The Influence of Different Saliency on Full-Reference Sonar Image Quality Evaluation

IOP Conf. Series: Materials Science and Engineering **569** (2019) 052093 doi:10.1088/1757-899X/569/5/052093

The Influence of Different Saliency on Full-Reference Sonar Image Quality Evaluation

Huiqing Zhang [1, 2], Shuo Li [1, 2, *], Weiling Chen[3], and Yutao Liu[4]

[1]Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

[2]Engineering Research Center of Digital Community, Ministry of Education, Beijing 100124, China

[3]College of Physics and Information Engineering, Fuzhou University, Fuzhou, 350108, China

[4]The Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China

[*]Corresponding author's e-mail: shuoli1025@163.com

Abstract. With the advent of sonar technology, our understanding of the ocean has become more comprehensive, especially for deep sea biology and geology. However, the sonar image is easily degraded during the underwater acoustic channel acquisition process, which affects the later research work. To this end, this paper compares and analyzes multiple saliency models and combines them with PSNR, SSIM and GSIM to explore an effective sonar image quality evaluation method. Finally, an experimental analysis on the newly established sonar image quality database shows that the difference of the significance model in predicting human attention has a performance gain effect on the image quality evaluation method when fused with the saliency model.

1. Introduction

It is well known that ordinary optical camera systems are completely unusable in the dark and in mixed water. With the advent of sonar technology, underwater rescue, mineral resource exploration and protection of aquatic animals have been carried out smoothly. As an intermediate medium, the sonar image is able to present underwater images to us. Due to the uncertainty of the underwater acoustic channel propagation environment, the sonar image is easily interfered by several typical distortions in the acquisition process, resulting in image degradation, which has a very negative impact on subsequent processing such as image feature extraction and automatic target recognition. Therefore, it is necessary to explore effective sonar image quality evaluation (IQE) methods.

The earliest proposed IQE metrics are the mean square error (MSE) and the peak signal to noise ratio (PSNR), which measure the degree of difference between the distorted image and the original image. However, as image complexity continues to increase, their relevance to human judgment of images is also decreasing. Later, people began to explore the IQE method from different angles. For example, the structural similarity index (SSIM) [1] predicts the image quality by comparing the brightness, contrast, and structural similarities between the lossless image and the lossy image. The gradient similarity index (GSIM) [2]-[3] is a comparison of the gradient difference between the lossless image and the lossy image to predict the image quality.

The research progress of image quality evaluation shows that adding visual attention to objective evaluation has potential added value. For example, in [4], a special saliency model was developed to optimize two IQE methods, namely visual information fidelity (VIF) and MSSIM, thus significantly improving their performance. In both literature [5] and [6], the saliency is introduced in IQE design to improve its correlation with subjective quality evaluation. In addition, saliency also plays a significant role in improving the contrast and sharpness of images. Guided by saliency, the contrast enhancement algorithms QMC [7] and RIQMC [8] are superior to the recently developed enhancement techniques. The ARISM model was proposed in [9] based on the idea of saliency, which effectively solved the problem of sharpness degradation. Therefore, it is usually effective to use a specific saliency model to specifically optimize the target IQE.

The sonar image has different characteristics from the natural scene image (NSIs) [10]-[11]. NSIs are mostly color images taken by ordinary cameras, which have large pixel values and complex texture content. The sonar image is a grayscale image formed by special equipment, which has small changes in pixel value, less detail, and concentrated information distribution. Therefore, the methods used to evaluate the quality or distortion of natural scene images are not necessarily applicable to sonar images. As far as we know, NSIs have been studied by most researchers, but the research on sonar IQE is very limited [12]-[13]. In order to solve the problem of sonar IQE, we analyze eight kinds of saliency models in this paper, and combine with PSNR, SSIM and GSIM to explore an effective sonar IQE method.

The rest of this paper is organized as follows. In section 2, we first analyze the eight saliency models and the three IQE metrics and then discuss how to fuse them. Section 3 gives the experimental results on the sonar image quality database (SIQD) [14] and discusses which fusion method is better for the prediction of sonar image quality. Finally, we conclude in the section 4.

2. Methodology

In order to explore an effective evaluation method for sonar image quality, we fused eight saliency models and three IQE methods, respectively, and tested the performance of the fused model and the original model on the newly established sonar image quality database.

The eight state-of-the-art models of saliency model, namely AIM [15], SUN [16], SR [17], PFT [18], SeR [19], SWD [20], HFT [21], and CovSal [22], are used in this paper. AIM and SUN are implemented in slightly different ways. They use shannon's visual feature self-information metric to calculate significance. SR and PFT are simple and effective models for studying the Fourier transform phase spectrum. The SeR method first computes the local regression kernel of the image, which is used to measure the similarity between the pixel and its surrounding environment. The 'self-similarity' measure is then used to calculate visual significance. SWD method measures the significance of image by calculating the difference between image blocks and the spatial distance between image blocks and center offset, and then negatively weighting the difference according to the corresponding spatial distance. HFT is to reconstruct the two-dimensional signal by using the original phase and amplitude spectrum, and filter the scale selected according to the minimum entropy of significance graph to get the significance graph. CovSal solves the influence of different feature dimensions on the overall visual significance, and proposes to use the covariance matrix of simple image features as the meta feature of significance estimation.

As shown in Figure 1, the left-most image is an original image in the SIQD database, and the remaining eight images are the saliency maps generated by AIM, CovSal, HFT, PFT, SeR, SR, SUN, SWD. Next, we use the common similarity measurement method to detect the saliency difference between the lossless image and the lossy image, and define the saliency comparison as:

$$S(x,y)=\frac{2S_x\times S_y+C_1}{S_x^{\,2}+S_y^{\,2}+C_1} \tag{1}$$

where x and y are lossless and lossy sonar images, S_x and S_y are the saliency maps of the lossless and lossy sonar images, and C_1 is used here to prevent the denominator of Eqs. 1 from going to zero.

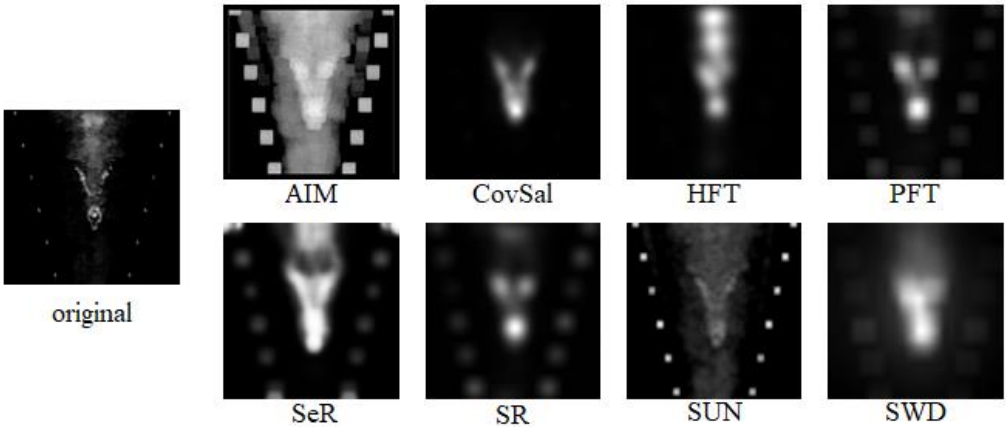

Figure 1. Illustration of saliency maps generated by 8 state-of-the-art saliency models for one of the original images in SIQD database.

We use three of the most classical IQE algorithms metrics, namely PSNR, SSIM, and GSIM, Specific as follows. The PSNR of image is based on the mean square error (MSE) of lossy image and lossless image on the basis of each pixel, which is an improvement of MSE. MSE is defined as:

$$MSE(x,y)=(x-y)^2. \tag{2}$$

Next, it is fused with the saliency maps to explore a better sonar IQE method:

$$F_{MSE}=\frac{\sum S(x,y)\times MSE(x,y)}{\sum S(x,y)} \tag{3}$$

where $\sum$ is the summation symbol. Finally, PSNR is defined as:

$$PSNR=10log_{10}\left[\frac{\left(2^n-1\right)^2}{F_{MSE}}\right] \tag{4}$$

where n is the upper limit of the pixel value in the image, and we take $n=255$. In order to prevent the case of $F_{MSE}=0$, we set an upper limit on PSNR: $\overline{PSNR}=min\left(1000,PSNR\right)$.

The SSIM metric is used to evaluate the image quality based on the comparison of local brightness, contrast and structure of lossless and lossy images. Specifically, the similarity of brightness, contrast, and structure between two image blocks $\hat{x}$ and $\hat{y}$ extracted from lossless and lossy images is defined as:

$$Br(\hat{x},\hat{y})=\frac{2\mu_{\hat{x}}\times\mu_{\hat{y}}+C_2}{\mu_{\hat{x}}^2+\mu_{\hat{y}}^2+C_2}\qquad Co(\hat{x},\hat{y})=\frac{2\sigma_{\hat{x}}\times\sigma_{\hat{y}}+C_3}{\sigma_{\hat{x}}^2+\sigma_{\hat{y}}^2+C_3}\qquad St(\hat{x},\hat{y})=\frac{\mu_{\hat{x}}+C_4}{\mu_{\hat{x}}\mu_{\hat{y}}+C_4} \tag{5}$$

where μ, σ are mean and standard deviation, and $\mu_{\hat{x}}\mu_{\hat{y}}$ is the cross-correlation evaluations. The function of C_2 to C_4 are the same as C_1, and $C_4=C_3/2$. Then, SSIM is obtained by merging the above three kinds of information:

$$SSIM(\hat{x},\hat{y})=\frac{(2\mu_{\hat{x}}\times\mu_y+C_2)(2\mu_{\hat{x}}\mu_{\hat{y}}+C_3)}{(\mu_{\hat{x}}^2+\mu_{\hat{y}}^2+C_2)(\sigma_{\hat{x}}^2+\sigma_{\hat{y}}^2+C_3)}. \tag{6}$$

Finally, SSIM and saliency maps are fused to evaluate the sonar image quality:

$$\overline{SSIM} = \frac{\sum S(x,y) \times SSIM(\hat{x},\hat{y})}{\sum S(x,y)} \quad . \tag{7}$$

The GSIM metric is used to predict the quality of sonar image by comparing the gradient difference between lossless image and lossy sonar image. In order to respect the integrity of the original algorithm, we still use Prewitt gradient operator here:

$$g_h = \frac{1}{3} \times \begin{pmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{pmatrix} \quad g_v = \frac{1}{3} \times \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{pmatrix} \tag{8}$$

where g_h and g_v are separately indicating Prewitt convolution masks in the horizontal and vertical directions. The typical Prewitt gradient operator is applied to gradient extraction:

$$Gm_x = \sqrt{(g_h * x)^2 + (g_v * x)^2} \qquad Gm_y = \sqrt{(g_h * y)^2 + (g_v * y)^2} \tag{9}$$

where Gm_x and Gm_y represents the gradient magnitude map of lossless and lossy images, and ' * 'represents the convolution operator. Similarly, we use the similarity measurement method to detect the difference between the two on a large scale:

$$GSIM(x,y) = \frac{2Gm_x \times Gm_y + C_5}{Gm_x^2 + Gm_y^2 + C_5} \tag{10}$$

where, the function of C_5 is the same as C_1 to C_4. Finally, GSIM and saliency maps are fused to evaluate the sonar image quality:

$$\overline{GSIM} = \frac{\sum S(x,y) \times GSIM(x,y)}{\sum S(x,y)} \quad . \tag{11}$$

3. Experimental Results and Analysis

In this section, we will test the performance of each fusion method on the newly established SIQD database. The SIQD database includes 40 lossless sonar images captured by acoustic lens sonar or side-scan sonar (as display in Figure 2, which includes swimmers, shipwrecks, seabed and underwater creatures) and 800 lossy sonar images generated from lossless sonar images with 4 distortion types at four to six distortion levels. For more information, readers can refer to [14].

Due to the nonlinearity in the process of subjective scoring, we first used the five-parameter logistic regression function to carry out the nonlinear mapping for the objective quality prediction score [23]-[24]:

$$f(s) = \alpha_1 \left(0.5 - \frac{1}{1+e^{\alpha_2 (s - \alpha_3)}} \right) + \alpha_4 s + \alpha_5 \tag{12}$$

where s and $f(s)$ are the objective quality prediction score and its regression version, and α_1 to α_5 are five free parameters.

Then, we use five performance indicators suggested by the video quality experts group, to evaluate the performance of the fused model. Firstly, the Pearson Linear Correlation Coefficient (PLCC) is used to predict the accuracy between subjective quality score and $f(s)$. Second, the Spearman Rank Order Correlation Coefficient (SROCC) and the Kendall's Rank Order Correlation Coefficient (KROCC) are used to predict the monotonicity between subjective quality score and objective quality prediction score. Prediction consistency can be estimated using the Root Mean-Square (RMS) error and the Mean Absolute Error (MAE) between subjective quality score and $f(s)$. The larger the value

of SROCC, KROCC and PLCC, and the smaller the value of RMS and MAE, the better the performance of IQE model [25]-[26].

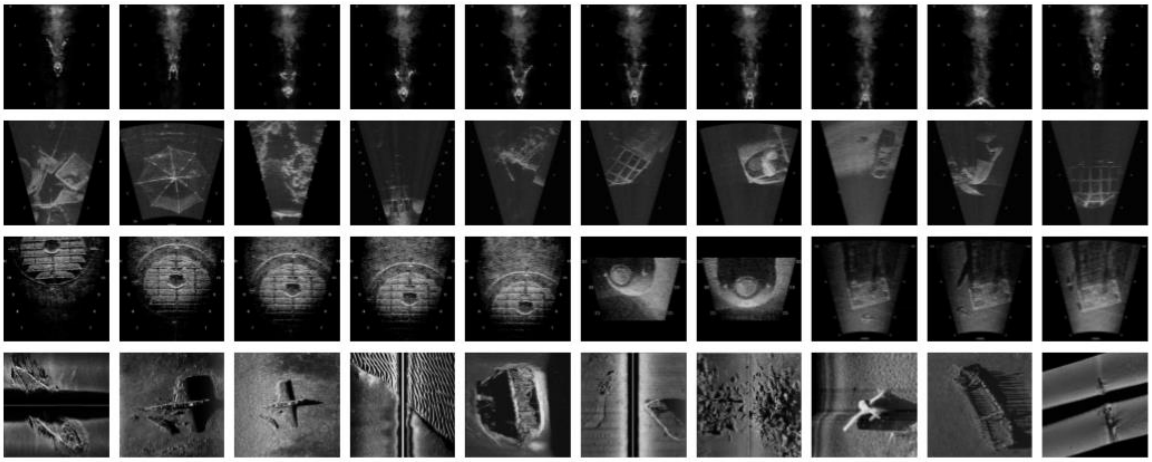

Figure 2. Lossless sonar images.

Table 1: Test results of SROCC, KROCC, PLCC, RMS, and MAE metrics on SIQD database

Criteria	Metric	original	AIM	covSal	HFT	PFT	SeR	SR	SUN	SWD
SROCC	PSNR	**0.2754**	**0.2759**	**0.2275**	0.2754	**0.2755**	0.2754	**0.2755**	**0.2767**	**0.2755**
	SSIM	**0.5545**	0.4177	0.5445	0.1708	0.1708	0.5443	0.5444	0.0993	0.1708
	GSIM	**0.7127**	0.6550	0.7127	**0.7136**	**0.7137**	**0.7142**	**0.7138**	0.6180	**0.7133**
KROCC	PSNR	**0.2914**	**0.3094**	0.2914	0.2912	0.2912	0.2909	0.2912	**0.3065**	0.2913
	SSIM	**0.5674**	**0.5688**	0.5674	0.5672	0.5672	0.5672	0.5673	**0.5691**	0.5672
	GSIM	**0.7017**	0.6444	0.7017	**0.7027**	**0.7027**	**0.7033**	**0.7027**	0.6042	**0.7022**
PLCC	PSNR	**0.2078**	**0.2215**	0.2078	0.2077	0.2077	0.2074	0.2077	**0.2170**	0.2077
	SSIM	**0.4033**	**0.4035**	0.4033	0.4026	0.4026	0.4031	0.4033	**0.4038**	0.4026
	GSIM	**0.5063**	0.4634	0.5063	**0.5072**	**0.5073**	**0.5079**	**0.5073**	0.4315	**0.5068**
RMS	PSNR	**11.148**	**11.146**	**11.147**	**11.147**	**11.147**	**11.147**	**11.147**	**11.144**	**11.147**
	SSIM	**9.3926**	10.308	9.3927	11.381	11.380	**9.3923**	9.3928	11.540	11.380
	GSIM	**7.6736**	8.4797	**7.6736**	**7.6654**	**7.6627**	**7.6556**	**7.6622**	8.7875	**7.6671**
MAS	PSNR	**13.444**	**13.441**	**13.443**	**13.443**	**13.443**	**13.443**	**13.443**	**13.438**	**13.443**
	SSIM	**11.730**	12.706	**11.729**	13.779	13.778	11.731	11.730	13.915	13.779
	GSIM	**9.8094**	10.567	9.8094	**9.7972**	**9.7959**	**9.7882**	**9.7940**	10.994	**9.8016**

The performance comparison of the fused model is shown in Table 1. The boldface type represents the performance score of the original model, and the red font type represents the model whose performance increases after the fusion with the saliency model. As listed in Table 1, the performance of PSNR and GSIM is improved after the fusion with saliency model, while the performance gain of SSIM is not obvious. Specifically, for PSNR, the saliency model AIM and SUN can improve the performance of PSNR on five indicators in terms of predicting sonar image quality. For GSIM, the performance of GSIM in predicting sonar image quality on five indicators can be improved when fused with saliency models (namely HFT, PFT, SeR, SR, and SWD). However, for SSIM, the addition of the significance model did not increase the performance of SSIM, which may be caused by the high

similarity between the brightness features extracted from SSIM and the significance model. In a word, saliency model is of guiding significance for IQE.

4. Conclusion

This paper is a stage study on sonar IQE, which is devoted to find a better method for sonar IQE. We compares and analyses eight saliency models and combines them with PSNR, SSIM and GSIM to explore an effective sonar IQE method. Experimental results on the SIQD database show that most of the saliency models have positive effects on IQE, and in the future work, we will consider fusing GSIM with the significance model (such as HFT, PFT, SeR, SR, or SWD) to explore the IQE method with better performance.

Acknowledgment

This work is supported by the Major Science and Technology Program for Water Pollution Control and Treatment of China (2018ZX07111005)

References

[1] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., (2004) Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process., 13: 600–612.
[2] Gu, K., Wang, S., Zhai, G., Lin, W., Yang, X., and Zhang, W., (2016) Analysis of distortion distribution for pooling in image quality prediction IEEE Trans. broadcasting, 62:446-456.
[3] Gu, K., Zhai, G., Yang, X., and Zhang, W., (2014) An efficient color image quality metric with local-tuned-global model. In: IEEE Int. Conf. Image Process. Paris, pp. 506-510.
[4] Ma, Q. and Zhang, L., (2008) Image quality assessment with visual attention. In: 15th Int. Conf. Pattern Recognit., Tampa, pp. 1–4.
[5] Barland, R. and Saadane, A., (2006) Blind quality metric using a perceptual importance map for JPEG-20000 compressed images. In: 13th IEEE Int. Conf. Image Process. Atlanta, pp. 2941–2944.
[6] Rao, D. V., Sudhakar, N., Babu, I. R., and Reddy, L. P., (2007) Image quality assessment complemented with visual regions of interest. In: Int. Conf. Comput, Theory Appl., Kolkata, pp. 681–687.
[7] Gu, K., Zhai, G., Yang, X., Zhang, W., and Chen C., W., (2015) Automatic contrast enhancement technology with saliency preservation. IEEE trans. circuits and systems for video technology. 25: 1480-1494.
[8] Gu, K., Zhai, G., Lin, W., and Liu, M., (2016) The analysis of image contrast: from quality assessment to automatic enhancement. IEEE trans. cybernetics. 46: 284-297.
[9] Gu, K., Zhai, G., Lin, W., Yang, X., and Zhang, W., (2015) No-reference image sharpness assessment in autoregressive parameter space. IEEE Trans. Image Process., 24: 3218-3231.
[10] Bovik, A. C. and Sheikh, H. R., (2006) Image information and visual quality. IEEE Trans. Image Process., 15: 430-444.
[11] Gu, K., Li, L., Lu, H., Min, X., and Lin,W., (2017) A fast reliable image quality predictor by fusing micro- and macro-structures. IEEE Trans. Industrial Electronics, 64: 3903-3912.
[12] Chen, W., Gu, K., Min, X., Yuan, F., Cheng, E., and Zhang, W., (2018) Partial-Reference Sonar Image Quality Assessment for Underwater Transmission. IEEE Trans. aerospace and electronic systems. 54: 2776-2787.
[13] Chen, W., Gu, K., Lin, W., Xia, Z., Callet, P., L., and Cheng,E.,(2019) Reference-free quality assessment of sonar images via contour degradation measurement," IEEE Trans. Image Process., to be published.
[14] Chen, W., Fang, Y., Gu, K., Yuan, F., and Cheng, E., (2019) Subjective quality assessment of sonar images based on underwater acoustic transmission. Signal, Image and Video Process., to be published, https://arxiv.org/submit/2648465/view.
[15] Tsotsos, J. K. and Bruce, N. D. B., (2009) Saliency, attention, and visual search: An information

theoretic approach. Journal of Vision, 9: 5-5.
[16] Zhang, L., Tong, M. H., Marks, T. K., Shan, H., and Cottrell G. W., (2008) SUN: A Bayesian framework for saliency using natural statistics. Journal of Vision, 8: 32-32.
[17] Hou, X. and Zhang, L., (2007) Saliency detection: A spectral residual approach. In: 20th IEEE Conf. Comput. Vis. Pattern Recognit., Minneapolis, pp. 1–8.
[18] Guo, C. and Zhang, L., (2008) Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. In: 21th IEEE Conf. Comput. Vis. Pattern Recognit., Anchorage, pp. 1–8.
[19] Seo, H. J. and Milanfar, P., (2009) Nonparametric bottom-up saliency detection by self-resemblance. In: 22th IEEE Conf. Comput. Vis. Pattern Recognit., Miami, pp. 1–8.
[20] Duan, L., Wu, C., Miao, J., Qing, L., and Fu, Y., (2011) Visual saliency detection by spatially weighted dissimilarity. In: 24th IEEE Conf. Comput. Vis. Pattern Recognit., Colorado Springs, pp. 1–8.
[21] Li, J., Levine, M. D., An, X., Xu, X., and He, H., (2013) Visual saliency based on scale-space analysis in the frequency domain. IEEE transactions on pattern analysis and machine intelligence, 35: 996-1010.
[22] Erdem, E. and Erdem, A., (2013) Visual saliency estimation by nonlinearly integrating features using region covariances. Journal of Vision, 13: 1-20.
[23] Gu, K., Tao, D., Qiao, J., F., and Lin, W., (2018) Learning a no-reference quality assessment model of enhanced images with big data. IEEE trans. neural networks and learning systems. 29: 1301-1313.
[24] Sheikh, H., R., Bovik, A., C., and Veciana, G., de., (2005) An information fidelity creation for image quality assessment using natural scene statistics. IEEE Trans. Image Process., 14: 2117–2128.
[25] Gu, K., Zhai, G., Yang, X., and Zhang, W., (2015) Using free energy principle for blind image quality assessment. IEEE trans. Multimedia, 17: 50-63.
[26] Gu, K., Qiao, J., and Li, X., (2019) Highly efficient picture-based prediction of PM2.5 concentration. IEEE trans. Industrial Electronics, 66: 3176-3184.