is reported, their method needs to be taken with a degree of skepticism, considering that the ratio of edge pixels to the whole image is usually very low and varies a lot between images containing different content. Despite all the shortcomings, these methods cast lights on the potential of enhancing the SSIM indexes by giving more voice to the edges in the image quality assessment.

III. Proposed Method

In this paper, we improve the performance of the SSIM indexes by incorporating edge quality in an adaptive manner. As claimed in [7], the lower the quality of a distorted image, the more likely the HVS to judge its quality based on the difficulty of recognizing its content. To this end, we first develop an edge-quality term aiming to measure this difficulty. Then, we follow the adaptive combination method in [7] to regularize the SSIM indexes using the edge-quality term.

A. Preservation of Edge Direction

Though edge magnitude and direction are both important in visual perception, we believe that the preservation of edge direction is more critical when people try to recognize the subject matter of an image. For example, contrast change in an image often results in drastic change of the edge magnitude (but little effect on the edge direction). In most cases, people can still readily recognize the content of the image. On the contrary, the disarrangement of edge directions often leads to difficulty in understanding an image. To this end, we design a simple quality measure based on the preservation of edge direction. Firstly, both the reference image and the distorted image are convolved with the eight Kirsch edge operators [16], where each operator responds maximally to an edge oriented in a particular direction. Direction with the maximum edge magnitude is chosen as the direction of that pixel. Mathematically, given an arbitrary pixel x and the pixels in its neighborhood $\begin{matrix} a_0 & a_1 & a_2 \\ a_7 & x & a_3 \\ a_6 & a_5 & a_4 \end{matrix}$, the edge direction of x is given by $\arg\max_{i=0,1,\ldots,7} |5(a_i + a_{i+1} + a_{i+2}) - 3(a_{i+3} + \ldots + a_{i+7})|$, where all subscripts are evaluated modulo 8. Then, a pixel-wise comparison of the edge direction is carried out between the reference image and the distorted image, which yields an edge-quality measure with a simple form

$$Q_e = \frac{N_p}{N}, \quad (10)$$

where N_p is the number of edge pixels whose directions are correctly preserved in the distorted image (compared with that in the reference image), and N is the total number of edge pixels in the reference image. Here, we employ a canny edge detector beforehand to locate all the edge pixels, rather than

TABLE I
DESCRIPTION OF IMAGE DATABASES (N_{ri}: NUMBER OF REFERENCE IMAGES; N_{di}: NUMBER OF DISTORTED IMAGES; N_{dt}: NUMBER OF DISTORTION TYPES)

Name	N_{ri}	N_{di}	N_{dt}
LIVE	29	779	5
CSIQ	30	866	6
TID	25	1700	17
TID-7	25	700	7

simply applying a threshold on the gradient magnitude. The canny edge detector is chosen because it tends to find thin and true edges, which is considered to be important to the efficacy of the proposed edge-quality measure.

B. Regularization of SSIM

For regularization of the single-scale SSIM index, we define

$$R\text{-}SSIM = [Q_{ssim}]^{(1-\alpha)} \cdot [Q_e]^{\alpha}, \tag{11}$$

and for the multi-scale SSIM index,

$$R\text{-}MSSSIM = [Q_{msssim}]^{(1-\alpha)} \cdot [Q_e]^{\alpha}. \tag{12}$$

Here, the weight $\alpha \in [0, 1]$ controls the contribution of the edge quality Q_e to the overall quality prediction. It is worth mentioning that all the Q_{ssim}, Q_{msssim}, and Q_e are in the range of $[0, 1]$, and higher values for them indicate better quality. To let Q_e have more say as the image quality gets poorer, we define

$$\alpha = \frac{1}{1 + \beta_1 (Q_{ssim})^{\beta_2}} \tag{13}$$

for Eq. (11); and similarly

$$\alpha = \frac{1}{1 + \beta_1 (Q_{msssim})^{\beta_2}} \tag{14}$$

for Eq. (12). Here, $\beta_1 \geq 0$ and $\beta_2 \geq 0$ are free parameters.

IV. RESULTS

A. Image Quality Databases

There are seven publicly available image quality databases with subjective ratings from human viewers [3]. We choose to evaluate the proposed method on three of them, namely LIVE [17], CSIQ [18], and TID [19], which have larger sizes and more distortion types. A brief description of them is given in TABLE I. To better demonstrate the effectiveness of our method, we also extract a subset from the TID database. This subset is denoted as "TID-7", which contains the seven most common distortion types in practice, namely, additive Gaussian noise, Gaussian Blur, JPEG compression, JPEG2000 compression, JPEG transmission errors, JPEG2000 transmission errors, and contrast change. Some images from the TID database are presented in Fig. 1.

B. Performance Criteria

We adopt four metrics from [9] to compare the performance of different image quality measures. These metrics are:

- Spearman Rank Correlation Coefficient (SRCC):

$$SRCC = 1 - \frac{6 \sum_{i=1}^{N} d_i^2}{N(N^2 - 1)} \tag{15}$$

  where N is the size of the image dataset, d_i is the difference between the i-th image's ranks in objective quality evaluation and subjective quality evaluation.
- Kendall Rank Correlation Coefficient (SRCC):

$$KRCC = \frac{2(N_c - N_d)}{N(N - 1)} \tag{16}$$

  where N_c and N_d are the number of concordant and discordant pairs based on comparison between their objective ranks and subjective ranks.
- Pearson Linear Correlation Coefficient (PLCC) :

$$PLCC = \frac{\sum_i (q_i - \bar{q}) * (o_i - \bar{o})}{\sqrt{\sum_i (q_i - \bar{q})^2 * \sum_i (o_i - \bar{o})^2}} \tag{17}$$

  where q_i is the i-th objective score after nonlinear regression, and o_i is the corresponding subjective score.
- Root mean-squared error (RMSE):

$$RMSE = \sqrt{\frac{1}{N} \sum_i (q_i - o_i)^2}. \tag{18}$$

Among these metrics, SRCC and KRCC are used to evaluate the prediction monotonicity while the other two are used to measure prediction accuracy.

C. Performance of Quality Prediction

We compare the regularized SSIM indexes with the following image quality metrics:

- PSNR: the classic Peak Signal-to-Noise Ratio method,
- VSNR [6] and VIF [5]: two good perceptual image quality metrics as reported in [3],
- MAD [7] : the state-of-the-art image quality metric,
- EDHSSIM: the recently proposed structural similarity index based on edge direction histogram [13],
- SSIM [1] and MSSSIM [2].

All these quality metrics are evaluated on the four databases described in Table I. On each database, the optimal values for the β_1 and β_2 in the Eq. (13) and (14) are selected by test on 20% of the distorted images, and the reported performance is obtained by applying the methods on the entire database. In practice, these two parameters can be selected beforehand on a set of images that is properly constructed based on the practical distortion conditions. A cheap way to do that is to draw images from the publicly available image quality databases.

The performance on the LIVE, CSIQ, TID-7, and TID databases are shown in Table II, III, V, and IV, respectively. We first analyze the results on the three databases (LIVE, CSIQ and TID7) that contain only common distortion types. It can be observed that:

(a) Additive Gaussian noise (c) JPEG compression (b) JPEG2000 Compression

(d) JPEG transmission errors (e) JPEG2000 transmission errors (f) Contrast change

(g) Gaussian blur (h) Non eccentricity pattern noise (i) Local block-wise distortions

Fig. 1. Images of different distortion types from the TID database

1) The regularized SSIM indexes (R-SSIM and R-MSSSIM) greatly improve the performance of quality prediction compared with the original SSIM indexes.
2) In contrast, the EDHSSIM method that also incorporates edge quality is ineffective in most cases. Similarly, the PSNR and VSNR also perform poor constantly.
3) The proposed R-SSIM and R-MSSSIM consistently achieve comparable or even better performance, when compared with the state-of-the-art method MAD and another competitive method VIF.

The above observations strongly support that the proposed method is a very promising way to improve the performance of the SSIM indexes.

On the TID database, the performances of all methods are not as good as that on the other three databases. This may be attributed to the existence of some "exotic" distortion types in the database, e.g. the local block-wise distortions in Fig. 1(h) and (i). These distortions are extremely difficult to handle for a general-purpose image quality metric [19]. Nevertheless, the regularized SSIM indexes still achieve some performance gain over the original SSIM indexes.

TABLE II
PERFORMANCE ON LIVE

Metrics	SRCC	KRCC	PLCC	RMSE
PSNR	0.8756	0.6865	0.8723	13.36
VSNR	0.9271	0.7610	0.9229	10.52
VIF	0.9632	0.8270	0.9598	7.67
MAD	**0.9669**	**0.8421**	**0.9675**	**6.91**
SSIM	0.9479	0.7963	0.9449	8.95
SSIM(MS)	0.9513	0.8044	0.9489	8.62
EDHSSIM	0.9203	0.7583	0.9265	10.28
R-SSIM	**0.9635**	**0.8305**	**0.9622**	**7.44**
R-MSSSIM	**0.9633**	**0.8302**	**0.9619**	**7.47**

D. Comparison of Computation Time

We compare the computation time by running each algorithm over the same "reference image - distorted image"

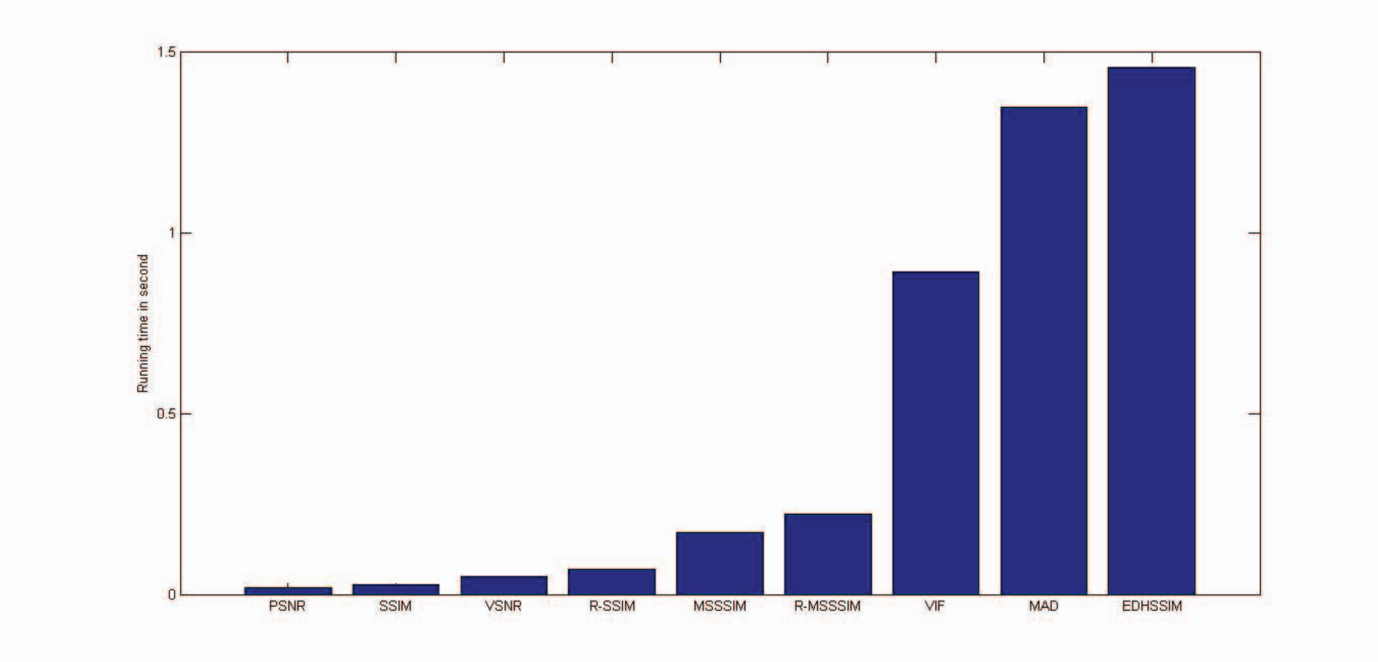

Fig. 2. Running time for the image quality metrics (in second)

TABLE III
PERFORMANCE ON CSIQ

Metrics	SRCC	KRCC	PLCC	RMSE
PSNR	0.8058	0.6084	0.7512	0.1733
VSNR	0.8109	0.6248	0.7355	0.1779
VIF	0.9195	0.7537	0.9277	0.0980
MAD	**0.9466**	**0.7970**	**0.9500**	**0.0820**
SSIM	0.8756	0.6907	0.8613	0.1334
MS-SSIM	0.9133	0.7393	0.8991	0.1149
EDHSSIM	0.7431	0.5633	0.8323	0.1455
R-SSIM	**0.9288**	**0.7655**	**0.9350**	**0.0931**
R-MSSSIM	**0.9453**	**0.7916**	**0.9467**	**0.0846**

TABLE IV
PERFORMANCE ON TID-7

Metrics	SRCC	KRCC	PLCC	RMSE
PSNR	0.6393	0.4612	0.6111	1.2099
VSNR	0.6261	0.4648	0.5940	1.2296
VIF	**0.8861**	**0.7103**	**0.9088**	**0.6378**
MAD	0.8237	0.6563	0.8416	0.8255
SSIM	0.8560	0.6584	0.8337	0.8440
SSIM(MS)	0.8521	0.6568	0.8385	0.8330
EDHSSIM	0.7123	0.5292	0.7520	1.0075
R-SSIM	**0.8796**	**0.6875**	**0.8689**	**0.7565**
R-MSSSIM	**0.8911**	**0.7038**	**0.8840**	**0.7146**

pair for 100 times on a 64-bit Windows machine with Intel Core i7 CPU (@2.20 GHz) and 8GB of RAM [1]. Each image has 512*512 pixels. The average time needed for each run is presented in Fig. 2. We can see that the proposed R-SSIM and R-MSSSIM indexes require much less computation time than the two methods MAD and VIF that demonstrate outstanding capability of quality prediction. Particularly, the R-SSIM index (0.07 second per run) is nearly 20 times faster than the MAD algorithm (1.35 seconds per run). Without doubt, the high efficiency makes our method very appealing in many practical situations, e.g. real-time monitoring of multimedia service quality at the receiver end which could easily be a mobile device with limited computing power.

[1]MATLAB implementation of all the methods under comparison can be found online except that the EDHSSIM, R-SSIM and R-MSSSIM methods were implemented by ourselves (also in MATLAB). The online sources are as follow.
SSIM : https://ece.uwaterloo.ca/~z70wang/research/ssim/ssim.m
MS-SSIM: https://ece.uwaterloo.ca/~z70wang/research/ssim/mssim.zip
PSNR, VSNR and VIF:
http://foulard.ece.cornell.edu/gaubatz/metrix_mux/metrix_mux_1.1.zip
MAD: http://vision.okstate.edu/mad/MAD_index_2011_10_07.zip

TABLE V
PERFORMANCE ON TID

Metrics	SRCC	KRCC	PLCC	RMSE
PSNR	0.5531	0.4027	0.5223	1.1435
VSNR	0.7064	0.5340	0.6820	0.9815
VIF	0.7496	0.5863	0.8090	0.7888
MAD	**0.8340**	**0.6445**	**0.8306**	**0.7473**
SSIM	0.7749	0.5768	0.7732	0.8511
SSIM(MS)	**0.8542**	**0.6568**	**0.8451**	**0.7173**
EDHSSIM	0.5593	0.4003	0.6226	1.0501
R-SSIM	0.7863	0.5950	0.8067	0.7930
R-MSSSIM	**0.8569**	**0.6605**	**0.8543**	**0.6975**

V. CONCLUSION

In this paper, we propose a very simple method to improve the performance of the widely used SSIM indexes. The proposed model explicitly measures the edge quality based on preservation of edge direction, and then adaptively combines it with the original SSIM indexes to make a final quality prediction. Extensive evaluation on multiple publicly available image quality databases shows that our methods achieve remarkable correlation with the human perception of image quality. Moreover, the proposed method is computationally much more efficient than other competitive methods, which makes it a better option for quality assessment tasks that desire both effectiveness and efficiency.

REFERENCES

[1] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[2] Z. Wang, E.P. Simoncelli, and A.C. Bovik, "Multiscale structural similarity for image quality assessment," *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers*, vol. 2, no. 1, pp. 1398–1402, 2003.

[3] W. Lin and C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey," *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, May 2011.

[4] A. K. Moorthy and A. C. Bovik, "Visual quality assessment algorithms: what does the future hold ," *Multimedia Tools Appl.*, vol. 51, no. 2, pp. 675–696, 2011.

[5] H. R. Sheikh and A. C. Bovik, "Image information and visual quality.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 15, no. 2, pp. 430–44, Feb. 2006.

[6] D. M. Chandler and S. S. Hemami, "VSNR: a wavelet-based visual signal-to-noise ratio for natural images.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 16, no. 9, pp. 2284–98, Sept. 2007.

[7] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 011006, 2010.

[8] J. Redi, H. Liu, R. Zunino, and I. Heynderickx, "Interactions of visual attention and quality perception," *Imaging*, vol. 7865, pp. 78650S–78650S–11, 2011.

[9] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 20, no. 5, pp. 1185–98, May 2011.

[10] C. Li and A. C. Bovik, "Content-partitioned structural similarity index for image quality assessment," *Signal Processing: Image Communication*, vol. 25, no. 7, pp. 517–526, Aug. 2010.

[11] G. H. Chen, C. L. Yang, and S. L. Xie, "Gradient-based structural similarity for image quality assessment," *ICIP*, Atlanta, GA, 2006.

[12] G. H. Chen, C. L. Yang, L. M. Po, and S. L. Xie, "Edge-based structural similarity for image quality assessment," *ICASSP*, Hubei, China, 2006.

[13] X. Chen, R. Zhang, and S. Zheng, "Image quality assessment based on local edge direction histogram," *International Conference on Image Analysis and Signal Processing*, 2011.

[14] J. H. Elder and . W. Zucker, "Evidence for boundary-specific grouping," *Vision Research*, vol. 38, no. 1, pp. 143–152, 1998.

[15] D. Marr, E. Hildreth, "Theory of edge detection," *Proceedings of the Royal Society of London*, Series B 207, pp.187–217, 1980.

[16] R. Kirsch, "Computer determination of the constituent structure of biological images," *Computers and Biomedical Research*, vol. 4, pp. 315–328, 1971.

[17] A. K. Moorthy Z. Wang A. C. Bovik H. R. Sheikh, K. Seshadrinathan and L. K. Cormack, "Image and video quality assessment research at LIVE," http://live.ece.utexas.edu/research/quality/.

[18] E. C. Larson and D. M. Chandler, "Categorical image quality (CSIQ) database," http://vision.okstate.edu/csiq.

[19] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, J. Astola, M. Carli, and F. Battisti, "TID2008: A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics," *Advances of Modern Radioelectronics*, vol. 10, pp. 30-45, 2009.