

Three-Component Weighted Structural Similarity Index

Chaofeng Li^{1,2} and Alan C. Bovik,² *Fellow, SPIE*

(¹School of Information Technology, Jiangnan University, Wuxi, 214122, China)

(²Laboratory for Image and Video Engineering (LIVE), The University of Texas at Austin, Austin, TX 78712, USA)

The assessment of image quality is very important for numerous image processing applications, where the goal of image quality assessment (IQA) algorithms is to automatically assess the quality of images in a manner that is consistent with human visual judgment. Two prominent examples, the Structural Similarity Image Metric (SSIM) and Multi-scale Structural Similarity (MS-SSIM) operate under the assumption that human visual perception is highly adapted for extracting structural information from a scene. Results in large human studies have shown that these quality indices perform very well relative to other methods. However, the performance of SSIM and other IQA algorithms are less effective when used to rate amongst blurred and noisy images. We address this defect by considering a three-component image model, leading to the development of modified versions of SSIM and MS-SSIM, which we call *three component SSIM* (3-SSIM) and *three component MS-SSIM* (3-MS-SSIM).

A three-component image model was proposed by Ran and Farvardin,^[13] wherein an image was decomposed into edges, textures and smooth regions. Different image regions have different importance for vision perception, thus, we apply different weights to the SSIM scores according to the region where it is calculated. Thus, four steps are executed: (1) Calculate the SSIM (or MS-SSIM) map. (2) Segment the original (reference) image into three categories of regions (edges, textures and smooth regions). Edge regions are found where a gradient magnitude estimate is large, while smooth regions are determined where the gradient magnitude estimate is small. Textured regions are taken to fall between these two thresholds. (3) Apply non-uniform weights to the SSIM (or MS-SSIM) values over the three regions. The weight for edge regions was fixed at 0.5, for textured regions it was fixed at 0.25, and at 0.25 for smooth regions. (4) Pool the weighted SSIM (or MS-SSIM) values, typically by taking their weighted average, thus defining a single quality index for the image (3-SSIM or 3-MS-SSIM).

Our experimental results show that 3-SSIM (or 3-MS-SSIM) provide results consistent with human subjectivity when finding the quality of blurred and noisy images, and also deliver better performance than SSIM (and MS-SSIM) on five types of distorted images from the LIVE Image Quality Assessment Database.

Keywords: Three-Component, Human Visual System (HVS), Image Quality Assessment, Structural Similarity (SSIM), Multi-scale Structural Similarity (MS-SSIM)

1 Introduction

The role of images in present day communication has been steadily increasing, but digital images are subject to a wide variety of distortion during acquisition, processing, compression, storage, transmission and reproduction, so the assessment of image quality is very important for numerous image processing applications.

Over the years, numerous objective metrics have been proposed^[1-6] to assess image quality. The simplest and most widely used quality metrics are the mean square error (MSE), computed by averaging the squared intensity differences of

distorted and reference image pixels, and the related quantity of peak signal-to-noise ratio (PSNR). However, the MSE and its variants do not correlate well with subjective quality measures^{[7]-[10]}. There are a number of perceptual factors^[11] that can influence human perception of visual quality. Successfully incorporating these into objective IQA metrics can lead to an improved correlation with visual perception^[3].

Two prominent examples, the Structural Similarity Image index (SSIM)^[2,12] and the Multi-scale Structural Similarity index (MS-SSIM)^[1] operate under the assumption that visual perception is highly adapted for extracting structural information from a scene. Results in large human studies have shown that these quality indices perform very well relative to other methods. Gao et al.^[13] develop a content-based image quality assessment metric, and use complex fuzzy integral method to fuse SSIM of image three regions. However, the performance of SSIM is not as competitive when used to assess images between blurred and noise image. We address this defect by considering a three-component image model^[14], and use simple weighted mean SSIM of three regions, and lead to the development of modified versions of SSIM and MS-SSIM, which we call *three component SSIM* (3-SSIM) and *three component MS-SSIM* (3-MS-SSIM).

This paper is organized as follows. Section 2 reviews SSIM, MS-SSIM and a gradient-based SSIM. The proposed 3-SSIM (and 3-MS-SSIM) indices are described in Section 3. Section 4 gives experimental results. Finally, in Section 5, future thoughts are presented.

2 Single-scale SSIM and Multi-scale SSIM Indices

2.1 Single-scale SSIM

Based on the assumption that the HVS is highly adapted to extract structural information from the viewing field, a new philosophy for image quality measurement SSIM was proposed by Wang and Bovik^[12]. It defines the function for the luminance comparison of the signals, the contrast comparison of the signals, and the structure comparison of the signals respectively as follows.

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

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

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

where μ_x and μ_y are the sample means of x and y respectively, σ_x^2 and σ_y^2 are the sample variances of x and y respectively, and σ_{xy} is the sample correlation coefficient between x and y . The small constants C_1 , C_2 , C_3 are used to avoid instability when the denominator(s) might approach zero.

Combining the three comparison functions of (1)-(3) yields a specific form of the SSIM index:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (4)$$

In [12], the SSIM index is used for IQA via a 11*11 sliding window approach. The window moves pixel-by-pixel across the whole image space. At each step, the SSIM index is calculated within the local window, and the resulting SSIM index map can be viewed as the quality map of the distorted images. Finally, a mean SSIM index of the quality map is used to evaluate the overall image quality.

2.2 Multi-scale SSIM

Wang, Simoncelli and Bovik developed a multi-scale SSIM (MS-SSIM) index ^[1]. Denote the original image as Scale 1, and the highest scale as Scale M, which is obtained after M-1 iterations. At the j^{th} scale, the contrast comparison (2) and the structure comparison (3) are calculated and denoted as $c_j(x,y)$ and $s_j(x,y)$, respectively. The luminance comparison (1) is computed only at Scale M and denoted as $l_M(x,y)$. The overall MS-SSIM index is obtained by combining the measurement across scales via

$$SSIM(x, y) = l_M(x, y) \cdot \prod_{j=1}^M c_j(x, y) s_j(x, y) \quad (5)$$

MS-SSIM is able to capture distortions as they occur across scales, as well as better matching the human visual response. In human studies, MS-SSIM significantly outperforms single-scale SSIM.

2.3 Modified gradient-based SSIM (GSSIM)

Chen, Yang and Xie^[15] developed an improved SSIM algorithm called Gradient-based Structural Similarity (GSSIM), which compares edge information between the distorted image and the original image. GSSIM replaces the contrast comparison $c(x, y)$ in (2) and structure comparison $s(x,y)$ in (3) with gradient-based contrast comparison $c_g(x, y)$ and structure comparison $s_g(x, y)$ respectively.

Let X' and Y' denote the gradient maps of the original and the distorted images respectively, and let x' and y' be block vectors from X' and Y' . Then:

$$c_g(x, y) = \frac{2\sigma_{x'}\sigma_{y'} + C_2}{\sigma_{x'}^2 + \sigma_{y'}^2 + C_2} \quad (6)$$

$$s_g(x, y) = \frac{\sigma_{x'y'} + C_3}{\sigma_{x'}\sigma_{y'} + C_3} \quad (7)$$

where the sample statistics have the same definition as in (1)-(3), but applied on the gradient images. Subsequently GSSIM is given by:

$$GSSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{x'y'} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (8)$$

3 Modifying SSIM: 3-SSIM and 3-MS-SSIM

IQA algorithms generally operate without attempting to take into account image content. Since algorithms for image content identification remain in a nascent state, IQA algorithms that succeed in assessing quality as a function of content will await developments in that direction. However, low-level content of visual importance, sometime called salient image features, might be used to improve IQA algorithms. For example, intensity edges certainly contain considerable image information, and are perceptually significant.^[16] Using this observation we incorporate a three-component image model into SSIM (or MS-SSIM), and thereby develop three-component weighted SSIM (3-SSIM) and MS-SSIM (3-MS-SSIM) indices. We have also developed three-component indices using GSSIM (and MS-GSSIM), which we term the 3-GSSIM (and 3-MS-GSSIM) indices.

The development of 3-SSIM and 3-MS-SSIM follows four steps: (1) Calculate the SSIM (or MS-SSIM) map. (2) Independent of the SSIM results, segment the original (reference and distorted) image into three categories of regions (edges, textures and smooth regions). Edge regions are found where a gradient magnitude estimate is large, while smooth regions are determined where the gradient magnitude estimate is small. Textured regions are taken to fall between these two thresholds. (3) Apply non-uniform weights to the SSIM (or MS-SSIM) values over the three regions. (4) Pool the weighted SSIM (or MS-SSIM) values, e.g., their weighted average, thus defining a single quality index for the image (3-SSIM or 3-MS-SSIM). A diagram depicting calculation of 3-SSIM (or 3-MS-SSIM) is shown in Fig. 1.

3.1 Three-Component Image Model

In our approach, an image is partitioned into three parts: edges, textures and smooth regions as proposed in [14]. In our approach, we seek to more heavily weight degradations of intensity edges in the IQA process. Textured regions, by contrast, often mask degradations. Artifacts in smooth regions may be quite obvious, especially if they are high-frequency or edge-like.

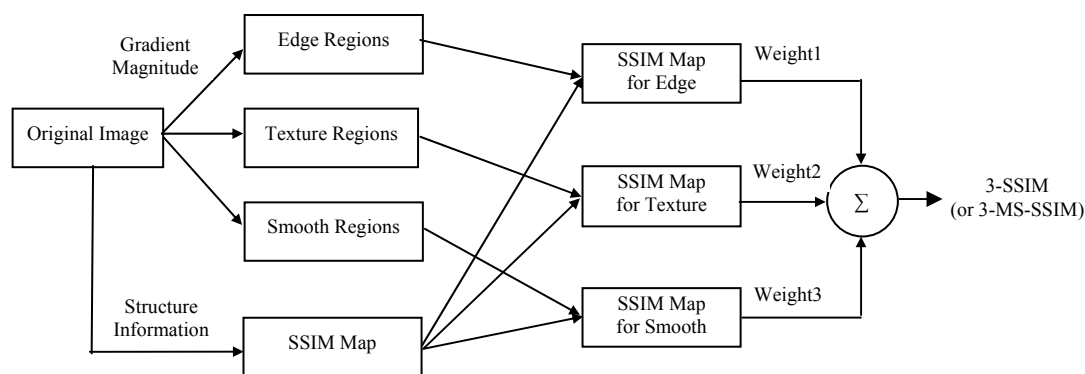


Fig.1 Diagram for calculating 3-SSIM (or 3-MS-SSIM)

We may partition an image into three components using the computed gradient magnitude; [15] gives a simple image partition method that we use to obtain a partition into three types of regions. The following steps explain the process:

Step 1: Compute the gradient magnitudes using a Sobel operator on the original and the distorted images.

Step 2: Determine thresholds $TH_1 = 0.12 * g_{max}$ and $TH_2 = 0.06 * g_{max}$, where g_{max} is the maximum gradient magnitude value computed over the original image.

Step 3: Assign pixels as belonging to edge, texture, and smooth regions, as follows.

Denoting the gradient at coordinate (i, j) on the original image by $p_o(i, j)$ and the gradient on the distorted image as $p_d(i, j)$, the pixel classification is carried out according to the following rules:

R₁: If $p_o(i, j) > TH_1$ or $p_d(i, j) > TH_1$, then the pixel is considered as an edge pixel.

R₂: If $p_o(i, j) < TH_2$ and $p_d(i, j) \leq TH_1$, then the pixel is regarded as being part of a smooth region.

R₃: Otherwise, the pixel is regarded as being in a textured region but is not an edge pixel.

Figure 2 shows an original image and a blurred version of it. Also shown are the identified edges, the identified smooth regions, and the identified textured regions.

3.2 Determining the Weights

Edges play an important role in the perception of images, and edges that are distorted, e.g., by blur, can greatly impact the perceived quality of an image. We therefore modify the SSIM and MS-SSIM indices by allocating greater weight to the scores on edge regions than on smooth and texture regions. To keep things simple, we fixed the weight for edge regions at 0.5.

Smooth regions are important also, and the eye is sensitive to such artifacts as false contouring, blocking, and high-frequency noise in smooth regions. While the distortion of textures can also be perceptually significant, some distortions can be obscured or masked by the presence of textures. As a simple approach, we apply the same weights on smooth and texture regions, fixing both at 0.25 (hence the sum of all weights is unity).

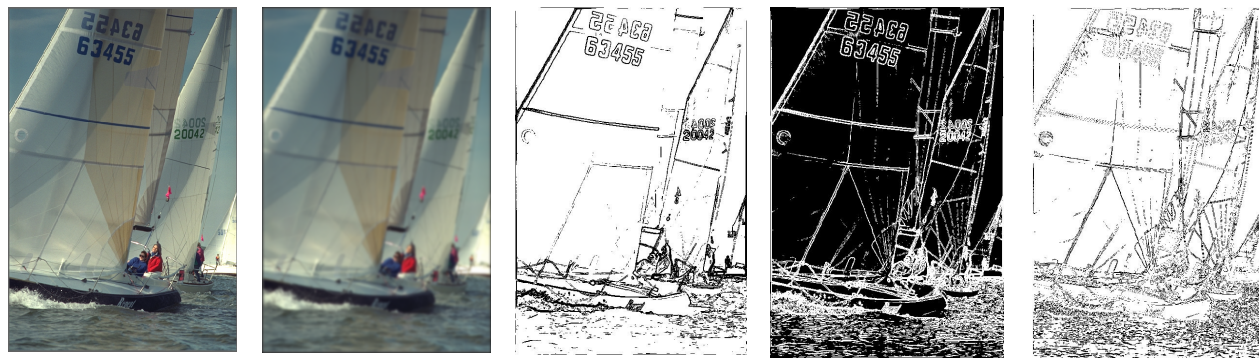


Fig. 2 From left-to-right: Original image; blurred image; edge pixel image; smooth pixel image; and texture pixel image. In the latter three cases black indicates membership in the classes edge, smooth, and texture, respectively.

5 Experimental Results and Discussion

5.1 Specific comparison for blurred image and noise image assessment

Extensive experimental results have shown that the SSIM and MS-SSIM indices correlate with the perception of visual quality much better than does the PSNR, and other earlier IQA algorithms.^[12] However, these studies also suggest that performance might be improved when assessing the quality of blurred and noisy images, as also shown in Fig. 3. Most observers would agree that Fig. 3(a) is of better quality than Fig. 3(b), yet SSIM and MS-SSIM give contrary assessment results. By comparison, 3-SSIM and 3-MS-SSIM yield assessments that appear to better agree with visual perception of quality.



Fig. 3 Comparison of IQA algorithms on blurred and noisy images (a) noise image: subjective mean opinion score (MOS) = 44.7534, SSIM = 0.3491, MS-SSIM = 0.8539, 3-SSIM = 0.6120, 3-MS-SSIM = 0.9391 (b) blurred image: MOS = 66.3322, SSIM = 0.6932, MS-SSIM = 0.8825, 3-SSIM = 0.4933, 3-MS-SSIM = 0.8364.

5.2 Experimental results on LIVE image database

In order to further provide more extensive quantitative comparisons, we evaluated the performances of 3-SSIM, 3-MS-SSIM, 3-GSSIM, 3-MS-GSSIM, SSIM, GSSIM, MS-SSIM, MS-GSSIM and PSNR on the LIVE Image Quality Assessment Database^[17]. This database includes five types of distorted images: JPEG, JPEG2000, white noise (WN), Gaussian Blur (Gblur) and Fastfading (FF). Difference Mean Opinion Scores (DMOS) are given for each distorted image.

Two performance metrics are used to evaluate the algorithms. The first is the linear correlation coefficient (LCC) between DMOS and the algorithm scores following nonlinear regression. The nonlinearity chosen for regression was a five-parameter logistic function^[18]. The second is the Spearman rank order correlation coefficient (SROCC). The results, presented in Tables I and II, are reported for the different distortion types as well as on the entire dataset. As can be seen, 3-SSIM outperforms SSIM and PSNR on all types of distorted images, and also performs better than MS-SSIM for WN, Gblur and FF distortions. Overall, 3-MS-SSIM outperforms MS-SSIM. GSSIM also outperforms SSIM, and 3-GSSIM far outperforms GSSIM. But GSSIM is not as effective in multiscale form, and MS-GSSIM is far inferior to MS-SSIM. Our three-component weighted method is still effective for MS-GSSIM, and 3-MS-GSSIM outperforms MS-GSSIM. Figures 4 and 5 depict scatter plots of DMOS versus 3-SSIM and 3-MS-SSIM for the different kinds of distortions.

5 Conclusions

In this paper, we presented a three-component weighted structural similarity metric. Using the idea that different image regions have different perceptual significance relative to quality, we used different weights for SSIM scores according to segmentation of images into regions. Our experimental results show that 3-SSIM (or 3-MS-SSIM) provide results consistent with human subjectivity when comparing the quality of blurred and noisy images, and also deliver better performance than SSIM (and MS-SSIM) on five types of distorted images from the LIVE Image Quality Assessment Database. Interesting, 3-SSIM performed better than either MS-SSIM or 3-MS-SSIM, suggesting that edge relevance might be of greater significance than multiscale.

Table I Linear correlation coefficient after nonlinear regression

	JPEG2000	JPEG	WN	Gblur	FF	All data
PSNR	0.8878	0.8596	0.9813	0.7840	0.8752	0.9293
SSIM	0.9368	0.9297	0.9793	0.8741	0.9452	0.9388
3-SSIM	0.9592	0.9323	0.9871	0.9686	0.9688	0.9479
GSSIM	0.9382	0.9343	0.9537	0.9076	0.9479	0.9563
3-GSSIM	0.9581	0.9414	0.9798	0.9587	0.9553	0.9701
MS-SSIM	0.9578	0.9426	0.9860	0.9579	0.9346	0.9435
3-MS-SSIM	0.9588	0.9417	0.9903	0.9735	0.9423	0.9440
MS-GSSIM	0.9378	0.9137	0.9669	0.9731	0.9445	0.9492
3-MS-GSSIM	0.9513	0.9320	0.9833	0.9763	0.9496	0.9557

Table II Spearman rank order correlation coefficient

	JPEG2000	JPEG	WN	Gblur	FF	All data
PSNR	0.8898	0.8409	0.9853	0.7816	0.8903	0.9092
SSIM	0.9317	0.9028	0.9629	0.8942	0.9411	0.9250
3-SSIM	0.9536	0.9045	0.9801	0.9687	0.9624	0.9447
GSSIM	0.9326	0.9038	0.9367	0.9364	0.9451	0.9448
3-GSSIM	0.9507	0.9069	0.9651	0.9676	0.9607	0.9617
MS_SSIM	0.9536	0.9108	0.9780	0.9539	0.9350	0.9532
3-MS_SSIM	0.9538	0.9114	0.9881	0.9698	0.9440	0.9602
MS-GSSIM	0.9311	0.8804	0.9447	0.9692	0.9381	0.9514
3-MS-GSSIM	0.9432	0.8977	0.9728	0.9713	0.9466	0.9629

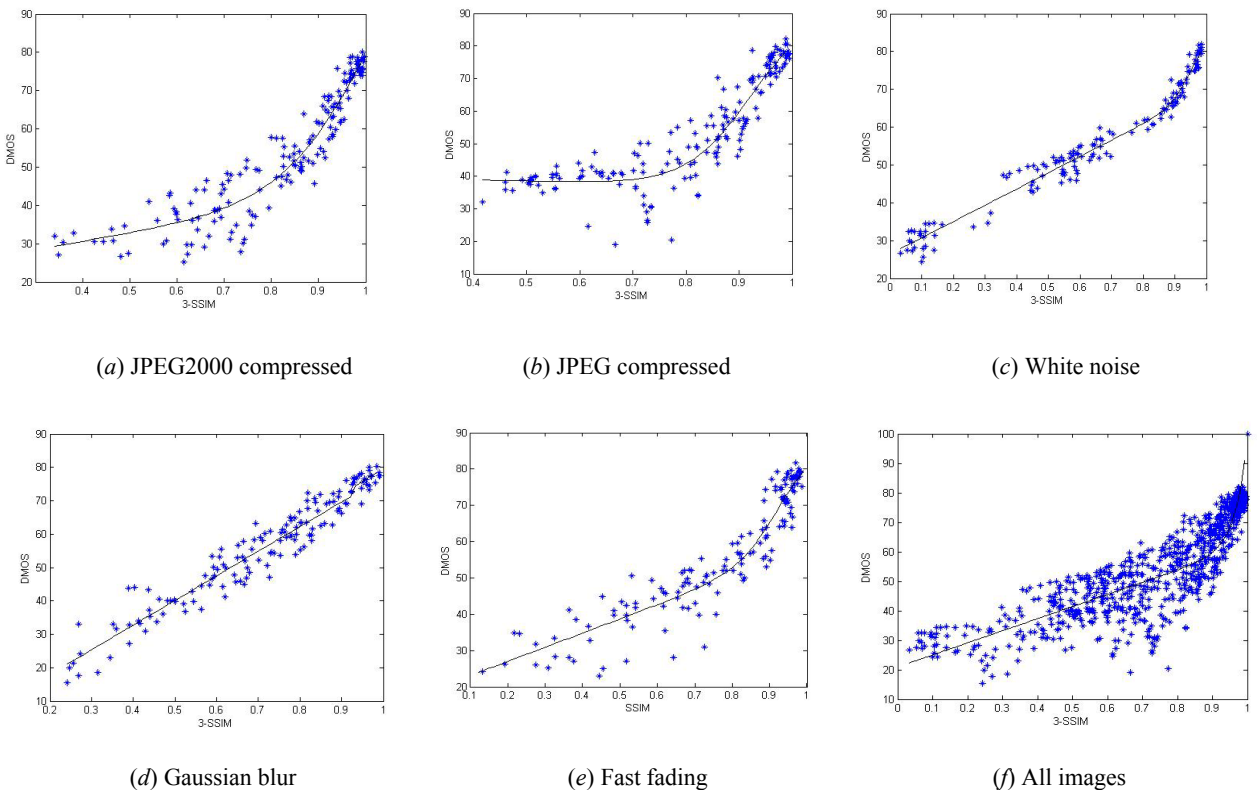


Fig. 4 Scatter plots for DMOS versus 3-SSIM for different kinds of distortions.

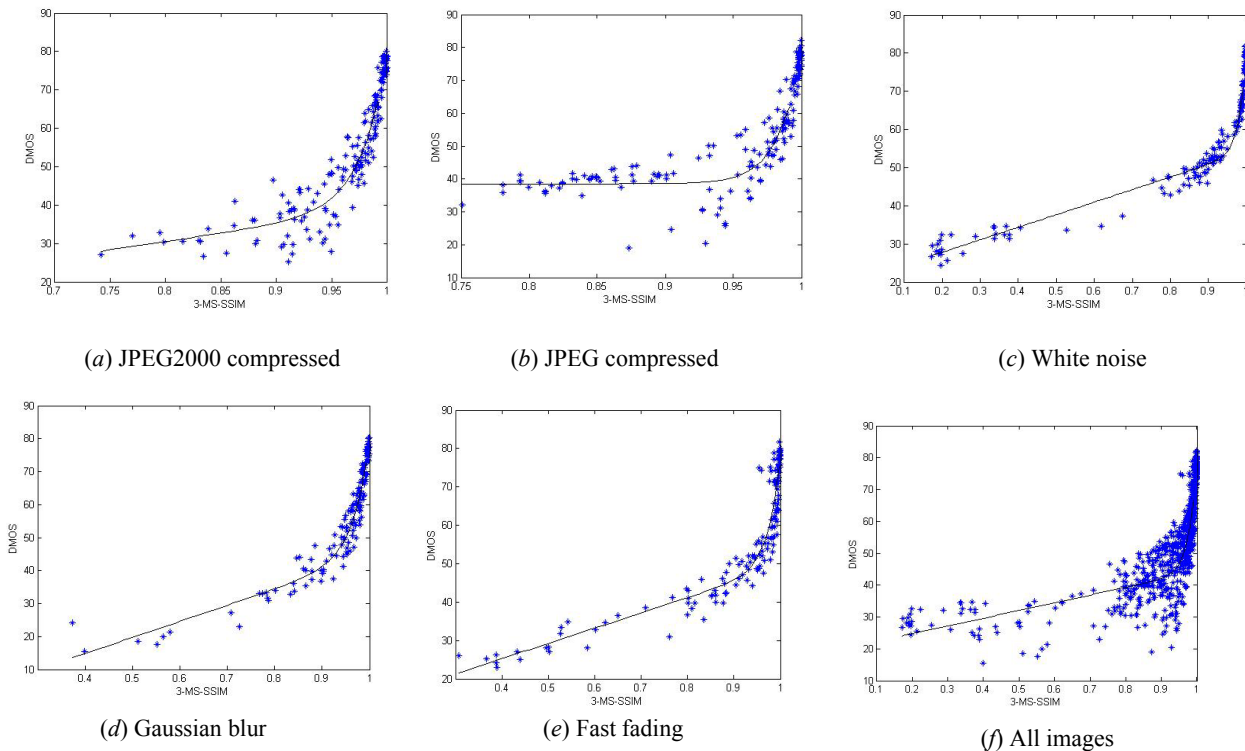


Fig.5 Scatter plots for DMOS versus 3-MS-SSIM for different kinds of distortions.

Acknowledgments

The Author would like to thank Dr. Kalpana and Anush for their help in programming.

References

- [1] Wang Z., Simoncelli E.P. and Bovik A.C., "Multi-scale structural similarity for image quality assessment," Proc. IEEE *Asilomar Conference on Signals, Systems, and Computers*, 1398-1402(2003).
- [2] Wang Z. and Bovik A. C., "A universal image quality index," *IEEE Signal Processing Letters*, 9, 81–84 (2002).
- [3] Eskicioglu A.M., Fisher P.S., "Image quality measures and their performance," *IEEE Transactions on Communications* , 43(12), 2959–2965 (1995).
- [4] Karunasekera, S.A., Kingsbury, N.G., "A distortion measure for blocking artifacts in images based on human visual sensitivity," *IEEE Transactions on Image Processing*, 4(6), 713–724 (1995).
- [5] Mill, N.B., "A visual model weighted cosine transform for image compression and quality assessment," *IEEE Transactions on Communications*, 33(6), 551–557 (1985).
- [6] Saghri, J.A., "Image quality measure based on a human visual system model," *Optical Engineering*, 28(7), 813–818 (1989).
- [7] Girod B., "What's wrong with mean-squared error," in *Digital Images and Human Vision*, A.B. Watson (Ed.) Cambridge, MA: MIT Press, 207–220 (1993).
- [8] Teo, P.C. and Heeger, D.J., "Perceptual image distortion," *Proc. SPIE* 2179, 127–141(1994).
- [9] Eckert M. P. and Bradley A. P., "Perceptual quality metrics applied to still image compression," *Signal Processing*, 70, 177–200 (1998).
- [10] Wang Z., Bovik, A.C., and Lu, L., "Why is image quality assessment so difficult," *IEEE Int. Conf. Acoustics, Speech and Signal Processing* 4, 3313–3316 (2002).
- [11] Wandell B.A., [*Foundations of Vision*], Sinauer Associates Inc, 1995.
- [12] Wang Z., Bovik, A.C., Sheikh, H.R. and Simoncelli E.P., "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, 13(4), 600-612 (2004).
- [13] Gao X.B., Wang T., and Li J., "An Objective Content-based Image Quality Assessment Metric," Proc LNAI 3642, 231–240(2005).
- [14] Ran X. and Farvardin N., "A perceptually-motivated three-component image model - Part I: description of the model," *IEEE Transactions on Image Processing*, 4(4), 401-415(1995).
- [15] Chen G.H., Yang C.L., Xie S.L., "Gradient-based structural similarity for image quality assessment," Proc. ICIP, 2929-2932(2006).
- [16] Li J.L., Chen G., Chi, Z.R., *et al.* "Image coding quality assessment using fuzzy integrals with a three-component image model," *IEEE Transactions on Fuzzy Systems*, 12(1), 99-106(2004).
- [17] Sheikh, H.R., Wang, Z., Cormack, L. and Bovik, A.C., "LIVE Image Quality Assessment Database," [online] <http://live.ece.utexas.edu/research/quality>.
- [18] Sheikh H.R., Sabir M.F., and Bovik A.C., "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, 15(11), 3441-3451(2006).