SUBJECTIVE AND OBJECTIVE QUALITY EVALUATION OF SONAR IMAGES FOR UNDERWATER ACOUSTIC TRANSMISSION

Weiling Chen†‡, Fei Yuan†‡, En Cheng†‡, and Weisi Lin§

†School of Information Science and Technology, Xiamen University, Xiamen, China
‡Key Laboratory of Underwater Acoustic Communication and Marine Information Technology
§School of Computer Science and Engineering, Nanyang Technological University, Singapore
E-mail: chengen@xmu.edu.cn

ABSTRACT

One of the most critical missions of sonar is to capture deep-sea pictures to depict sea floor and various objects, and provide an immense understanding of biology and geology in deep sea. Due to the poor condition of underwater acoustic channel, the captured sonar images very possibly suffer from several typical types of distortions before finally reaching to users. Unfortunately, very limited efforts have been devoted to collecting meaningful sonar image databases and benchmark reliable objective quality predictors. In this paper, we first generate a sonar image quality database (SIQD), including 840 images. All distorted images were collected without artificially introducing any distortions beyond those occurring during compression and transmission. The subjective quality assessment was conducted for gathering mean opinion score (MOS) to represent the image quality and existence of target (EOT) which describes whether the image is useful. Based on the built SIQD database, state-of-the-art general image quality metrics were found to poorly correlate with "ground-truth" MOS. As a consequence, this paper further develops a novel full-reference local entropy backed sonar image quality predictor (LESQP). The experimental results demonstrate the superiority of our LESQP metric over the available quality measures.

Index Terms— Sonar image, underwater acoustic transmission, quality assessment, database, human visual system

1. INTRODUCTION

One of the most significant sonar missions is to photograph seafloor for later processing in laboratory or industry. Owing to the poor condition of the underwater acoustic channel, images however very possibly suffer from some typical types of distortions before finally reaching to human [1]. Image quality assessment (IQA) method plays an important part in maintaining a satisfactory quality of received images. Though it is time-consuming to conduct subjective assessments for distorted sonar images, the data obtained from human viewers provides the ground truth, which can be used for comparing objective IQA metrics. There exist several well-established image quality databases for typical pictures [2]-[7]. These databases are inclined to collect the mean opinion scores (MOSs) or differential mean opinion scores (DMOSs) so as to explore the variations occurred in natural images due to distortion introduction.

To the best of our knowledge, we are only aware of the initial effort made in establishing turbid underwater image quality dataset [8], and the work uses quality index called Structural Degradation Index (SDI) defined from the Structure SIMilarity Index (SSIM) [9] rather than the subjective scores. Since the SSIM is designed for natural images which may not agree with the subjective quality of sonar image, a sonar image quality database (SIQD) is constructed in this paper, which consists of 40 reference images, 800 test images contaminated via compression or transmission and their subjective qualities. We gather the MOS for each distorted image as its quality index. Besides, considering the utility of sonar image [10], EOT is also collected for each image. This index shows the subjective opinion about whether the target is included in the sonar image using labels: "with target" and "without target". In the proposed database, targets include underwater creatures, seabed, shipwrecks and so on. Since a sonar image is useless without an important object even though it has high quality, the EOT is added in the proposed database to describe the usefulness of a sonar image from the aspects of authentic applications. More detail about EOT will be discussed in the following sections.

Apart from subjective quality assessment for sonar images, a novel full-reference (FR) local entropy backed sonar image quality predictor (LESQP) is developed as well. Entropy measures the disorder of a system. When distortion is added to the image transmission system, the system tends to go from a state of order (low entropy) to a state of disorder (high entropy) [11]. This change will be reflected in image entropy. On this basis, the local entropy is used to describe the disorder of an image. Experiments demonstrate the superiority of our IQA method over existing IQA models for synthetic aperture sonar images [12]-[14].

The rest of this paper is organized as follows. Section 2 introduces the subjective assessment methodology and compositions of the SIQD database. Section 3 describes the proposed LESQP metric. In Section 4, a comparison of the LESQP and state-of-the-art FR-IQA algorithms is conducted for comparison. Section 5 concludes this paper.

2. SUBJECTIVE QUALITY ASSESSMENT

2.1. Reference and source images

For the purpose of this study, 40 reference images captured by five kinds of sonar have been chosen as reference images. It is supposed that images are displayed at sonar monitor. Because of this, these grayscale 8-bit images are of the same fixed resolution (320×320 pixels) which are suitable for the display equipment for sonar. They are of different content, some of them are quite textural whilst others contain large homogeneous regions. The objects in these images include swimmer, shipwrecks, underwater creatures, seabed, etc. Most of these objects are common underwater, although there are also some images containing human-made objects like an umbrella. Fig. 1 shows the sample reference images in the constructed SIQD database.

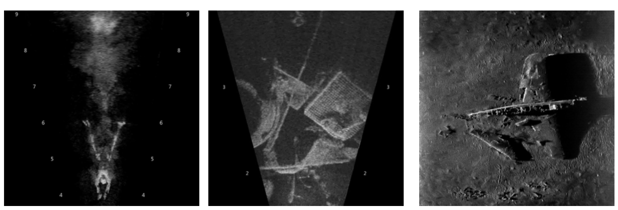


Fig. 1. Sample reference images in the SIQD database.

Although there have been great improvements in compression and transmission techniques, each stage of communication may introduce perceivable distortions. In this paper, the source compression before transmission and packet loss via transmission is considered. We choose SPIHT [15] and ComGBR [16] to make compressed images, and simulate packet loss by making man-made bit error on coding streams of two abovementioned compression coding methods. The bit error ratios (BERs) are selected according to the recent achievements about underwater acoustic communication [17]-[18]. For each distortion, four or five levels are considered, resulting in 20 distorted images created from a single pristine image.

2.2. Subjective test methodology

Since the number of images in this SIQD database is prohibitively high for a double stimulus setup, the single stimulus with multiple repetition (SSMR) as specified in [19] is employed in this database. All the images are divided into 20 groups, each has 42 images including reference images and distorted images. The same image is not presented twice in succession with the same level of impairment. In order to get a stable result, each group is tested by the same subjective viewer twice in different order, i.e. each subjective viewer rates the quality of same images in two different sessions, the order of images in these two sessions are different. The scores assigned to the images are obtained by taking the mean of the data issued from these two sessions. Each session includes two presentations: first one includes five images (different from the 42 images) to stabilize the viewers' opinion, the second one includes the aforementioned 42 images waiting for subjective scores and 5 repeated images for rejection.

In order to make the viewers be more certain about the quality rating, a 5-category discrete scale is obtained. In addition, viewers are asked to label the image using "with target" when the target is included in the image and is clear enough to be recognized, or "without target" when the target is not included in the image or has been destroyed too severely to be recognized. The viewers are asked to take the subjective test according to their habit for viewing screen. Each group of image is tested by at least 25 viewers. All of them are occupied in underwater acoustic communication-related works.

2.3. Image Indices

In order to measure the image quality, there are two indices included in the proposed database. First, the MOS for each image is obtained as a quality index. This index shows the visual feel of subjective viewers for each test image. The evaluation criteria are based on [19], the difference is that when measuring the subjective quality of sonar images, viewers take more account of the information that can be extracted from the test image. Then considering the utility of sonar image, EOT is also collected for each image. This index reflects whether the target is able to be recognized in the sonar-captured image by two labels: "with target" and "without target". When one test image is labeled with "with target", it means that there is at least one target included in this test image. On the contrary, the test image will be labeled with "without target" if there is not a target in the image or the target has been afflicted too severely to be recognized. So there are two kinds of images may be labeled with "without target", high-quality images without containing any important object and images with poor qualities. We choose the label that has the largest probability to be chosen for a test image. The EOT is defined as Eq. 1

$$EOT = \underset{Q}{\arg\max}\, P_Q(i). \tag{1}$$

In Eq. 1, i denotes test image. $P_Q(\cdot)$ is the fraction of viewers who label the image i with label Q, it approximates to the probability when image i is labeled Q, $Q \in \{\textit{with target}, \textit{without target}\}$.

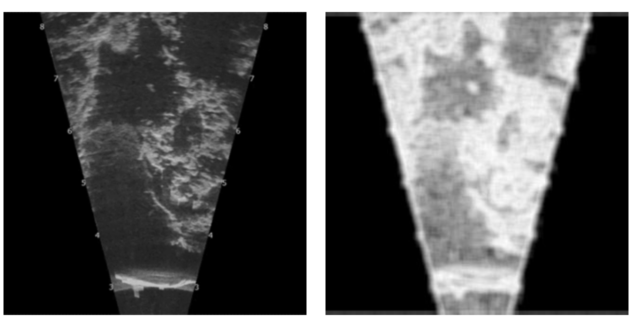
3. OBJECTIVE QUALITY ASSESSMENT

In most sonar missions, with the absence of prior knowledge about the target, the receiving echoes can be modelled as the output of a stochastic source. In this context, entropy (more specifically, Shannon entropy) is the expected value (average) of the information contained in each echo [20], and can be reflected by the entropy of obtained sonar image. When distortion is added to the image transmission system, the MOS of the distorted image will be lower and image entropy will be different since the information contained in this image has been changed by distortion. Then the change of image entropy can measure the distortion. In this paper, LESQP metric is proposed to predict the MOS value using local entropy of sonar image.

At typical viewing distances, only the object in a local area within an ”uncrowded window” can be recognized clearly [21], so the local entropy is employed in this paper. First we define the local entropy of the central position (x, y) in $2M + 1 \times 2M + 1$ image block as Eq. 2:

$$H_f(x, y) = -\sum_{i=0}^{255} P_i log P_i. \tag{2}$$

where P_i denotes gray-level distribution in this image block. In this paper, the local entropy is calculated within a local $2k_1 + 1 \times 2k_1 + 1$ window. Then the entropy map as Fig. 4 shows can be calculated.



(a) (b)

Fig. 2. Entropy map of the sonar image: (a) Original image; (b) Entropy map (brighter pixels indicates higher entropy).

Since the visual attention is sensitive to image edges, key locations are marked by a feature mask which is simply generated by an edge operator with a dilation operation. Suppose M_r and M_d are the result of edge detection perform on reference image f_r and distorted image f_d respectively. H_{f_r} and H_{f_d} are the entropy map of reference image f_r and distorted image f_d respectively. Then the masked entropy map can be defined as Eq. 3:

$$\begin{aligned} \hat{H}_{f_r}(x, y) &= H_{f_r}(x, y) \cdot M_r(x, y) \\ \hat{H}_{f_d}(x, y) &= H_{f_d}(x, y) \cdot M_d(x, y) \end{aligned} \tag{3}$$

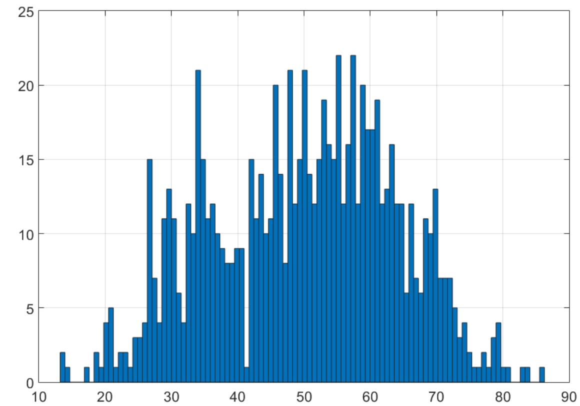


Fig. 3. Histogram of MOSs for images in the SIQD database.

The similarity between two entropy maps at corresponding location (x, y) is defined as Eq.4:

$$s(x, y) = \frac{2\hat{H}_{f_r}(x, y) \cdot \hat{H}_{f_d}(x, y) + c}{\hat{H}^2_{f_r}(x, y) + \hat{H}^2_{f_d}(x, y) + c(x, y)} \tag{4}$$

where c is a constant to avoid instability when $\hat{H}^2_{f_r}(x, y) + \hat{H}^2_{f_r}(x, y)$ is close to zero, and it is modified considering visual masking proposed in [22], the c is defined as Eq. 5, where K is chosen as 50 and δ is a small positive constant:

$$c(x, y) = K * min(H_{f_r}(x, y), H_{f_d}(x, y)) + \delta. \tag{5}$$

When comes to pooling method, the image activity measurement (IAM_0) [23] at each image location is calculated in a $k_2 \times k_2$ window to form a normalized effectivity map I. This naturally leads to Eq. 6 to defined the $LESQP$ as:

$$LESQP = \sum\sum s(x, y) \cdot I(x, y). \tag{6}$$

4. EXPERIMENTAL RESULTS

4.1. Data processing and indices obtaining

We employ the quartiles of the subjective scores for each image, normalized cross correlation (NCC) and the Euclidean distance (EUD) between every two subjective ratings to analyze the subject agreement. The results show that most images in this database have shown the agreement among viewers. Then the repeated image rule and kurtosis value [7] for rejection of subjective viewers are employed. Fig. 3 shows the histogram of MOSs for images in the SIQD database after rejection. It shows that MOSs span most of the quality range. The percentage of images labeled by ”with target” and ”without target” is 56% and 44% respectively. There are just as many images with target as images without target.

Table 1. Performance comparison for IQA algorithms on the SIQD database

Criteria	FSIM	VSNR	MAD	ADD-SSIM	GSM	SSIM	GMSD	VSI	PSIM	CPCQI	LESQP
SROCC	0.686	0.433	0.701	0.709	0.615	0.627	0.707	0.720	0.727	0.549	**0.785**
CC	0.707	0.476	0.726	0.731	0.633	0.650	0.714	0.736	0.738	0.567	**0.796**
RMSE	9.631	11.980	9.369	9.298	10.541	10.356	9.532	9.219	9.433	11.517	**8.474**
KROCC	0.490	0.299	0.509	0.509	0.429	0.417	0.503	0.523	0.528	0.377	**0.593**
MAE	7.539	9.772	7.307	7.331	8.316	8.554	7.621	7.148	7.492	9.241	**6.427**

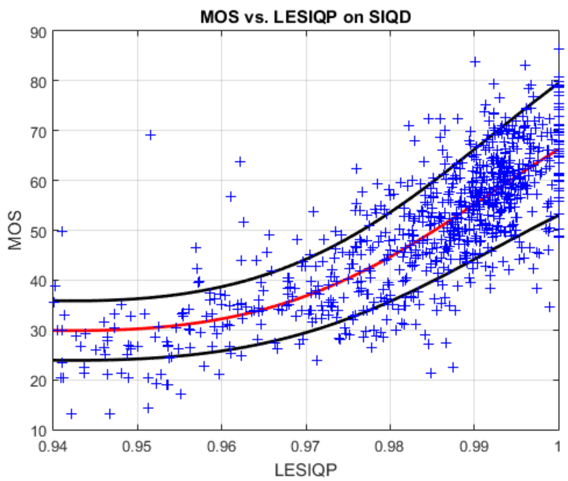


Fig. 4. Scatter plots of MOS versus predicted quality.

4.2. Comparative analysis for quality metrics

A five parameter logistic mapping is employed to remove nonlinearity because of the subjective rating process and to finish the comparison of the algorithms in a common analysis space. Fig. 4 shows the scatter plot of MOS versus predicted quality using LESQP. The middle curve in red is the five-parameter logistic mapping curve between LESQP and MOS, while the upper and lower curve in black denoted 20% the upper bound and lower bound respectively. As can be seen from Fig. 4, most data points are between the bounding curves. That means if 20% measurement error is allowed, most images will be correctly predicted from the LESQP.

We have evaluated the performance of 11 state-of-the-art FR-IQA algorithms [3], [9], [22], [24]-[30] and the proposed LESQP on the SIQD database. Five commonly used performance metrics are employed to evaluate the performance of the IQA algorithms. They are Spearman rank order correlation coefficient (SROCC), Pearson linear correlation coefficient (CC), root mean square error (RMSE), Kendall rank-order correlation coefficient (KROCC) and mean absolute prediction error (MAE). Table 1 lists the comparison of the performances of different IQA algorithms.

As can be seen from Table 1, the proposed LESQP performances better than other algorithms in the SIQD database. The performance is as much as about 11% better than the most of the competitors. Apart from proposed LESQP, VSI which includes visual saliency lays over the most of the others in performance. While LESQP is based on entropy and shows the best performance among the proposed IQA methods, it can be concluded that information performs better in the representative of sonar images' quality. Besides, MAD, ADD_SSIM, GMSD and PSIM also show better performance than the rest of the methods on the proposed SIQD database. Among these IQA methods, PSIM evaluates the quality based on gradient magnitude and color information. Since all sonar images are in grayscale, so the PSIM performs effectively mainly based on the gradient magnitude on the proposed SIQD database. And GMSD computes the local quality map by comparing the gradient magnitude maps for the reference and distorted images. It seems that gradient is also a good quality predictor of sonar images because it can capture the structure and contrast changes which contain the most information of sonar images.

5. CONCLUSION

This paper proposes a kind of image quality database designed for sonar images (SIQD). All the distortions contained in the SIQD database are collected within the actual progress of compression coding and transmission. After subjective IQA for images in this database, MOS and EOT are obtained as the quality indices. Apart from subjective quality assessment, a FR objective IQA algorithm based on local entropy (LESQP) is designed. The performance of 11 state-of-the-art FR IQA algorithms are compared with the proposed LESQP in SIQD database. Among the 12 algorithms compared in section 5, the proposed LESQP shows the better performance than other 11 FR IQA algorithms, while image gradient shows the potential to be a good quality predictor for sonar images.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (61571377, 61471308).

6. REFERENCES

[1] Kim, J., Park, K., Park, J., & Yoon, J., "Coherence bandwidth effects on underwater image transmission in multipath channel," *Japanese Journal of Applied Physics.*, vol. 50, no. 7S, pp. 07HG05, 2011.

[2] Sheikh, H., Wang, Z., Cormack, L., & Bovik, A., "LIVE Image Quality Assessment Database," http://live.ece.utexas.edu/research/quality/subjective.htm.

[3] Larson, E., & Chandler, D., "Most apparent distortion: full-reference image quality assessment and the role of strategy," *JEI*, vol. 19, no. 1, pp. 011006-011006, 2010.

[4] Gu, K., Zhai, G., Yang, X., & Zhang, W., "Hybrid no-reference quality metric for singly and multiply distorted images," *IEEE Transactions on Broadcasting*, vol. 20, no. 3, pp. 555-567, Sep. 2014.

[5] Gu, K., Liu, M., G. Zhai, X. Yang, and W. Zhang, "Quality assessment considering viewing distance and image resolution," *IEEE Transactions on Broadcasting*, vol. 61, no. 3, pp. 520-531, Sep. 2015.

[6] Gu, K., Zhai, G., Lin, W., & Liu, M., "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Transactions on Cybernetics*, vol. 46, no. 1, pp. 284-297, Jan. 2016.

[7] Wang, S., Gu, K., Zhang, X., Lin, W., Zhang, L., Ma, S., & Gao, W., "Subjective and objective quality assessment of compressed screen content images," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 6, no. 4, pp. 532-543, Dec. 2016.

[8] Duarte, A., Codevilla, F., Gaya, J., & Botelho, S., "A dataset to evaluate underwater image restoration methods," *OCEANS 2016-Shanghai. IEEE*, Apr. 2016.

[9] Wang, Z., Bovik, A., Sheikh, H., & Simoncelli, E., "Image quality assessment: from error visibility to structural similarity," *IEEE TIP*, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[10] Rouse, D., & Hemami. S., "Nature image utility assessment using image contours," *International Conference on Image Processing (ICIP)*. pp. 2217-2220, 2009.

[11] Kalwa, J., & Madsen, A., "Sonar image quality assessment for an autonomous underwater vehicle," *Automation Congress, 2004. Proceedings. World.*, vol. 15, pp. 33-38, Jun. 2004.

[12] Leier, S., & Zoubir, A., "Quality assessment of synthetic aperture sonar images based on a single ping reference," *OCEANS 2011 IEEE-Spain.*, pp. 1-4, May. 2011.

[13] Debes, C., Engel, R., Zoubir, A., & Kraft, A., "Quality assessment of synthetic aperture sonar images," *OCEANS 2009-EUROPE.*, pp. 1-4, May. 2009.

[14] Callow, H. J., Saebo, T. O., & Hansen, R., "Towards robust quality assessment of SAS imagery using the DPCA algorithm," *Europe Oceans 2005.*, vol. 2, pp. 1095-1100, Jun. 2005.

[15] Said, A., & Pearlman, W., "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on circuits and systems for video technology.*, vol. 6, no. 3, pp. 243-250, Jun. 1996.

[16] Chen, W., Yuan, F., & Cheng, E., "Adaptive Underwater Image Compression with High Robust Based on Compressed Sensing," *IEEE International Conference on Signal Processing, Communications and Computing*, 2016, in press.

[17] Yu, Y., Zhou, F., Qiao, G., & Nie, D., "Orthogonal M-ary code shift keying spread spectrum underwater acoustic communication," *Chinese Journal of Acoustics.*, vol. 62, no. 6, pp. 279-288, 2013.

[18] He, C., Huang, J., Han, J., & Zhang, Q., "Cyclic shift keying spread spectrum underwater acoustic communication," *Acta Physica Sinica.*, vol. 134, no. 12, pp. 8379-8385, 2009.

[19] Methodology for the subjective assessment of the quality of television pictures. International Telecommunication Union Recommendation ITUR BT.500-13, 2012.

[20] Shannon, C. E., "A mathematical theory of communication," *Bell System Technical Journal.*, vol. 27, no. 3, pp. 379-423, Oct. 1948.

[21] Pelli, D., & Tillman, K., "The uncrowded window of object recognition," *Nature neuroscience.*, vol. 11, no. 10, pp. 1129-1135, 2008.

[22] Liu, A., Lin, W., & Narwaria, M., "Image quality assessment based on gradient similarity," *IEEE TIP*, vol. 21, no. 4, pp. 1500-1512, 2012.

[23] Saha, S., & Vemuri, R., "An analysis on the effect of image activity on lossy coding performance," *IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 3, pp. 295-298, 2000.

[24] Zhang, L., Zhang, L., Mou, X., & Zhang, D., "FSIM: a feature similarity index for image quality assessment," *IEEE TIP*, vol. 20, no. 8, pp. 2378-2386, 2011.

[25] Chandler, D., & Hemami, S., "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE TIP*, vol. 16, no. 9, pp. 2284-2298, 2007.

[26] Gu, K., Wang, S., Zhai, G., Lin, W., Yang, X., & Zhang, W., "Analysis of distortion distribution for pooling in image quality prediction," *IEEE Transactions on Broadcasting*, vol. 62, no. 2, pp. 446-456, Jun. 2016.

[27] Xue, W., Zhang, L., Mou, X., & Bovik, A., "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE TIP*, vol. 23, no. 2, pp. 684-695, 2014.

[28] Zhang, L., Shen, Y., & Li, H., "VSI: A visual saliency-induced index for perceptual image quality assessment," *IEEE TIP*, vol. 23, no. 10, pp. 4270-4281, 2014.

[29] Gu, K., Li, L., Lu, H., Min, X., & Lin, W., "A fast reliable image quality predictor by fusing micro-and macro-structures," *IEEE Transactions on Industrial Electronics*, vol. 64,. no. 5, pp. 3903-3912, May 2017.

[30] Gu, K., Tao, D., Qiao, J., & Lin, W., "Learning a no-reference quality assessment model of enhanced images with big data," *IEEE Transactions on Neural Networks and Learning Systems*, 2017, in press.