

Performance Evaluation of Distortion Measures for Retinal Images

Nirmala S.R, Dandapat S. and Bora P.K
Department of Electronics and Communication Engineering
Indian Institute of Technology Guwahati
Guwahati-781039, Assam, India

ABSTRACT

Evaluating the quality of processed retinal images is an important issue in applications such as telemedicine. The traditional image quality measures are having limitation in emphasizing the loss of clinically significant information. We previously proposed a wavelet weighted blood vessel distortion measure (WBVDM) for retinal images. The WBVDM gives more importance to the distortion in clinical features (blood vessels) and less importance to the clinically nonsignificant distortion. This paper presents a statistical evaluation of the performance of a number of image quality measures in quantifying the distortion in retinal images. The measures are then investigated in terms of their correlation with subjective evaluation using the Pearson linear correlation coefficient (PLCC) and Spearman rank order correlation coefficient (SROCC). Their statistical behavior is also evaluated in terms of how discriminating they are to distortion artifacts when tested on a variety of images using the analysis of variance (ANOVA) method. The experimental results indicate that WBVDM performs better by showing higher values of PLCC, SROCC and ANOVA analysis.

General Terms

Medical image analysis

Keywords

Image quality, DWT, subband coefficients, objective measures, WBVDM, MSSIM.

1. INTRODUCTION

The advanced digital imaging and computing technology is used with full potential in modern ophthalmology as it is dependent on digital retinal images [1]. The diagnostic features of retinal image are the blood vessels, optic disc and macula. The digital retinal images are essential tools in identifying and diagnosing eye disorders such as diabetes, glaucoma or age-related macular degeneration. Processing of retinal images for storage, compression, transmission and reproduction leads to distortions in the retinal features. This may result in the degradation of image quality. The clinical usefulness of a medical image is highly dependent on its quality. Hence it is important to quantify the degradation occurring in the processed retinal images and to devise suitable measurement methods. Image quality evaluation can be done either subjectively or objectively. Subjective evaluation is the most accurate and reliable way of assessing the quality of an image. However, this method is slow, inconvenient and expensive for practical usage. Thus, objective image quality metrics that can automatically predict the perceived image quality are preferred [2],[3]. The objective quality measures are simple to compute and useful in many ways. The commonly used mean square error (MSE),

root mean squared error (RMSE), normalized mean squared error (NMSE) and the peak signal-to-noise ratio (PSNR) are computed by averaging the squared intensity differences of distorted and reference image pixels. They perform a global measurement of the quality of the processed image and do not provide information about the local distortion. A lower value of the objective measure does not ensure the clinically agreeable quality. Hence they are not well matched with the medical expert evaluation results.

For medical images, a quality/distortion measure has to be defined from the diagnostic perspective which takes proper account of the medical nature of the images [4]. Certain regions of a medical image are rich in clinical information and some regions are clinically nonsignificant. In retinal images the features like blood vessels, optic disc and macula constitutes the diagnostically important information. The retinal background is generally characterized as containing no useful information. Hence the quality measure for retinal image is expected to emphasize any distortion in the diagnostic features and ignore the effects in nondiagnostic regions.

This paper is organized as follows. Section 2 presents an overview of the techniques to measure retinal image quality from a general perspective and discusses some of the wavelet based image quality measures. Section 3 gives the details of the subjective experiment and the different objective measures used to evaluate the quality of retinal images. The performance evaluation results of the study are discussed in Section 4. The conclusion of the analysis is made in Section 5.

2. IMAGE QUALITY ASSESSMENT

This section presents an overview of the quality assessment metrics for fundus retinal image. The limitations of the conventional quality assessment metrics and need for a more meaningful distortion measure for the retina image is also discussed.

2.1 Retinal image quality assessment

The quality assessment (QA) of retinal images has more constraints than the traditional image quality evaluation methods presented in the previous section. The QA has to be performed to simulate the opinion of expert ophthalmologists. The research on retinal image QA started with histogram based methods [5]. The approach starts with an intensity histogram template generated from excellent quality retinal images. The target image quality is assessed by convoluting its histogram with the template histogram. A quality index Q , normalized between 0 and 1 is obtained and used to define the quality of the given image. But the subsequent literature [6]

found that some good quality images having histograms different from the template histogram whereas the histogram of some poor quality images resembled the template histogram. Hence the authors tried to modify the QA by using two different sets of features: the distribution of the edge magnitudes in the image (global edge histogram) and the local distribution of the pixel intensity (local histogram), as compared to the global histogram of [5]. The QA in [7] considers features unique to retinal images. The authors found a correlation between image blurring and visibility of the vessels. The blood vessel segmentation procedure is applied to the retinal image and the area of the detected vessels is computed. The image quality is estimated using the total vessel area. But the limitation of the method is: even if some vessels are not detected and the main thick vessels with significant area may mislead the retinal image QA algorithm. In the subsequent publication [8], image quality was defined by two aspects: image clarity and field definition. The visibility of the macular vessels was used as an indicator of image clarity, since these vessels are known to be narrow and become less visible with any image degradation. An image with adequate field definition was defined as one that shows the full 45° field of view, the optic disc, and at least two optic disc diameters of visible retina around the fovea. This type of QA methods have been used to assess the quality of retinal images for use in diabetic retinopathy screening. The most recent method for QA [9] employ two sets of feature to represent image quality: colour and second order image structure invariants (ISI). The authors employed filter banks to generate features invariant to rotation, position or scale. The authors call the whole QA process as image structure clustering (ISC).

2.2 Image compression considerations

In applications like telemedicine, large image databases and bandwidth limitations make lossy image compression techniques a necessity. Hence the evaluation of medical image degradation due to compression is an important stage in telemedicine. A list of literature is available for retinal image compression for various applications [10]-[15]. In [10], the authors have analyzed the effects of various degrees of JPEG compression on the automatic and manual diagnosis of diabetic retinopathy based on the number of detected microaneurysms. They found that the distortion introduced by the JPEG compression affect the diagnosis only at relatively high compression ratios. The RMSE, a pixel based measure and a variant of MSE is used in [12] to quantify the distortion in retinal image. But the pixel based measures estimate average global error and do not show the localized errors. They fail to reflect the degradation in diagnostic information. Hence it is required to develop and evaluate a diagnostically meaningful objective measure for digital retinal images.

2.3 Discrete wavelet transform (DWT) based image quality measure

The discrete wavelet transform (DWT) provides an efficient spatial and frequency localization of an image. Wavelet based image quality assessment schemes have been proposed based

on the idea to exploit the characteristics of the human visual system (HVS) [16]-[18]. These works are aimed at evaluating the visual quality of a multimedia image. The DWT is used in various biomedical image processing applications such as noise removal, enhancement and detection of diagnostic features [19]. The multiresolution analysis (MRA) or multiscale property of DWT helps to extract local image features effectively. The MRA technique offers, features that are difficult to detect at one scale but may be easily detected at another scale [20]. The recent compression techniques for medical images such as JPEG2000 and DICOM (Digital Imaging and Communication in Medicine) [21], [22] use the discrete wavelet transform (DWT). Hence a wavelet based quality measure will be useful for these compression techniques.

The retinal images contain fine anatomical features such as blood vessel structure and coarse features or regions with slowly varying pixel values such as the optic disc and homogeneous retinal background. The anatomical structure of the retina is dominated by the blood vessel structure and they have a high diagnostic impact. In a retinal image, different blood vessels have different resolutions; main thick vessel branches into thin vessels and spreads all over the retina. The most important criteria for determining the retinal image quality is whether the blood vessels can be easily distinguished from the background in the processed (reconstructed) images. A good quality retinal image is the one in which main vessels (emerging from the optic disc) and the fine blood vessels (near the macula) can be observed [8]. Since the blood vessels in a retinal image are having different resolution (varying thickness), multiresolution property of wavelets can be used to analyze the retinal vessel features. The directional property is another important characteristic of the vessel structure. Since two dimensional DWT (2D DWT) can decompose the image in several directions, the vessel details in different orientations can also be studied. However in the course of processing, changes may arbitrarily distort the blood vessels such as, smoothing of blood vessels in case of lossy image compression [23]. Any changes in the retinal blood vessels need to be quantified for diagnostic analysis of retinopathy diseases. Hence a distortion measure which captures the changes in the blood vessel structure effectively and gives less importance to the distortion in clinically nonsignificant background regions, is more meaningful. In this direction, we have developed a wavelet based distortion measure for blood vessels in our previous work [24] and it is briefly described in the following paragraph.

The N-level DWT of an image consists of one low frequency (approximation) subband and (3x N) high frequency (detail) subbands. The wavelet analysis of retinal image shows that the retinal image information is spread over the different wavelet subbands. Hence it is expected that certain subbands may contain relatively more information about the retinal blood vessel features than the other subbands. The diagnostic importance of each subband is examined by zeroing the coefficients of the subband and keeping all other subband coefficients unchanged. The effect is analyzed by

reconstructing the image with the altered subband coefficients. The original and the reconstructed images are examined to assess the image quality from the point of view of the blood vessels. The qualitative analysis shows the degree of influence of each subband on these features and diagnostic accuracy. For example: the large vessels which form major structures, considered as relatively lower frequency components are influenced more by the lower resolution wavelet subband coefficients. Whereas the fine structures such as thin vessels, considered as relatively higher frequency components and can be extracted from higher resolution images. Since blood vessels are the major components in retinal image, the retinal image quality can be considered to be dependent on blood vessels. To emphasize the diagnostically significant error from the nonsignificant error, different subband errors are weighted differently. The weight for a subband is computed using significant wavelet coefficients which lie within the circular region of interest (CROI) from that subband. The wavelet coefficients that lie outside the CROI are made zero to reduce their effect on the distortion measure. The detailed discussion of the weight computation and the distortion measure is given in [24].

In this paper, we present the results of subjective quality assessment study and evaluate the performance of ten image quality assessment algorithms. The subjective study contained images distorted using four different distortion types and subjective image quality evaluations. This study was distinct in terms of image data sets, contains different types of distortion, distortion strength and subjects participated in the evaluation of distorted images.

3. EXPERIMENTAL DETAILS

This section presents the details of the experiments for subjective study such as reference images used as the original images, distortion sources, the human subjects involved and different image quality assessment algorithms used in the objective evaluation.

3.1 Image database

The image database is derived from two sources: twenty images are considered from the publicly available DRIVE database [25] and ten images are considered from the clinical data obtained from a local eye care hospital¹. They are called as set-1 and set-2 images. These images are resized to (512 x 512) and all distorted images are derived from the resized images.

¹Shri Sankaradeva Netralaya (SSN), Guwahati, Assam, India.

3.2 Types of image distortion

The reference images are distorted using four different types of distortion. In the study of the quality measures, two compression algorithms: the DCT based JPEG [26] and wavelet based set partitioning in hierarchical trees (SPIHT) [27] are used. The other types of distortions are the Gaussian blurring filters (Gblur) and the addition of Gaussian white noise (GWN) at various values of standard deviation. The level of distortion is adjusted manually so that the distorted images obtained spans the entire range of quality.

The Figure 1 shows the variation of the subjective quality (set-1 images only) with the distortion strength for the Gaussian blurring distortion.

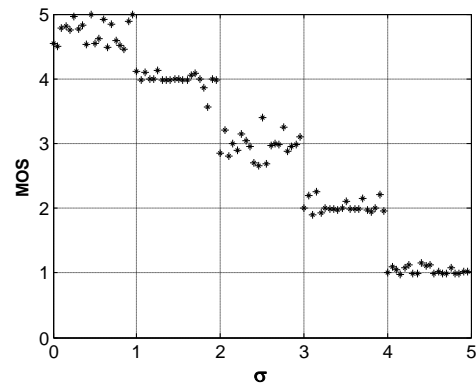


Figure 1: Subjective MOS versus σ for Gaussian blur

3.3 Subjective testing

The subjective quality ratings are obtained using a double stimulus continuous procedure for different distorted images. It is a method to measure the quality of a system relative to a reference. The continuous scale provides the subject to indicate fine degradations in image quality. This is more convenient for the assessor compared to the five-level absolute category rating. The continuous scales are divided into five equal lengths, which correspond to the normal five-point quality scale of [28]. The subjective test is conducted by displaying a pair of images on the screen. The first image is always the original (reference) image without any artifacts and considered to be of excellent quality. The second image is the reconstructed version of the original which possess some difference with respect to original. The assessor is asked to give the score for the second image taking the first image as reference. The absence of defects in the reference part of the presentation pair helps to obtain optimal results. The viewers are allowed to assess without any constraint on viewing distance, time and lighting conditions. The subjective mean opinion score (MOS) values are obtained from retinal specialists and research scholars working in different areas of signal and image processing.

3.3 Objective image quality measures

The objective image quality evaluation is of major concern in the field of medical image processing systems, such as those for acquisition, compression, restoration and enhancement. Over the years, a number of researchers have developed general purpose objective image quality assessment algorithms [2], [3]. The various objective image quality measures investigated in this paper are listed in Table 1.

The table gives the image quality assessment algorithms belonging to different classes. The traditional pixel difference based global quality measures includes MSE and PSNR. They are easy to compute but are not correlated with the visual image quality [29]. The performance of the objective measures can be improved by incorporating the properties of human visual system (HVS). A new image distortion measurement called Structural Similarity (SSIM) based on the degradation of structural information was proposed in [30]. The SSIM metric is based on the assumption that the HVS is highly adapted to extract structural information from natural images. The SSIM is computed locally and then averaged to obtain a single mean SSIM (MSSIM) value. It is proved that

the SSIM can better quantify the perceived image distortion than the traditional MSE or PSNR. Multiscale SSIM (MS-SSIM) method is an extension of the SSIM and offers more flexibility in incorporating the variations of viewing conditions. MS-SSIM is a convenient way to incorporate image details at different resolutions [31]. The SSIM and MS-SSIM takes a value of 1 for perfect similarity and 0 for no similarity. Recently, a method based on wavelets tries to improve the HVS model by incorporating the low-level and mid-level properties of human vision [32]. The computation of VSNR involves two stages. In the first stage, a threshold for the distortion of the degraded images is set using the wavelet based visual masking models. The distorted image is assumed to be perfect ($VSNR=\infty$) if distortions are below the threshold. In the second stage where the distortions are above the threshold, the low level visual property of perceived contrast and the mid level visual property of global precedence are used. These properties are used to determine Euclidean distances in distortion-contrast space of multiscale wavelet decomposition. Then VSNR is calculated as a linear combination of these distances. A higher VSNR indicates that the tested image is less degraded. Information theoretic approaches [33], [34] try to find alternative methods to imitate the HVS. The information fidelity criterion (IFC) [33] considers the mutual information between the wavelet subband coefficients of the original and distorted images to evaluate the visual quality. It theoretically ranges from zero (no fidelity) to infinity (perfect fidelity). The visual information fidelity (VIF) [34] criterion uses two facts of image information to define the perceptual quality: the mutual information and the information content of the reference(original) image itself. This measure can be between 0 and 1, where 1 means perfect quality and 0 means worst quality. The new wavelet based image quality measure IQM is computed by using Watson’s model of noise visibility in different wavelet subbands. The IQM is defined as perceptual weighted difference between coefficients of original and degraded image [17]. The distortion measure for quantifying the distortion in blood vessel structure of retinal images is computed using WBVDM [24].

Table 1. Image quality assessment (IQA) algorithms tested in this paper

IQA Algorithms	Comment
MSE	Pixel difference based
PSNR	Function of MSE
SSIM	[30]
Multiscale SSIM	[31]
VSNR	[32]
IFC	[33]
VIF	[34]
IQM	[17]
WBVDM	[24]

4. EVALUATION RESULTS

Performance of the objective models was evaluated with respect to three aspects of their ability to estimate subjective assessment of image quality [35]. They are:

- prediction accuracy – the ability to predict the subjective quality ratings with low error
- prediction monotonicity – the degree to which the model’s predictions agree with the relative magnitudes of subjective quality ratings and
- prediction consistency – the degree to which the model is robust with respect to a variety of image artifacts.

The relationship between each model predictions and the corresponding subjective ratings was estimated using a nonlinear regression. This removes any nonlinearity due to the subjective rating process and facilitates comparison of the models in a common analysis space. The following logistic function was used to fit the subjective data to the objective data.

$$f(x) = \frac{b1}{1 + e^{-b2(x-b3)}} \quad (1)$$

The nonlinear regression function was used to transform the set of objective measure values to a set of predicted MOS values, which were then compared with the actual MOS values from the subjective tests. Once the nonlinear transformation was applied, the objective model’s prediction performance was then evaluated by computing various metrics.

Metrics relating to Prediction Accuracy of a model:

- Metric 1: The Pearson linear correlation coefficient (PLCC) between predicted MOS and actual MOS.

Metrics relating to Prediction Monotonicity of a model:

- Metric 2: Spearman rank order correlation coefficient (SROCC) between predicted MOS and actual MOS.

Metrics relating to Prediction Consistency of a model:

- Metric 3: Outlier Ratio of outlier-points to total points N.

Twice the MOS Standard Error was used as the threshold for defining an outlier point.

- Metric 4: RMSE between actual MOS and predicted MOS values.

4.1 Performance of image quality measures

The performance of all objective quality measures in terms of the PLCC, the SROCC and the RMSE after nonlinear regression are shown in Table 2-4 respectively, for each distortion type and for the entire database. Figure 2 shows the scatter plots of objective quality measure values versus difference of MOS (DMOS) along with the best fitting logistic function. The scatter plots are shown only for compression (SPIHT) distortion. The results in the three tables demonstrate that the MSE and PSNR have limitations in quantifying the distortion in clinically important features. The IFC and VIF show improved performance compared to SSIM and VSNR. The WBVDM performs best in terms of both PLCC and SROCC among all the quality measures tested in this paper. The WBVDM also shows a smaller RMSE and with all these observations, it can be considered as a better measure to quantify the diagnostic distortion in retinal images.

4.2 Analysis of variance and F-ratio

Analysis of Variance (ANOVA) [36] was used as a statistical tool to evaluate the merits of the quality measures. It compares the means of several groups of observations and determines

whether or not a number of data groups are statistically different. In an ANOVA, the F-ratio is the statistic used to test the hypothesis that the means of all the groups of observations are significantly different from one another. In its simplest form, ANOVA is called one-way analysis of variance. The output of the ANOVA is the identification of those image quality measures that are most consistent and discriminative of the distortion artifacts due to compression, blur and noise. The one-way ANOVA results of the image quality measures for the data obtained from set-1 images and for SPIHT compression artifact are listed in Table 5. The larger value of the F-ratio indicates that the means of all the groups of observations are significantly different and the objective model performance is more close to the subjective evaluation. From the Table 5, it is observed that the WBVDM has the highest F-ratio among all other image quality models. Thus, in terms of statistical significance, the WBVDM outperforms at predicting the retinal image distortion.

5. CONCLUSION

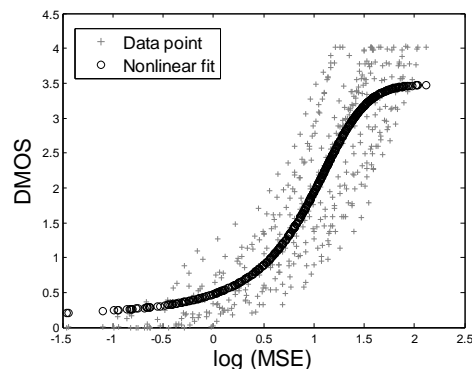
This paper presents a statistical evaluation of the performance of a number of image quality measures in quantifying the distortion in retinal images. It was found that the WBVDM, which is simple to compute and more responsive to diagnostic distortions, shows better performance. The WBVDM can be considered as a method for objective quality assessment of digital retinal images which agrees well with evaluations made by the human observers and may be useful in clinical practice.

6. REFERENCES

- [1] N. Patton, T. M. Aslam, T. MacGillivray, I. J. Deary, B. Dhillon, R. H. Eikelboom, K. Yogesana, I. J. Constable, Retinal image analysis: Concepts, applications and potential, *Progress in retinal and eye research*, 25(1), 2006, 99-127.
- [2] A. M. Eskicioglu, Quality measurement for monochrome compressed images in the past 25 years, 2000, *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 4, 1907-1910, Istanbul, Turkey, Jun 2000.
- [3] H. R. Sheikh, M. F. Sabir, A. C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, *IEEE Tans. Image Process.* 15(11), 3441-3452, 2006.
- [4] P. Cosman, R. M Gray, R. A Olshen. Evaluating the Quality of Compressed Medical Images: SNR, Subjective Rating and Diagnostic accuracy, *Proc. IEEE*, 82 (6), 919-932, 1994.
- [5] S. Lee and Y.Wang, Automatic retinal image quality assessment and enhancement, *Proceedings of SPIE Image Processing*, 1581–1590, 1999.
- [6] M. Lalonde, L. Gagnon, and MC Boucher, Automatic visual quality assessment in optical fundus images, *Proceedings of Vision Interface*, 259-264, 7-9, 2001.
- [7] D. B Usher, M. Himaga, and M. J. Dumskyj, Automated assessment of digital fundus image quality using detected vessel area, *Proceedings of Medical Image Understanding and Analysis*, 81–84, British Machine Vision Association (BMVA), 2003.
- [8] Alan D Fleming, Sam Philip, Keith A Goatman, John A Olson, and Peter F Sharp, Automated assessment of diabetic retinal image quality based on clarity and field definition, *Invest Ophthalmol Vis Sci.*, 47(3), 1120–1125, 2006.
- [9] Meindert Niemeijer, Michael D Abramoff, and Bram van Ginneken, Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening, *Med Image Anal*, 10(6), 888–898, 2006.
- [10] M. J. Cree, H. F. Jelinek, The effect of JPEG compression on automated detection of microaneurysms in retinal images, *Proc. of SPIE-IS&T Electronic Imaging*, 6813, 68130M-68130M 10, 2008.
- [11] P. Hansgen, P. E. Undrill, M. J. Cree, The application of wavelets to retinal image compression and its effect on automatic microaneurysm analysis, *Computer Methods and Programs in Biomedicine*, 56(1), 1-10, 1998.
- [12] R. H. Eikelboom, K. Yogesana, C. J. Barry, I. J. Constable, M. T. Kearney, L. Jitskaia, P. H. House, Methods and Limits of Digital Image Compression of Retinal Images for Telemedicine, *Investigative Ophthalmology and Visual Science*, 41(7), 1916-1924, 2000.
- [13] D. Beauregard, J. Lewis, M. Piccolo, H. Bedell, Diagnosis of glaucoma using telemedicine - the effect of compression on the evaluation of optic nerve head cup-disc ratio, *Journal of Telemedicine and Telecare*, 6, 123-125, 2000.
- [14] J. Conrath, A. Erginay, R. Giorgi, A. Lecleire-Collet, E. Vicaut, J-C Klein, A. Gaudric, P. Massin, Evaluation of the effect of JPEG and JPEG2000 image compression on the detection of diabetic retinopathy, *Eye, Official journal of The Royal College of Ophthalmologists*, 21, 487–493, 2006.
- [15] A. Basu , A. D. Kamal, W. Illahi, M. Khan, P. Stavrou, R. E. J. Ryder, Is digital image compression acceptable within diabetic retinopathy screening?, *Diabetic Medicine*, 20(9), 766 - 771, 2003.
- [16] Y. K. Lai and C. C. J. Kuo, A Harr wavelet approach to compressed image quality measurements, *J. of Visual Commun. and Image Representation*, 11, 17-40, 2000.
- [17] E. Domic, S. Grgic, M. Grgic, New image-quality measure based on wavelets, *J. of Electronic Imaging* 19(1), 011-018, 2010.
- [18] Z. Gao and Y. F. Zheng, Quality constrained compression using DWT-based image quality metric, *IEEE Trans. Circuits and Systems for video technology*, 18(7), 910-922, 2008.
- [19] M. Unser, A. Aldroubi, A review of wavelets in biomedical applications, in *Proc. IEEE*, 84(4), 626 -638, 1996.
- [20] S. Mallat, S. Zhong, Characterization of signals from multi-scale edges, *IEEE Trans. Pattern Analysis and machine Intelligence*, 14 (7), 710-732, 1992.
- [21] B. Erickson, A. Manduca, P. Palisson, K. Persons, F. Earnest, IV, V. Savcenko, N. Hangiandreou, Wavelet compression of medical images, *Radiology*, 206, 599-607, 1998.
- [22] H. Ringl, R. E. Scherthaner, A. A. Bankier, M. Weber, M. Prokop, C. J. Herold, C. S. Prokop, JPEG2000 compression of thin-section CT images of the lung:

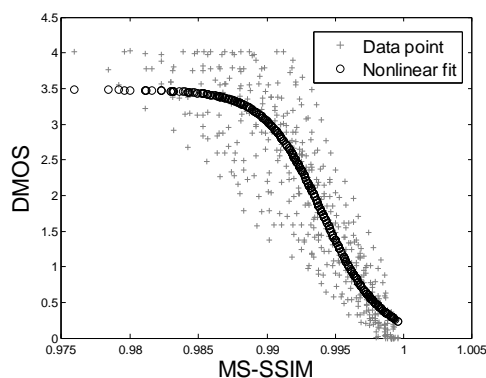
Effect of compression ratio on image quality, *Radiology*, 240, 869-877, 2006.

- [23] TJ Chen, KS Chuang, Wu J, SC Chen , IM Hwang, ML Jan, Quality degradation in lossy wavelet image compression, *J Digit Imaging* 16(2), 210-215, 2003.
- [24] S. R. Nirmala, S. Dandapat, P. K. Bora, Wavelet weighted blood vessel distortion measure for retinal images, *Biomedical Signal processing and Control*, 5(4), 282-291, 2010.
- [25] M. Niemeijer, J. J. Staal, B. van Ginneken, M. Loog, M.D. Abramoff, Comparative study of retinal vessel segmentation methods on a new publicly available database, *SPIE Medical Imaging*, 5370, 648-656, 2004.
- [26] G. K. Wallace, The JPEG still picture compression standard, *IEEE Trans. Consum. Electron*, 38(1), 18–34, 1992.
- [27] A. Said and W. A. Pearlman, A new fast and efficient image codec based on set partitioning in hierarchical trees, *IEEE Trans. Circuits Syst. Video Technol.* 6(3), 243–250, 1996.
- [28] ITU-R Recommendation BT.500-11, Methodology for the subjective assessment of the quality of the television pictures, International Telecommunication Union, Geneva, Switzerland, 2002.
- [29] S. Wong, L. Zaremba, D. Gooden, H. K. Huang, Radiological image compression: A review, *Proc. IEEE*, 83 (2), 194-219, 1995.
- [30] Z. Wang, A. C. Bovik, H. Sheikh, E. P. Simoncelli, Image Quality Assessment: From Error Visibility to Structural Similarity, *IEEE Trans. Image Process.*, 13(4), 600-612, 2004.
- [31] Z. Wang, E. P. Simoncelli, and A. C. Bovik, Multi-scale structural similarity for image quality assessment, *IEEE Asilomar Conf. Signals, Systems, and Computers*, Nov. 2003.
- [32] D. M. Chandler and S. S. Hemami, VSNR: a wavelet-based visual signal-to-noise ratio for natural images, *IEEE Trans. Image Process.* 16(9), 2284–2298, Sept. 2007.
- [33] H. R. Sheikh, A. C. Bovik, and G. de Veciana, An information fidelity criterion for image quality assessment using natural scene statistics, *IEEE Trans. Image Process.*, 14, (12), 2117–2128, 2005.
- [34] H. R. Sheikh and A. C. Bovik, Image information and visual quality, *IEEE Trans. Image Process.*, 15(2), 430–444, 2006.
- [35] Final report from the video quality experts group on the validation of objective models of video quality assessment, PHASE-II, Aug 2003.
- [36] A. C. Rencher, *Methods of Multivariate Analysis*, New York, John Wiley (1995).

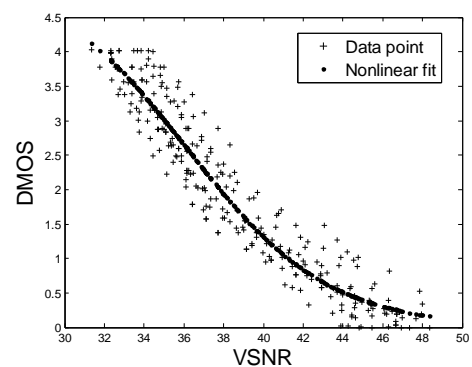


(a)

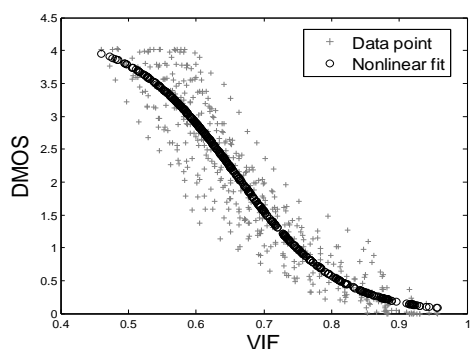
Figure 2. Scatter plots showing the nonlinear fit for the a) MSE (b) MS-SSIM (c) VSNR (d) VIF (e) IQM and (f) WBVDM versus DMOS.



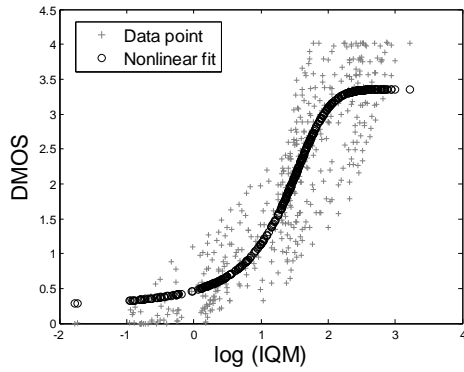
(b)



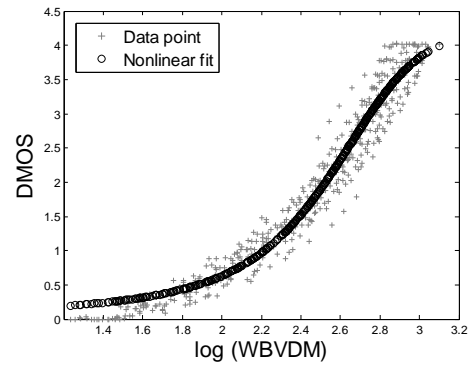
(c)



(d)



(e)



(f)

Table 2. Pearson linear correlation coefficient(PLCC) after nonlinear regression

Distortion \ Metrics	MSE	PSNR	SSIM	MS-SSIM	VSNR	IFC	VIF	IQM	WBVDM
JPEG	0.8653	0.8746	0.9028	0.9051	0.9567	0.9414	0.9334	0.8890	0.9604
SPIHT	0.8714	0.8744	0.8932	0.8911	0.9441	0.9143	0.9199	0.8818	0.9662
Gblur	0.8456	0.8628	0.9401	0.9415	0.9505	0.9690	0.9663	0.8770	0.9695
GWN	0.8354	0.8490	0.8623	0.8601	0.9475	0.9252	0.9203	0.8296	0.9720

Table 3. Spearman Spearman rank order correlation coefficient (SROCC) after nonlinear regression

Distortion \ Metrics	MSE	PSNR	SSIM	MS-SSIM	VSNR	IFC	VIF	IQM	WBVDM
JPEG	0.8754	0.8754	0.9087	0.9211	0.9592	0.9423	0.9360	0.9036	0.9697
SPIHT	0.8737	0.8737	0.8942	0.8943	0.9434	0.9155	0.9210	0.8876	0.9659
Gblur	0.8663	0.8663	0.9481	0.9549	0.9590	0.9721	0.9703	0.9034	0.9756
GWN	0.8511	0.8511	0.8659	0.8684	0.9505	0.9282	0.9217	0.8471	0.9766

Table 4. Root mean squared error (RMSE) after nonlinear regression

Distortion \ Metrics	MSE	PSNR	SSIM	MS-SSIM	VSNR	IFC	VIF	IQM	WBVDM
JPEG	0.6247	0.6051	0.5358	0.5302	0.3646	0.4208	0.4475	0.5705	0.3502
SPIHT	0.6038	0.5976	0.5537	0.5591	0.4048	0.4989	0.4831	0.5822	0.3173
Gblur	0.6305	0.5980	0.4022	0.3980	0.3666	0.2922	0.3037	0.5671	0.2896
GWN	0.6786	0.6526	0.6252	0.6298	0.3952	0.4690	0.4830	0.6894	0.2929

Table 5. ANOVA results (F-scores) for the SPIHT compression distortions

Metrics	MSE	PSNR	SSIM	MS-SSIM	VSNR	IFC	VIF	IQM	WBVDM
SPIHT	85	87	96	108	236	118	142	78	288