Sensors & Transducers, Vol. 169, Issue 4, April 2014, pp. 193-198

Sensors & Transducers

Analysis of Image Quality Assessment with Markov Random Field Oriented on Low Dose CT Images

[1] Jiehang DENG, [2] Min QIAN, [2] Guoqing QIAO, [1] Xiaoping LIN, [1] Chao Wang

[1] School of Computers, Guangdong University of Technology Guangzhou, 510006, China
[2] X-ray Department, General Hospital of Guangzhou Military Command Guangzhou, 510010, China
[1] Tel.: +86-020-39322279, fax: +86-020-39322279
E-mail: dengjiehang@gdut.edu.cn

Received: 29 March 2014 Accepted: 28 April 2014 Published: 30 April 2014

Abstract: In order to provide precision and objective image quality measures (IQMs) for the low dose CT (Computed Tomography) images, various general IQMs need to be validated and analyzed. The IQM based on Markov Random Field (MRF) has not been check and validated by a comprehensive distorted database. First choose a standard distorted image database of LIVE (Laboratory for Image & Video Engineering) to validate and analyze the performance of various IQMs. Then assess various low dose phantom CT images by the IQMs. Experimental results show that the mutual information based on MRF is more obvious and precision than other measures to reflect the quality changes of the LIVE database and low dose CT phantom images. It can provide effective reference for quality assessment of low dose CT images. *Copyright © 2014 IFSA Publishing, S. L.*

Keywords: Low dose CT, Image quality assessment, Markov random field, Mutual information.

1. Introduction

The CT (Computed Tomography) plays an important role in disease prevention and clinical diagnosis [1]. Compared with other imaging modalities, CT has the advantages of higher resolution and faster imaging. However, CT examination of the high-dose X-ray radiation may induce cancer, leukemia or other genetic diseases [2]. To mitigate this hazard, it is an efficient way to lower the dose level of the X-ray exposure. On the same time, the CT imaging quality degrades with the decrease of the X-ray exposure. How to maintain the diagnosis quality in the low dose CT images is the key point in the medical CT researching. The performance analysis of the image quality assessments can be used to validate and optimize the algorithms of image quality improvement on low dose CT images.

The image evaluation methods can be categorized into two groups. One is subjective human evaluation, the other is objective measures [3, 4]. In the subjective human evaluation, the radiologists need to evaluate the images repeatedly. That is a costly, time-consuming job. The evaluated result is vulnerable to the background knowledge, motivation and clinical experience for different radiologists. The observation may be affected by the observation distance, light conditions, effects of fatigue and other objective factors [5].

With respect to the subjective methods, objective quality assessment algorithms are simple, low cost,

easy to parse. In this paper, the objective assessment methods are discussed and analyzed only. The objective quality assessment algorithms are represented by many image quality measures (IQMs). The IQMs can be divided into two major groups: those based on the human visual system (HVS) and those based on arbitrary signal fidelity criteria [6].

The methods based on HVS are modeled by simulating its low-level structures in a bottom-up way. These researches are focused on how to enhance the consistency between objective evaluation of quality and the subjective one. Wei Xuehui [7] constructed respond functions based on the extracted perceived intensity, frequency and edges. He integrated these functions into an equation and established an image quality assessment mode1.

However, HVS is an extremely complex system. Simulating the composition and structure of the HVS usually causes the high complexity and large amount of computation for the algorithms [8]. The HVS based algorithms did not outperform the widely used Mean Square Error (*MSE*), Peak Signal to Noise Ratio (*PSNR*) [9] and other error standard based metrics in the aspect of the overall performance.

Among the objective image quality evaluation methods, *MSE*, *PSNR*, structural similarity (*SSIM*) [10], etc. are widely used. The methods such as *MSE*, *PSNR* are mainly based on the difference of gray-scale images between the distorted image and the original one. They only consider the images to be comprised of different isolated pixels. The spatial correlations between pixels are not taken into account in these methods. The relationship between the human visual characteristics and the pixels are ignored also. That leads to the poor consistency between subjective perception results and objective evaluated ones. Wang et al. [10] proposed *SSIM* after analyzing the limitations of HVS based methods. SSIM assess the image quality in three comparisons: luminance, contrast and structure. However, it fails to measure the blurred images with a lot of flat regions [11].

Espen Volden [12] proposed a clique-vector model by combining the mutual information theory and the Markov Random Field (MRF) to evaluate the image quality. He applied this metric to assess the redundancy between two images and achieve inspiring results. However, this index has not been tested and verified by a comprehensive image database yet. In this paper, the image database of the second edition established by the laboratory of image and video engineering (LIVE) provided by the University of Texas [13] is applied to test and validate the IQMs' performance. Espen's IQM is compared and analyzed with *MSE*, *PSNR* and *SSIM*. In order to analyze and evaluate the low dose CT images, all the IQMs in this paper are applied to assess the low dose phantom CT images, further.

2. The Algorithms of Image Quality Assessment

2.1. MSE

The *MSE* can be expressed as:

$$MSE = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[f`(x,y) - f(x,y)]^2, \quad (1)$$

where M, N are the height and the width of an image, respectively. $f(x, y)$ and $f'(x, y)$ are the intensity of coordinates (x, y) both in the reference and the test images, respectively.

2.2. PSNR

The *PSNR* can be expressed as:

$$PSNR = 10 \lg \frac{L^2}{MSE}, \quad (2)$$

where L is the maximum gray level of images. It is 255 for eight bits images.

2.3. SSIM

The *SSIM* can be expressed as

$$SSIM(f, f`) = [l(f, f`)^{\alpha}[c(f, f`)]^{\beta}[s(f, f`)]^{\gamma}, \quad (3)$$

where f, f' denote the reference image and the test image. $l(f, f')$ represents the luminance comparison factor. $c(f, f')$ denotes the contrast comparison factor, $s(f, f')$ is the structure comparison factor. This model can be adjusted by the parameters α, β and γ, respectively.

2.4. The Mutual Information *I*

The mutual information I can be expressed as

$$\begin{aligned} I(f, f`) &= H(f) - H(f \mid f`) = \\ &= \sum_{g_1 \in \wedge}\sum_{g_2 \in \wedge} P(f = g_1, f` = g_2)\log\frac{P(f = g_1, f` = g_2)}{P(f = g_1)P(f` = g_2)} = \\ &= \sum_{g_1 \in \wedge}\sum_{g_2 \in \wedge} h_{ff`}(g_1, g_2)\log\frac{h_{ff`}(g_1, g_2)}{h_f(g_1)h_f(g_2)}, \end{aligned} \quad (4)$$

where f, f' represent the reference and the test images, also. The pixels can take value g from the set Λ. For images with eight bits, $\Lambda = \{0, 1, 2, 3,..., 254, 255\}$. $h_f(\bullet)$ is the normalized histogram of the image f. $h_{ff'}(\bullet)$ is the normalized joint histogram between images f and f'. $P(\bullet)$ denotes the probability. $H(f)$ represents the entropy of the image f. $H(f|f')$ is the conditional entropy between f and f'.

2.5. The MRF Based Mutual Information I_{MRF}

Since the pixels of an image are not isolated, they correlate with each other in a neighborhood. The MRF can be used to analyze the correlation between physical phenomena in space or time. Therefore, MRF was applied to analyze the correlation in the pixels neighborhood.

I_{MRF} can be depicted as

$$I_{MRF}(f,f') = H(f) + H(f') - H(f,f')$$
$$= \sum_{g_1 \in \Lambda} \sum_{g_2 \in \Lambda} \sum_{g_1 \in \Lambda^v} \sum_{\vec{g_2} \in \Lambda^v} P(f_{s_1},...,f_{s_v},f_s,f'_{s_1},...,f'_{s_v},f'_s) = (\vec{g_1},g_1,\vec{g_2},g_2)] \times$$
$$\log \frac{P(f,f') = (g_1,g_2) \mid (f_r,f'_r), r \in V_s]}{P(f = g_1 \mid f_r, r \in V_s) P(f' = g_2 \mid f'_r, r \in V_s)}$$
$$= \sum_{g_1 \in \Lambda} \sum_{g_2 \in \Lambda} \sum_{g_1 \in \Lambda^v} \sum_{\vec{g_2} \in \Lambda^v} h_{ff'}(\vec{g_1},g_1,\vec{g_2},g_2) \log \frac{h_f(\vec{g_1}) h_{f'}(\vec{g_2}) h_{ff'}(\vec{g_1},g_1,\vec{g_2},g_2)}{h_f(\vec{g_1},g_1) h_{f'}(\vec{g_2},g_2) h_{ff'}(\vec{g_1},\vec{g_2})} \quad (5)$$

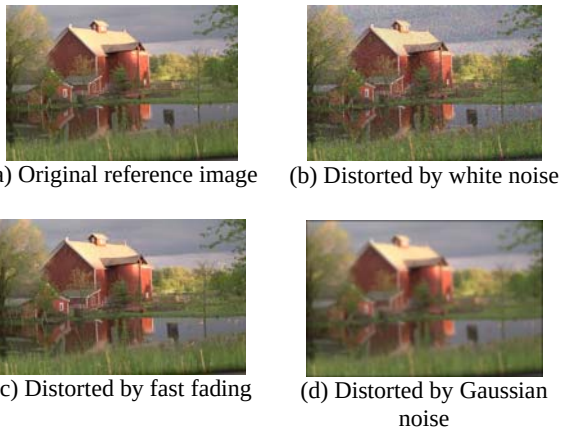
where f, and f' are the reference and the test images, respectively. V_s is the neighborhood of the s pixel, denoted by a vector $\vec{g} = (g_1,...,g_v)$. The other symbols are similar to Eq. (4). The nearest four neighbors of s are considered in this paper.

3. Experimental Results and Discussions

Without loss of generality, the LIVE standard database of images in the second version provided by University of Texas is applied to analyze the performance of IQMs in this paper. In the image database, images from different scenes are shown in original and the corresponding distorted ways. For each of the image, a subjective difference score DMOS (Difference Mean Opinion Scores) was provided. Researchers can check the IQMs' performance by comparing their values and the corresponding DMOS. The white noise, fast fading and Gaussian blur images in LIVE database are applied in this paper. One of the original and the corresponding distorted images in the database are shown in Fig. 1.

Generally speaking, it is acceptable for IQMs to stably predict subjective quality with a nonlinear mapping [14]. In fact, this nonlinear mapping is allowed in the testing and validating experiments of Video Quality Experts Group (VQEG) [15, 16]. The logarithm mapping is applied in this paper.

According to the standard of VQEG, four objective performance metrics are applied to evaluate the performance of IQMs. They are correlation coefficient (CC), mean absolute error (MAE), root mean square error (RMSE) and outlier ration (OR). The larger of the CC value, the higher correlation between the nonlinear mapped IQM and the DMOS. The values of MAE and RMSE represent the residual error between the nonlinear mapped IQMs and DMOS. The smaller of MAE and RMSE, the higher performance of the IQM. The OR value denotes the consistency between the mapped IQMs and DMOS. The smaller of OR, the better of the IQM. In order to show the relationship between the mapped IQMs and DMOS in detail, scatter maps are drawn in this paper, also.



(a) Original reference image (b) Distorted by white noise

(c) Distorted by fast fading (d) Distorted by Gaussian noise

Fig. 1. Certain images of original reference and distorted images.

3.1. Experiments for the Distorted Images by White Noise

There are 145 images in the database of white noise. Their experimental results are shown in Table 1. The parameters of *SSIM* are the same to Ye's settings [17]. In Table 1, the CC value between mutual information I and the DMOS is 0.9659. It is larger than all the value of *MSE*, *PSNR* and *SSIM*. For the values of MAE, RMSE and OR, the I values are less than *MSE*, *PSNR* and *SSIM*. That means the evaluated results of I are closer to the subjective results DMOS. This indicates that it is effective to evaluate the image quality from the information theory aspect. After introducing the clique-vector model to the I, the correlations among the neighborhood influence the value of the mutual information obviously, i.e. I_{MRF}. It leads to that the CC value of I_{MRF} is biggest and it is closest to the DMOS. This validates that it is more accurate to use the I_{MRF} to assess the image quality in the distorted database by the white noise.

Table 1. CC, MAE, RMSE and OR values between nonlinear mapped IQMs and DMOS for the distorted database by white noise.

Model	CC	MAE	RMSE	OR
MSE	0.8528	6.7814	8.3373	0.0601
PSNR	0.8778	6.2497	7.6486	0.0547
SSIM	0.8747	6.5101	7.7392	0.0488
I	0.9659	3.5100	4.1357	0.0395
I_{MRF}	0.9752	2.8434	3.5372	0.0332

3.2. Experiments for the Distorted Images of Fast Fading

There are 145 images in the distorted database of fast fading, also. The CC, MAE, RMSE and OR values between the mapped IQMs and DMOS are shown in Table 2. In Table 2, the CC value of I_{MRF} is larger than the rest of the CC values of other IQMs. The MAE, RMSE and OR values of I_{MRF} are 6.9105, 8.4250 and 0.0414, respectively, which are the lowest ones after comparing to the rest of the IQMs. All the CC, MAE, RMSE and OR values of I_{MRF} show that the evaluated results of I_{MRF} are closest to the subjective human evaluation in the distorted database of fast fading, also.

Table 2. CC, MAE, RMSE and OR values between nonlinear mapped IQMs and DMOS for the distorted database by fast fading.

Model	CC	MAE	RMSE	OR
MSE	0.8119	7.4571	9.6026	0.0598
PSNR	0.8290	7.0825	9.1995	0.0554
SSIM	0.8360	7.0533	9.0872	0.0476
I	0.8241	7.4442	9.3181	0.0498
I_{MRF}	0.8589	6.9105	8.4250	0.0414

3.3. Experiments for the Distorted Images of Gaussian Blur

The calculated results of the five IQMs based on the distorted database of Gaussian blur are shown in Table 3. According to Table 3, similar results that the I_{MRF} outperforms the other four IQMs can be achieved for the distorted database of Gaussian blur.

Table 3. CC, MAE, RMSE and OR values between nonlinear mapped IQMs and DMOS for the distorted database by Gaussian blur.

Model	CC	MAE	RMSE	OR
MSE	0.7197	8.9574	10.9167	0.0493
PSNR	0.7291	8.7886	10.7631	0.0465
SSIM	0.6446	9.9057	12.0204	0.0523
I	0.6634	9.6715	11.7654	0.0514
I_{MRF}	0.7565	8.4070	10.7631	0.407

3.4. Discussions for all Databases and all IQMs

The scatter plots of the results for all 3 databases and all 5 IQMs are shown in Fig. 2. In Fig. 2, the black curve represents the nonlinear mapping of the IQMs' values base on white noise dataset. The purple circles, red stars and blue pluses are the IQMs' values based on the white noise, Gaussian blur and fast fading datasets, respectively.

In Fig. 2(a) and Fig. 2(b), the values of *MSE* and *PSNR* increase or decrease when the subjective scores of DMOS increase. However, the points appear a little divergent. That shows the *MSE* and *PNSR* care reflect the image quality variation basically.

In Fig. 2(c), the *SSIM* values decrease when the subjective scores of DMOS increase. The *SSIM* values distribute in the region near to the fitted black curve, especially for the distorted database of white noise. However, the *SSIM* values fail to follow the changing trends of the subjective score DMOS for the database of Gaussian blue and fast fading. That agrees with Yang's [11] results: *SSIM* failed to measure the blurred images with a lot of flat regions.

Among these five scatter plots, the varying of I_{MRF} shows more consistent to the changing of DMOS. That means the calculated results of I_{MRF} are more consistent with the subjective human evaluation. It indicates that it is effective and accurate to evaluate the image quality by introducing the MRF to the mutual information.

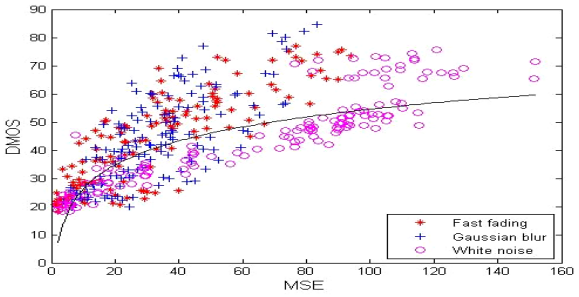
4. Experiments on Low Dose CT Images

In order to validate the performance of the IQMs on medical low dose CT images, the low dose CT phantom images are tested in this paper, further. The CT phantom images were created by a 64-row dual-source computed tomography system (Somatom Definition 2008G; SIEMENS). The helical scanning was under the following conditions: tube voltage of 120 kV, tube current of 420 mA, image slice thickness of 1 mm, and the field of view of 374 mm with a reconstruction function of B30f. The CT phantom Images with five X-ray dose levels were created under the X-ray exposure: 350, 280, 210, 140 and 70 mAs. They are corresponding to the 100, 80, 60, 40 and 20 % dose level. In Fig. 3, images in the same slice from different dose level datasets are shown. In the low dose phantom image, there are five circles varies from 30 mm to 60 mm in diameter. The contents inside each circle corresponds to the air, bone, calcified material, parenchyma and fat materials in a clockwise manner.

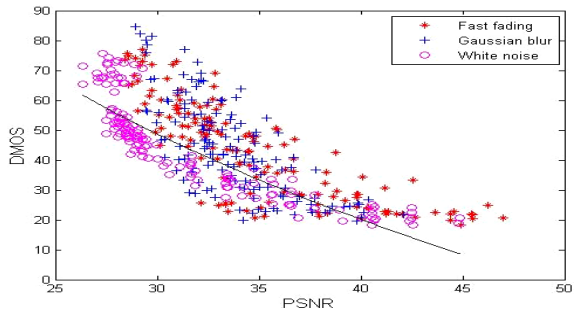
The *MSE*, *PSNR*, *SSIM*, *I* and I_{MRF} are applied to evaluate the low dose phantom CT images. The calculated results are shown in Table 4. In Table 4, the *PSNR* value increases from 60.0 to 70.5 when the X-ray dose level increases from 20 % to 80 %. The variation of *PSNR* value reflects the change of image quality in the isolated pixels aspect.

Table 4. Calculated results of IQMs based on low dose CT phantom images.

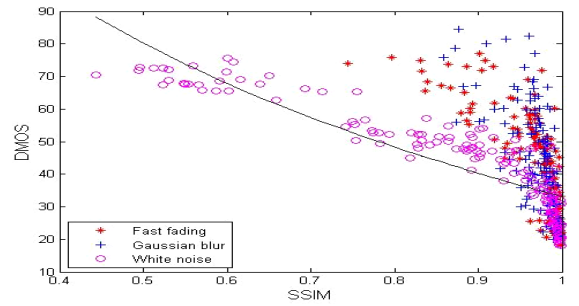
Model	Dose 20 %	Dose 40 %	Dose 60 %	Dose 80 %
MSE	1074.4	614.9	467.1	385.3
PSNR	66.0	68.4	69.6	70.5
SSIM	0.99963	0.99982	0.99987	0.9999
I	1.9889	2.1105	2.1793	2.2230
I_{MRF}	2.1502	2.3004	2.3687	2.5103



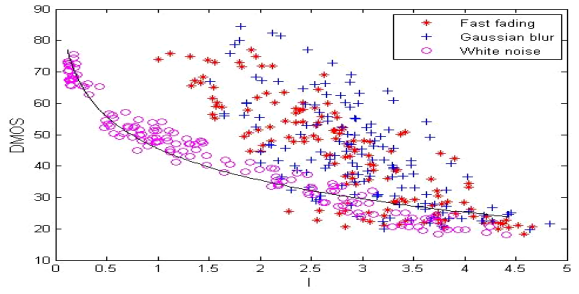
(a) *MSE* versus DMOS



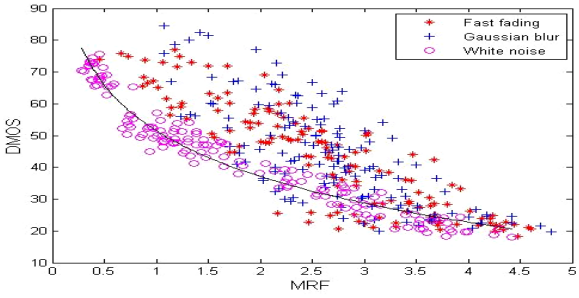
(b) *PSNR* versus DMOS



(c) *SSIM* versus DMOS



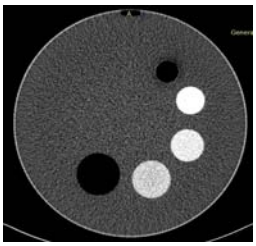
(d) *I* versus DMOS



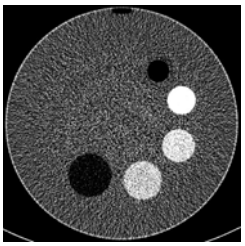
(3) I_{MRF} versus DMOS

Fig. 2. Scatter plots of the five IQMs for the different distorted databases.

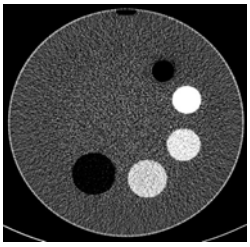
The mutual information *I* increase from 1.9889 to 2.2230. It reflects the changes of image quality in the information theory aspect. The value of I_{MRF} increases from 2.1502 to 2.5103 when the clique-vector model of MRF is introduced. The changing interval is enlarged from 0.2341 of *I* to 0.3601 of I_{MRF}. The value of *SSIM* remains almost unchanged when the X-ray dose level changes from 20 % to 80 %. In this sense, the I_{MRF} outperforms *I* and *SSIM* when evaluating the image quality of low dose CT dataset by introducing the local correlation.



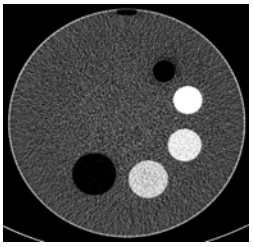
(a) Dose 100 %



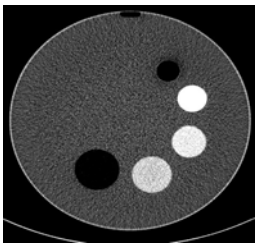
(b) Dose 20 %



(c) Dose 40 %



(d) Dose 60 %



(e) Dose 80 %

Fig. 3. The phantom CT images imaged in different dose levels.

5. Conclusions

In this paper, the LIVE databases and the phantom CT images created in different X-ray dose levels are applied to analyze and validate different IQMs: *MSE*, *PSNR*, *SSIM*, *I* and I_{MRF}. The experimental results show that *MSE*, *PSNR*, *SSIM* and *I* can reflect the variations of the common images. However, they did not consider the local correlations of images. Therefore, their values of CC, MAE, RMSE and OR between them and DMOS are poorer than that of I_{MRF}. In the experiments based on low dose phantom CT images, the changes of I_{MRF} can reflect the variation of image quality more accurately along with the changes of the X-ray dose level. It can evaluate the low dose CT images as an effective reference.

Acknowledgements

This work is supported by the National Nature Science Foundation of China (No. 61202267).

References

[1]. L. Wang, L. Ling, L. Xiang, et al., Enhancement method of multidirectional medical CT image based on wavelet laid, *Computer Engineering and Design,* Vol. 29, Issue 14, 2008, pp. 3693-3695.

[2]. M. G. M. Hunink, G. S. Gazelle, CT screening: a trade-off of risks, benefits, and costs, *Journal of Clinical Investigation*, Vol. 111, Issue 11, 2003, pp. 1612-1619.

[3]. Y. Duan, J. Ma, W. Chen, et al., Improved SSIM medical image quality assessment, *Computer Engineering and Applications*, Vol. 46, Issue 2, 2010, pp. 145-149.

[4]. L. Yang, H. Chen, L. Chen, Model of no reference image quality assessment based on steerable pyramid, *Computer Engineering and Design*, Vol. 34, Issue 8, 2013, pp. 2769-2773.

[5]. G. Jiang, D. Huang, et al., Overview on image quality assessment methods. *Journal of Electronics & Information Technology*, Vol. 32, Issue, 1, 2010, pp. 219-226.

[6]. H. R. Sheikh, A. C. Bovik, Image information and visual quality, *IEEE Transactions on Image Processing*, Vol. 15, Issue, 2, 2006, pp. 430-444.

[7]. X. Wei, J. Li, G. Chen, A perception based image quality assessment model, *Journal of Computer-Aided Design & Computer Graphics*, Vol. 12, Issue 1, 2007, pp. 1540-1545.

[8]. W. Zhou, A. C. Bovik, Why is image quality assessment so difficult?, *IEEE International Conference on Acoustics, Speech and Signal Processing*, Orlando, USA, 2002, pp. 3313-3316.

[9]. W. Zhou and A. C. Bovik, Mean Squared Error: Love it or leave it?, *IEEE Signal Processing Magazine*, Vol. 26, Issue 1, 2009, pp. 98-117.

[10]. W. Zhou, A. C. Bovik, H. R. Sheikh, et al., Image quality assessment: from error visibility to structural similarity, *IEEE Transactions on Image Processing*, Vol. 13, Issue 4, 2004, pp. 600-612.

[11]. C. Yang, G. Chen, S. Xie, Gradient information based image quality assessment, *Acta Electronica Sinia*, Vol. 35, Issue 7, 2007, pp. 1313-1317.

[12]. E. Volden, G. Giraudon, M. Berthod, Modelling image redundancy, in *Proceedings of the IEEE International Symposium Geoscience and Remote Sensing* (*IGARSS'95*), 1995, 3, pp. 2148-2150.

[13]. H. R. Sheikh, Z. Wang, L. Cormack and A. C. Bovik, LIVE Image Quality Assessment Database Release 2, http://live.ece.utexas.edu/research/quality.

[14]. H. R. Sheikh, M. F. Sabir, A. C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, *IEEE Transactions on Image Processing*, Vol. 15, Issue 11, 2006, pp. 3440-3451.

[15]. Video Quality Experts Group, Final report from the video quality experts group on the validation of objective models of video quality assessment, *VQEG,* Mar, 2000.

[16]. Video Quality Experts Group, Final report from the video quality experts group on the validation of objective models of video quality assessment, (2003-08) http://www.vqeg.org/

[17]. S. Ye, K. Su, C. Xiao, et al., Image quality assessment based on structural information extraction, *Acta Electronica Sinica*, Vol. 5, Issue 5, 2008, pp. 856-861.