Assessment of Diverse Quality Metrics for Medical Images including Mammography

T.Venkat

JNTU Kakinada and GNITC, Hyderabad, A.P, India

Narayana Rao

CSE, JNTUH, Hyderabad, A.P, India

Abstract

This paper presents the comparative analysis of various quality metrics for medical image processing. Measurement of image quality is vital for numerous image-processing applications. Image quality measurement is closely related to image resemblance assessment in which quality is based on the differences (or similarity) between a degraded image and the original, unmodified image chiefly in mammographic images. Simple verifiable techniques have been used for representing the image quality because we have employed automatic and mathematically defined procedure distinct detailed methods, which are costly, time consuming and observer dependent. In this paper images have been subjected to various degrees of blur, noise, compression and contrast levels and quality has been measured in terms of well known metrics such as Mean Squared Error (MSE), Structural Similarity Index Metrics (SSIM) Peak Signal-to-Noise Ratio (PSNR), Maximum difference (MD) including new measures of image qualities for low cost medical image analysis.

Keywords

Image quality analysis, Mean Square Error (MSE), Structural Similarity Index Metric(SSIM), Peak Signal to Noise Ratio(PSNR), Mean Absolute Error(MAE)

1. INTRODUCTION

Digital images have become very vital entities in medical analysis and diagnosis for mammography and scanned image techniques. Digital images are subjected to a broad variety of distortions during acquisition, processing, storage, compression, transmission and reproduction, any of which may effect in a degradation of visual quality [1]. Identifying the image quality measures that have highest sensitivity to these, distortions would help in systematic design of coding, communication and imaging systems and improvising or optimizing the picture quality for a desired quality of service at a minimum cost. For image quality measurement, there are basically two approaches:-

    1) Subjective measurements.
    2) Objective measurements.

Subjective measurements are the result of human experts providing their view of the image quality and objective measurements are performed with mathematical algorithms. In the applications in which images are ultimately to be viewed by human beings is the only exact method of quantifying visual image quality, through subjective evaluation. In reality, subjective evaluation is usually too inconvenient, time-consuming and expensive. The motto of research in objective image quality assessment is to extend quantitative measures that can automatically foresee perceived image quality [2].

An objective image quality metric can play a variety of roles in image processing applications. It can be used to animatedly supervise and adjust image quality. This approach could be used to optimize algorithms and parameter settings of image processing systems. It can be used to benchmark image processing systems and algorithms [3].

Objective image quality metrics can be classified according to the availability of an original (distortion-free) image, with which the distorted image is to be compared. The majority approaches are known as full-reference, this mean that a complete reference image is assumed to be known. In numerous practical applications the reference image is not accessible, and a no-reference or "blind" quality assessment approach is desirable. In third type of method, the reference image is only partially accessible, in the form of a set of extracted features made available as side information to help evaluate the quality of the distorted image. This is denoted to as reduced-reference quality assessment [4]. This new method focuses on full-reference image quality assessment.

The simplest and most extensively used full-reference quality metric is the mean squared error (MSE), computed by averaging the squared intensity differences of distorted and reference image pixels, along with the popular quantity of Peak Signal-to- Noise Ratio (PSNR). These are appealing because they are straightforward to calculate, have clear purpose and meaning and are mathematically suitable in the background of optimization. However, they are not extremely well matched to apparently visual quality. MSE and PSNR lack a critical feature and the ability to assess image similarity across distortion types. In the past three decades, a great deal of attempt one has to put to development of quality assessment methods that take the advantage of known characteristics of the human visual system (HVS) [5]. The majority of the proposed perceptual quality assessment models have followed a strategy of modifying the MSE measure so that errors are penalized in

agreement with their visibility. Among many of objective measures of picture quality, based on quantifiable distortion measures, illustrated in Table I.

Where, x(i, j) represents the original (reference) image and (i, j) represents the distorted (modified) image. Two Human visual systems (HVS) based image quality metrics are the universal image quality index (Q) and the Structural Similarity Index (SSIM) [6]. .

2. Popular Methods To Measure The Quality Of Image

Universal image quality metric: Let x = { xi │ i= 1,2,3,..., N} , y= { yi │ i= 1,2,3,..., N } be the original and the test images, respectively.

$$Q = \frac{4 \times \sigma_{xy} \times \bar{x} \times \bar{y}}{(\sigma_x^2 + \sigma_y^2) \times ((\bar{x})^2 + (\bar{y})^2)}$$

$\bar{x}$, $\bar{y}$, σ_x^2, σ_y^2 and σ_{xy} are given as:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i$$

$$\sigma_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$$

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$$

The dynamic range of Q is [-1, 1], with the best value achieved when yi = xi, i = 1,2,…,n. The index is computed for every window, which leads to a quality map of the image. The total quality index is the average of all the Q values in the quality map

$$Q = \frac{1}{M} \sum_{j=1}^{M} Q_j$$

M = total number of windows.

SSIM (Structural similarity index metric): Q produces unstable results when either $\left((\bar{x})^2 + (\bar{y})^2 \right)$ or ($\sigma_x^2 + \sigma_y^2$) is very near to zero. In order to avoid this problem, the measure has been generalized to the Structural Similarity Index (SSIM): [1]

$$SSIM = \frac{(2 \times \bar{x} \times \bar{y} + C1)(2 \times \sigma_{xy} + C2)}{(\sigma_x^2 + \sigma_y^2 + C2) \times ((\bar{x})^2 + (\bar{y})^2 + C1)}$$

C1 and C2 are constant. As in the case of Q, the general image quality MSSIM is acquire by computing the average of SSIM values over all windows:

$$MSSIM = \frac{1}{M} \sum_{j=1}^{M} SSIM_j$$

M = total number of windows.

3. THE COMPARATIVE PROPOSED APPROACH

The evaluation of image quality is the central aim for many image processing systems, such as those for enhancement, compression, transmission and also reproduction. A great deal of attempt has gone into designing quality assessment methods that take advantage of known characteristics of the HVS. Natural image signals are highly structured. The most basic principle to image quality assessment is that the HVS is highly adapted to extract structural information from the visual scene, and consequently a measurement of structural similarity (or distortion) should provide a good approximation to perceptual image quality. Basing on how structural information and structural distortion are defined, there may be different means to develop image quality assessment algorithms. Following steps are taken in proposed approach: 1. Study the existing metrics. 2. Simulate the various quality metrics with MATLAB. 3. Execute the various quality metrics by augmenting noise, blur, compression, and contrast applications with medical images such as mammographic and MRI etc.

4. rEsults and Findings

The functions illustrated in table I are being evaluated with diverse inputs as shown. Based on this a comparative evaluation of various quality metrics has been done with three different kinds of images, which are commonly used in areas of medical image processing:
A. MRI image (noise removal and with compression)
B. X-Ray image (for blur image applications)
C. Ultra sound image (with contrast application)

MRI image has been chosen for application of noise. These types of images have small size particulars, which are vulnerable to random noise. Gaussian noise has been selected for adding noise to the image. In Gaussian Noise, power is directly proportional to the variance. Therefore, different values of noise variances are taken for the purpose of adding different level of noise powers to the image. The images corresponding to diverse noise powers are shown in Figure 1 and values of various quality metrics is shown in Table II.

As can be seen from Table II, all quality metrics acknowledge the changes to the noise powers variation. MD appears to be the mostly sensitive metric whereas SC & PMSE being the least sensitive to the variation in noise power. From the Tables it is obvious that SSIM give results with greater accuracy.
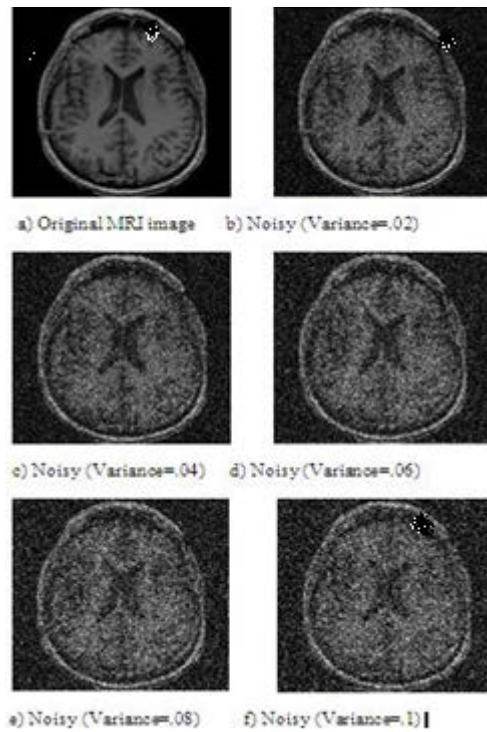
Figure 1. Different MRi Images with variance

Figure 2. Different x-ray blurred variance images

Table III and Figure 1 shows different values of quality metrics for MRI compression image. The compression depend upon the block processing means image is bigger we can use smaller block of image at a time .To compress the image, the first step is to subdivide the input image into non overlapping pixel blocks of size 8×8. They are subsequently processed left to right, top to bottom. The two-dimensional DCT is calculated for every block. The DCT coefficients are then quantized, coded, and transmitted. By applying block processing image signal firstly converted from space domain to frequency domain, then we use mask as a frequency domain filter. By multiplying frequency domain filter to frequency, domain signal and we get compression signal, which will depend upon the no. of elements in the filter. Table III depicts that by decreasing no. of elements are less in filter percentage of compression is changes [7].
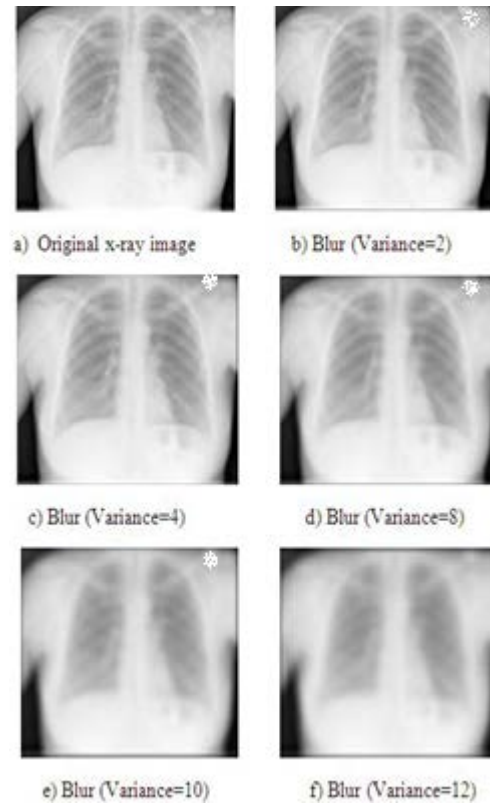
The images corresponding to different degrees of blur have been shown in Figure 2 and values of various quality metrics is shown in Table IV. From the table it is understandable that most of the quality metrics acknowledge the change in degree of blur except NK, SC , MD, and PMSE. So MSE, AD, MAE ,PSNR, IF and SSIM are additional significant whereas MD, NK, SC and PMSE have little significance in image blur applications. As depicted in Table IV it is apparent that SSIM give results with greater consistency[8] . The Contrast of an image shows the distribution of pixel values of an image from minimum to maximum values. A low contrast image has this distribution confined in a small range whereas a high contrast image has pixel values distribution over the full range from minimum to maximum probable pixel value. We have chosen an ultrasound image for variation in its contrast. Contrast of the image has been varied with the help of function, which is inherent in MATLAB. It changes the limiting range of pixel of output image as per specification in its arguments as shown in figure 3.

a) Original ultrasound image   b) Contrast (Stretching limits=0.1-0.9)

c) Contrast (Stretching limits=0.2-0.8)   d) Contrast (Stretching limits=0.3-0.7)
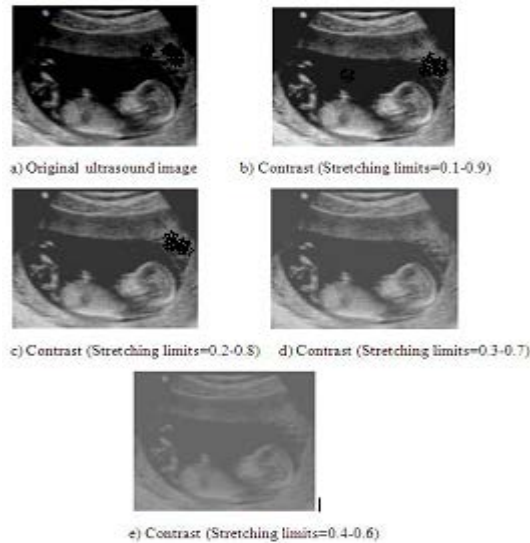
e) Contrast (Stretching limits=0.4-0.6)

Figure 3. Ultrasound Images with limits.

From the Table IV it is clear that most of the quality metrics recognize the change in degree of contrast except NK and IF. So MSE, PSNR, AD, MD, MAE, PMSE and SSIM are more significant whereas, NK, IF and SC have little significance in ultrasound image contrast applications. As can be seen from the table SSIM has elevated accuracy of results and grater values of SSIM indicate greater image similarity.

5. CONCLUSION

In this paper, comparative analysis of various quality metrics has been explored. Three different kinds of medical application images have been processed for different levels of compression, noise powers, blur and contrast. All image quality metrics described in this paper have been calculated for all four types of image processing operations. Output results have been evaluated for comparison of various quality metrics. These result shows that only a section i.e. a subset of quality metrics is suited for a particular type of image processing operation except the SSIM metric which is capable of expressing image quality irrespective of the kind of operation. Other quality metrics are easy but not the accurate for all kinds of image processing operations. SSIM has high performance, works precisely and provides a high-quality approximation of quality measurement however with moderately bigger computational time. The above work can be applied to meticulously analyze much impending breast cancer through mammographic diagnosis and detection cancer of type I and II with a lower cost.

REFERENCES

[1] Z.Wang, A.C Bovik, H.R Sheikh, and E.P Simoncelli , 2004. Image Quality Assessment: From Error Visibility to Structural Similarity. IEEE Transactions of Image Processing, (April 2004), vol. 13 1-12

[2] Zhou Wang and Alan C. Bovik, 2002. A Universal Image Quality Index. IEEE Signal Processing Letters, vol. 9, No. 3, 81-84.

[3] Z. Wang, A. C. Bovik, and L. Lu, 2002. Why is image quality assessment so difficult. In Proceedings of IEEE Int.Conf. Acoustics, Speech, and Signal Processing, (May 2002), vol. 4. 3313–3316.

[4] Eric Silva, Karen A. Panetta, Sos S. Agaian, 2007. Quantify similarity with measurement of enhancement by entropy, Proceedings: Mobile Multimedia/Image Processing for Security Applications, SPIE Security Symposium (April 2007), Vol. 6579, l.

[5] Z. Wang, H. R. Sheikh, and A. C. Bovik, 2003. Objective video quality assessment. The Handbook of Video Databases: Design and Applications, B. Furht and O. Marques, Eds. Boca Raton, FL: CRC Press,.

[6] Alan C. Brooks, Xiaonan Zhao, and Thrasyvoulos N. Pappas, 2008 .Structural Similarity Quality Metric in a Coding Context: Exploring the Space of Realistic Distortions, IEEE Transactions on image processing, (August 2008)Vol. 17, No. 8.

[7] A. M. Eskicioglu and P. S. Fisher, 1995 . Image quality measures and their performance, IEEE Trans. Commun.,(Dec 1995), vol. 43, 2959–2965.

[8] Sonja Grgic , Mislav Grgic,and Marta Mrak, 2004. Reliability Of Objective Picture Quality Measures, Journal of Electrical Engineering, (October 2004),Vol. 55, No. 1-2.

Table I. Conventional and new Metrics of Quality

	TYPE	DESCRIPTION
1.	**MSE(Mean Square Error)**	$\dfrac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(x(i,j)-y(i,j)\right)^2$
2.	**PSNR(Peak Signal to Noise Ratio)**	$PSNR = 10\log_{10}\dfrac{(2^n-1)^2}{\sqrt{MSE}}$
3.	**AD(Average Deviation)**	$AD = \dfrac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(x(i,j)-y(i,j))$
4.	**MD(Maximum Difference)**	$MD = MAX\,|(x(i,j)-y(i,j)|$
5.	**MAE(Mean Absolute Error)**	$MAE = \dfrac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}|(x(i,j)-y(i,j)|$

	TYPE	DESCRIPTION
6.	**NK(Normal Cross-Correlation)**	$NK = \dfrac{\sum_{i=1}^{M}\sum_{j=1}^{N}((x(i,j)-y(i,j))}{\sum_{i=1}^{M}\sum_{j=1}^{N}(x(i,j))^2}$
7.	**SC(Structural Content)**	$SC = \dfrac{\sum_{i=1}^{M}\sum_{j=1}^{N}(y(i,j))^2}{\sum_{i=1}^{M}\sum_{j=1}^{N}(x(i,j))^2}$
8.	**IF(Image Fiedelity)**	$IF = 1 - \dfrac{\sum_{i=1}^{M}\sum_{j=1}^{N}(x(i,j)-y(i,j))^2}{\sum_{i=1}^{M}\sum_{j=1}^{N}(x(i,j))}$
9.	**PMSE(peak Mean Square Error)**	$PMSE = \dfrac{1}{MN}\times\dfrac{\sum_{i=1}^{M}\sum_{j=1}^{N}(x(i,j)-y(i,j))^2}{(MAX(x(i,j))^2}$
10.	**SSIM(Structural Similarity Index Metrics)**	$SSIM = \dfrac{(2\times\bar{x}\times\bar{y}+C1)(2\times\sigma_{xy}+C2)}{(\sigma_x^2+\sigma_y^2+C2)\times((\bar{x})^2+(\bar{y})^2+C1)}$

Table II. Comparison of Quality Metrics for Noisy MRI Image

Variance	MSE	PSNR	AD	MD	MAE	NK	SC	IF	PMSE	SSIM
0.02	58.4048	39.2986	-6.6495	153	22.4545	1.0146	1.3525	.7066	.0012	.2036
0.04	63.9898	39.1042	-10.0501	190	30.5875	1.0263	1.5972	.4554	.0012	0.1327
0.06	65.2490	39.0579	-13.766	239	36.6935	1.0598	1.9323	.2172	.0013	.1072

| 0.08 | 66.5353 | 39.0155 | -16.6531 | 255 | 41.4679 | 1.0829 | 2.1662 | 0.0084 | .0014 | .0882 |
| 0.1 | 68.3621 | 38.9567 | -18.5784 | 255 | 45.1500 | 1.0900 | 2.3725 | -0.1925 | .0013 | .0741 |

Table III. Comparison of Various Quality Metrics for Compressed MRI Image

Compression%	MSE	PSNR	AD	MD	MAE	NK	SC	IF	PMSE	SSIM
15	42.1701	40.0568	.1255	162	10.7485	.8843	.8845	.8852	7.9233E-04	.5804
13	29.1606	40.8143	-.1179	152	7.1392	.9443	.9416	.9459	5.5913E-04	.7564
11	20.7699	41.5415	-.0837	137	5.0232	.9681	.9617	.9694	3.9982E-04	.8579
9	14.5008	42.1790	-.0326	114	3.7552	.9806	.9776	.9818	2.9813E-04	.9184
7	12.0426	42.9154	-.0148	92	2.7915	.9885	.9865	.9885	2.1241E-04	.9518

Table IV. Comparison of Various Quality Metrics For dissimilar Contrasts of Ultrasound Image

Stretch limits	MSE	PSNR	AD	MD	MAE	NK	SC	IF	PMSE	SSIM
0.1-0.9	.00289	60.8371	-31.3843	45	31.3858	1.3491	1.8728	.8241	5.1434E-08	.6849
0.2-0.8	.2238	51.3915	-41.5258	51	41.5541	1.3248	1.9363	1.0040	3.7094E-06	.6130
0.3-0.7	1.9789	46.6198	-52.1702	79	52.5986	1.2892	2.1567	.4233	3.5211E-05	.5215
0.4-0.6	10.607	43.1039	-62.4971	101	64.9350	1.2503	2.4578	1.0030	1.7661E-04	.3726