

# Statistical and Structural Information Backed Full-Reference Quality Measure of Compressed Sonar Images

Weiling Chen<sup>ID</sup>, Ke Gu<sup>ID</sup>, Weisi Lin<sup>ID</sup>, *Fellow, IEEE*, Fei Yuan<sup>ID</sup>, and En Cheng

**Abstract**—In sonar applications, important information such as distributions of minerals, underwater creatures has a high probability of being contained in sonar images. In many underwater applications such as underwater rescue and biometric tracking, it is necessary to send sonar images underwater for further analysis. Due to the bad conditions of underwater acoustic channel and current underwater acoustic communication technologies, sonar images very possibly suffer from several typical types of distortions. As far as we know, limited efforts have been made to gather meaningful sonar image databases and benchmark reliable objective quality model, so far. This paper develops a new objective sonar image quality predictor (SIQP), whose core is the combination of two features specific to a quality measure of sonar images. These two features, which come from statistical and structural information inspired by the characteristics of sonar images and the human visual system, reflect image quality from the global and detailed aspects. The performance comparison of the proposed metric with popular and prevailing quality evaluation models is conducted using a newly established sonar image quality database. The results of experiments show the superiority of our SIQP metric over the available quality evaluation models.

**Index Terms**—Sonar image, quality evaluation, local entropy, edge, underwater acoustic transmission, human visual system.

Manuscript received September 14, 2017; revised October 16, 2018; accepted December 25, 2018. Date of publication January 4, 2019; date of current version February 5, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61571377, Grant 61871336, Grant 61527804, and Grant 61703009, in part by the Beijing Advanced Innovation Center for Future Internet Technology under Grant 110000546619001, in part by the Young Elite Scientist Sponsorship Program by China Association for Science and Technology under Grant 2017QNRC001, and in part by the Nova Programme Interdisciplinary Cooperation Project under Grant Z161100004916041. This paper was recommended by Associate Editor A. Pizurica. (*Corresponding author: En Cheng.*)

W. Chen is with the Key Laboratory of Underwater Acoustic Communication and Marine Information Technology (Xiamen University), Ministry of Education, Xiamen University, Xiamen 361005, China, and also with the College of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China (e-mail: weiling.chen@stu.xmu.edu.cn).

K. Gu is with the Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing Advanced Innovation Center for Future Internet Technology, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China (e-mail: guke.doctor@gmail.com).

W. Lin is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: wslin@ntu.edu.sg).

F. Yuan and E. Cheng are with the Key Laboratory of Underwater Acoustic Communication and Marine Information Technology (Xiamen University), Ministry of Education, Xiamen University, Xiamen 361005, China (e-mail: yuanfei@xmu.edu.cn; chengen@xmu.edu.cn).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCSVT.2019.2890878

## I. INTRODUCTION

**B**ECAUSE of the ability to take images in relatively dim light underwater, sonar has been more and more widely used during recent years. Sonar images very possibly contain important information like submarine geomorphology, marine organism, wreck remains, and so on. In many (if not all) applications, it is necessary to send images captured by sonar to a remote location for further analysis. That is, sonar images will be transmitted via an underwater acoustic channel to users. One of the most complicated channels should be the underwater acoustic channel. Bad channel condition and current underwater acoustic communication technology only provide with relatively limited bandwidth and unstable link [1]. Sonar images are likely to suffer from different distortions during transmission. Image quality assessment (IQA) plays a pivotal role for monitoring by: (1) estimating the quality degradation due to nonideal transmission; (2) optimization of compression; (3) being the basis of retransmission; (4) being a benchmark in the process of image post-processing.

Numerous IQA approaches have been proposed particularly for camera-captured natural scene images (CC-NSIs). Based on the accessibility of reference information, IQA methods can be divided into three groups. If there is no reference signal available, then the IQA method is the so called no-reference (NR) method [2]–[5]. If the reference information is partially accessible, then it is termed as a reduced-reference (RR) one [6], [7]. The third one is the full-reference (FR) method as an antithesis of the NR method under the condition of complete reference signals accessible, and thus the current NR and RR methods cannot compete with FR methods [8]. When considering the optimization and restoration which are the two most common applications of sonar IQA, the reference signal is available and FR methods are more preferable. In multimedia signal processing, authors proposed a series of FR IQA methods based on the similarity measurements of features extracted from the distorted image and its reference signal for quality judgement [9]–[12]. Another kind of FR IQA model is based on the image fidelity which is defined according to the characteristics of human visual system (HVS) [13]. Other algorithms were devised using brain principle [14], analysis of distortion distribution [15], and Lin and Kuo [16] have provided a thorough survey. The above-mentioned IQA methods all perform well for CC-NSIs, but fail in quality prediction of sonar images.

Since there are different types of sonar devices, existing sonar image quality metrics are equipment-specific. Considering the application of synthetic aperture sonar (SAS) images, its quality can be represented by the correlation of sonar ping-returns by using measurements of sonar-platform motion and estimates of environmental characteristics [17]. Besides, the quality of SAS images can also be evaluated by the degree of navigation errors [18]. It can be seen that the above-mentioned works all lose sight of the influence of HVS. In [19], the quality of forward looking sonar images are measured by mathematical features as image mean value, entropy and the significance of echoes in a NR way. The method proposed in [19] only determines whether the image is sufficient for reliable obstacle recognition instead of providing a relatively accurately quality score. Actually few effort has been devoted to quality metrics of sonar images in accordance to the characteristics of HVS, so it is necessary to devise an effective FR sonar IQA method in consideration of both HVS and applications. Images captured using distinct kinds of sonars have different characteristics. Here acoustic lens sonar and side-scan sonar are selected, since images captured by them have reasonable resolution, precision and other parameters, and similar utilities, e.g. underwater searching, rescuing, seabed mapping, underwater biological detection, etc., which require human decision because it is difficult to achieve automatic image analysis according to the current state of the art. Those selected sonar images feature different gray levels, low contrast and less detail and the information extracted from them is mainly for detecting or mapping, whereas the CC-NSIs are mostly characterized by rich color variations, thick lines and complex texture content which forms the pattern primarily for entertainment [20], and these differences give rise to a big challenge to the assessment of sonar image quality. To tackle the problem, the characteristics of sonar images and HVS are both taken into account.

From the viewpoint of sonar image characteristics, the statistical characteristics of sonar images are considered due to their ability to represent the global information. For instance, skewness and kurtosis have been used to evaluate the quality of contrast-changed images [6], and image entropy has been used in [19] as one of the features for sonar IQA. In consideration of physical significance of sonar imaging and transmission system, the target of sonar can be modelled as a stochastic source. The information entropy is defined as the expected value of the information contained in a stochastic source, and the entropy of received sonar image is often different from source entropy. In this paper, the image entropy is extracted as one important feature to represent sonar image quality according to the fact that sonar images provide a visual representation of something that cannot be perceived visually. Aforementioned statistical features show the global information without considering the interplay between adjacent pixels, which may lead the loss of some information. In this paper, we hold the opinion that the local entropy (a kind of expression of statistical information) which shows the global information and takes the influence of adjacent pixels into account might be a better quality predictor of sonar images. Sonar images are usually captured for detection and

recognition in turbid water, and can be viewed as a typical grey-scale image with low contrast, less detail and centrally distributed information. Based on the above characteristics, the structures contribute more in the representation of sonar images than other image components and the edge is one of the important representation of structure. Besides, the important information in sonar images is relatively centralized, so the edge similarity before and after transmission in the most active areas can represent the changed quality via transmission. Since most CC-NSIs are characterized by rich color variations, thick lines and complex texture content, the significance of structure-relevant features may be more limited when compared to sonar images. And the edge map in the most active area is less typical for a CC-NSI than for sonar images, since the content of CC-NSIs are relatively more rich and complex.

From the HVS aspect, it is generally known that edge is important for visual perception and plays a crucial part in image content recognition. It has been demonstrated that the image edge which the visual attention is sensitive to can capture image structures [11]. As compared with color images, the structures contribute more in IQA of grayscale sonar images. Besides, image edge contains the majority of the information about the target that viewers are interested in, and it is one of the crucial features to identify the targets contained in sonar images. Moreover, multiple strategies are used by the HVS to infer image quality [21], [22]. When the image has relative high quality, the global perspective is frequently-used. However when a sonar image is afflicted with severe degradations, there only remain part of the structures that may encompass some useful information. So people tend to pay attention to the remaining structural information when image quality is poor. To incorporate the characteristics of sonar images with the HVS properties, a strategy combining both statistical and structural information is employed in this paper for quality prediction of sonar images.

In this paper, a sonar image quality predictor (SIQP) is proposed based on the similarities of local entropy map and edge map from the perspectives of statistical and structural information. Existing work has proved that entropy is an effective feature in the IQA [23]–[27]. Entropy represents the information included in an image from the aspect of information theory. Since sonar images are usually used for practical uses, high quality sonar images should contain enough useful information. Thus the entropy can be one of the indicators of sonar image quality. In this paper, local entropy map considering the ‘uncrowded window’ effect of object recognition [28] is extracted in our work. Then the changes occurred in global information which is measured by the similarity of local entropy maps between the distorted and reference images are calculated. Though entropy is a good quality index, it tends to represent the image quality from a global perspective which reflects statistical information without considering the spatial characteristics of pixels. As mentioned earlier, the HVS employs multiple strategies for quality estimation, and statistical information on behalf of local entropy works when images have reasonable

qualities according to the HVS characteristics. When facing the poor-quality images, the structural information that from detailed aspect is more important in quality determination. To simulate such mechanism of the HVS, we detect structural information from the edge map as a supplement to statistical information. The variations of structural information reflected by similarity of edge maps in the most salient region is extracted. Based on the saliency-based pooling, the changes of statistical and structural information are pooled into two parameters. Finally, we map the proposed feature parameters to the subjective image quality by building a quadratic polynomial model.

Major contributions of this paper are as follows. First, entropy is employed in this paper considering the characteristics of sonar imaging and transmission system. Since sonar images are providing a visual representation of something that we cannot perceive visually, and photographers are providing a visual representation of what we can see, the target can be modelled as stochastic source, and the entropy can measure the information contained in an object and the relevant images. Though entropy has been widely used in the existing IQA methods, it was counted as a feature without considering saliency and the interplay between adjacent pixels of images in most cases. In this paper, local entropy map is employed by considering the interplay between adjacent pixels. In consideration of saliency, activity-based pooling is employed for local entropy similarity map in this paper. Second, the proposed SIQP method was designed in accordance with both sonar image applications and HVS. On most occasions, sonar images are used for object detection and recognition, so the observer pays more attention to image information (entropy) and structure, which are essential for sonar image applications. Since multiple strategies are used by the HVS to infer image quality, the SIQP method combines both statistical and structural information to incorporate the characteristics of sonar images with the HVS properties. Third, the proposed SIQP method employs common quadratic polynomial model without severe under-fitting or over-fitting. The quadratic polynomial model provides simple structure but superior performance. We arrange the remainder of this paper below. Section II illustrates the detail of our SIQP metric explicitly. In Section III, a comparison of the SIQP metric with the existing state-of-the-art quality models is performed using the sonar image quality database (SIQD). We conclude this paper in Section IV.

## II. METHODOLOGY

In this section, we first discuss the differences between CC-NSIs and sonar images. Then the IQA strategies for sonar images are introduced from the global and detailed aspects. From the global aspect, the similarity of local entropy maps between distorted image and its reference version is extracted as statistical information, while, as for detailed aspect, the similarity of edge maps between distorted image and its reference version in salient region presents the structural information. Finally, the statistical and structural information are systematically integrated.

### A. Differences Between CC-NSIs and Sonar Images

Nowadays, there are many research results concerning IQA methods for CC-NSIs. Whether a reference image is available or not, the existing IQA methods of CC-NSIs, especially the newly proposed methods, show good performances. It has been proved that the image structure [9], [29], the statistical information [24], [25], [30], drift of saliency [31], and image decomposition [32], etc., are highly relevant to the quality of CC-NSIs. Since sonar images are different from CC-NSIs in terms of both imaging mechanism and image characteristics, different IQA strategies are used for sonar images as compared with CC-NSIs. In what follows, the differences between sonar images and CC-NSIs will be discussed.

Actually, there are obvious differences between CC-NSIs and sonar images. First, the CC-NSIs are generated by the reflection of light while the sonar images are formed by converting the echoes into digital images. Second, most CC-NSIs are captured using a camera for human consumption while sonar images are usually deployed in underwater searching, rescuing, seabed mapping, and underwater biological detection, and so on. For human consumption relevant applications, viewers pay more attention to the aesthetical elements instead of the information contained in image. But for missions like underwater search and rescue, seabed mapping, etc., viewers care more about how much relevant information will be perceived from the images. Third, sonar images have a homogeneous characteristic which can be seen from the small variation of the pixel values, but the pixel values of CC-NSIs have wider dynamic range. Fourth, since the sonar images are taken and transmitted underwater, they have higher probability to be severely distorted than CC-NSIs. Finally, the CC-NSIs reflect what photographers see directly, while sonar images are usually used in turbid water that no one knows what will appear in the captured images.

The theoretical research and experimental results indicate that statistical and structural information is strongly correlated with sonar image quality. In the section below, from the theory and implementation, we will introduce the proposed statistical and structural information based SIQP metric, whose block diagram is shown in Fig. 1.

### B. Statistical Information Extraction

For sonar images, the target can be modelled as a stochastic source, and the echoes reflected by objects can be modelled as the outputs of this stochastic source because of uncertainty according to the information theory [33]. The flow chart of the sonar imaging and transmission system is demonstrated in Fig. 2. The more uncertainty a stochastic source contains, the more information it emits. The entropy can be used to measure the information of a stochastic source, that is, entropy can measure the information contained in an object. If there is no distortion introduced during transmission, the entropy of received image equals to the information contained in an object; otherwise, the distortion will either add some useless information or reduce the useful information. In recent decades, entropy has been showed as an effective statistical feature in the IQA of CC-NSIs. In some existing works, image

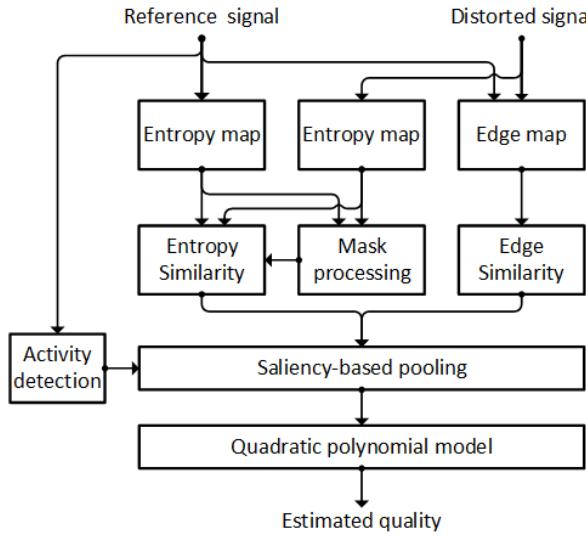


Fig. 1. Block diagram of the proposed SIQP metric.

entropy is used to be a weighting factor for pooling image blocks in IQA [27]. But most of the entropy-related studies are devoted to calculating spatial or spectral entropy as one of the quality features [23]–[26]. For example, Spatial-Spectral Entropy-based Quality (SSEQ) index [34] is one of the famous models based on spatial and spectral entropy histograms. In abovementioned studies, entropy was counted as a feature without considering the saliency and the interplay between adjacent pixels of images. And all these methods are designed without considering the characteristics of sonar images. For evaluating the quality of sonar images, we firstly extract the local entropy map which contains the statistical information of sonar images with some optimizations compared with the existing methods.

Research shows that the HVS identifies an object by feature detection and combination [35]. ‘Crowding’ occurs when objects are too close together and several features of the objects are composed into a jumbled percept. Most of the time, the majority of the human visual field is crowded, sparing only a central ‘uncrowded window’. At specific viewing distances, only the object in a local area within an ‘uncrowded window’ can be distinguished clearly, and outside of this window, objects are too crowded to be distinguished. To catch the objects that now lie outside of the window, we must move our eyes to put our window on those objects [28]. So we employ the local entropy in the proposed SIQP metric. The local entropy of the central position  $(x, y)$  is defined in an image block whose size is  $(2m+1) \times (2m+1)$ . Parameter  $m$  can be an arbitrary number for window size. We believe that its value is relevant to the content and resolution of image. Considering the relatively low resolutions of sonar images and the balance between requirements and cost,  $m$  is selected as 4, which has been proved to provide the best performance when compared with other values. For a sonar image  $I$ , the local entropy of the central position  $(x, y)$  in a  $(2m+1) \times (2m+1)$

image block is defined as:

$$H_I(x, y) = - \sum_{i=0}^{255} P_i \log P_i \quad (1)$$

where  $P_i$  denotes the gray-level distribution in the image block. Eq. (1) is used in each  $(2m+1) \times (2m+1)$  neighborhood around the corresponding pixel, and the output entropy value will be assigned to central position. In this way, we can obtain a local entropy map by moving this  $(2m+1) \times (2m+1)$  window pixel-by-pixel over the entire image. Since the HVS is highly tuned for the recognition of image edges [36], we mark the key locations based on a feature mask which is implemented by an edge detector followed a dilation operation. The feature mask is given by:

$$\begin{aligned} M_r &= h_s \circ Edge(I_r) = h_s \circ e_r \\ M_d &= h_s \circ Edge(I_d) = h_s \circ e_d \end{aligned} \quad (2)$$

where  $e_r$  and  $e_d$  denote the results of edge detection performed on reference sonar image  $I_r$  and distorted sonar image  $I_d$  respectively,  $h_s$  is a structural element, and  $\circ$  denotes the AND-operation between  $h_s$  and the binary edge map. Then we derive the masked entropy maps of  $I_r$  and  $I_d$ :

$$\begin{aligned} \hat{H}_{I_r}(x, y) &= H_{I_r}(x, y) \cdot M_r(x, y) \\ \hat{H}_{I_d}(x, y) &= H_{I_d}(x, y) \cdot M_d(x, y). \end{aligned} \quad (3)$$

One representative sonar image and its four corrupted images from the SIQD database as well as their corresponding local entropy maps and feature masks are shown in Fig. 3. For the reference sonar image, the local entropy map seems to be in a state of order that has clear edges. By comparison, the local entropy maps of distorted sonar images are in a state of disorder and the pixels are more disorganized. When the distortion is added to the image, the image tends to go from a state of order to a state of disorder [19]. The variation of entropy introduced by distortion will disturb the extraction of useful information. The lower the quality of a sonar image, the more disordered the local entropy map of this sonar image is. As shown in Fig. 3(k) to Fig. 3(o), the feature masks highlight the location of important objects (corals and fish) which are salient for viewers. Finally, the global information similarity is defined as:

$$\hat{s}(x, y) = \frac{2\hat{H}_{I_r}(x, y) \cdot \hat{H}_{I_d}(x, y) + c_1}{\hat{H}_{I_r}^2(x, y) + \hat{H}_{I_d}^2(x, y) + c_1} \quad (4)$$

where  $c_1$  is assigned as a small constant to avoid instability when  $\hat{H}_{I_r}^2(x, y) + \hat{H}_{I_d}^2(x, y)$  is very close to zero. Since small variation of information is easy to be neglected in an information clustered area,  $c_1$  is modified considering visual masking. Considering the physical significance of entropy map, each pixel represents the amount of information contained in the neighborhood of its location. In an information clustered area, small variation of information is easy to be neglected. Small variation of information will not influence the extraction of most of the information in an image block with high entropy, since the amount of variation only devotes a small percentage of total information contained in this block. For an image

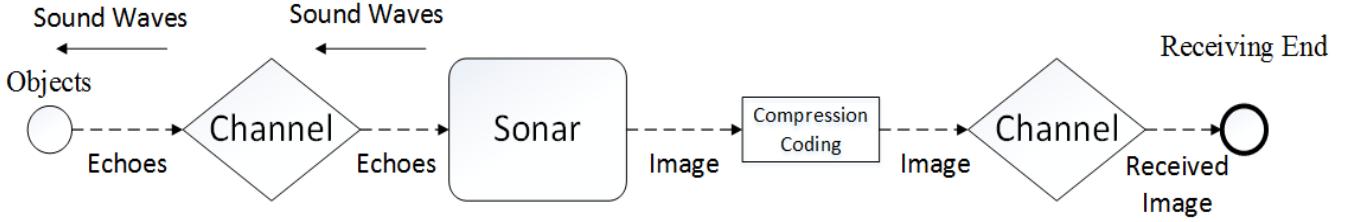


Fig. 2. Overview of the sonar imaging and transmission system.

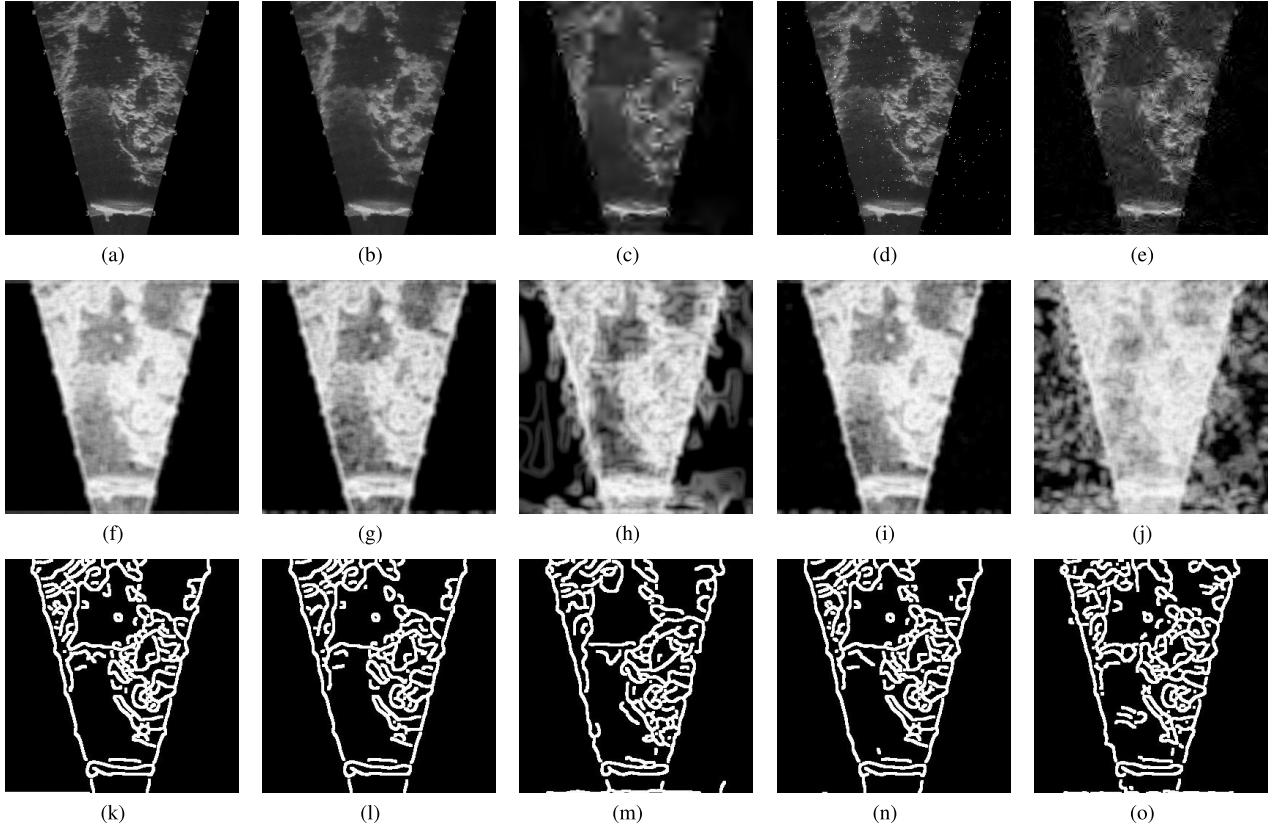


Fig. 3. (f)–(j) are local entropy maps extracted from sonar images (a)–(e), and (k)–(o) are feature masks of sonar images (a)–(e). (a) reference sonar image; (b) distorted sonar image whose mean opinion score (MOS) is 61.31; (c) distorted sonar image whose MOS is 30.4; (d) distorted sonar image whose MOS is 57.15; (e) distorted sonar image whose MOS is 29.39. (f)–(j) and (k)–(o) shows the associated local entropy maps and feature masks of sonar images (a)–(e), respectively.

block with low entropy, small variation of information will really affect the extraction of the most of the information in this block and easy to be perceived, since it will make up a large percentage of total information. The  $c_1$  is defined as:

$$c_1 = K * \min(H_{I_r}, H_{I_d}) \quad (5)$$

where  $K$  should be in a reasonable range; otherwise, the image distortion may be overestimated in some cases due to the visual masking. According to theoretical deduction and experimental results, the reasonable range for  $K$  is between 40 and 90, within which the differences of performance caused by varying  $K$  values are quite small. Readers can refer to [11] for theoretical analysis, and Section III-C for experimental results.

### C. Structural Information Extraction

In II-B, we proposed to use the image entropy measures the statistical information from the global aspect. When considering the HVS characteristics, the structural information

plays a crucial function in quantifying the quality of images. The effectiveness of edge/gradient based IQA methods along with their capability of representing image structure has been reported in [11], [37], and [38]. Besides, according to the HVS characteristics, when viewing a high-quality sonar image, the HVS attempts to extract the global information. As for low-quality images, the structural information draws more attention. An example is shown in Fig. 4. Fig. 4(a) indicates a pristine sonar image of plane wreckage, Fig. 4(b) is a high-quality sonar image ( $MOS = 69.32$ ), and Fig. 4(c) is a low-quality image ( $MOS = 34.88$ ). Obviously, Fig. 4(c) retains most information in Fig. 4(a). In situations in which the image quality is so poor as shown in Fig. 4(c), most details are destroyed and only the plane contour can be identified. The contour of the main object in the image is one of the representations of structural information and can be extracted using edge map. To approximate the HVS characteristics for low-quality sonar images, we combine an edge-based method with the aforementioned local-entropy-based method.

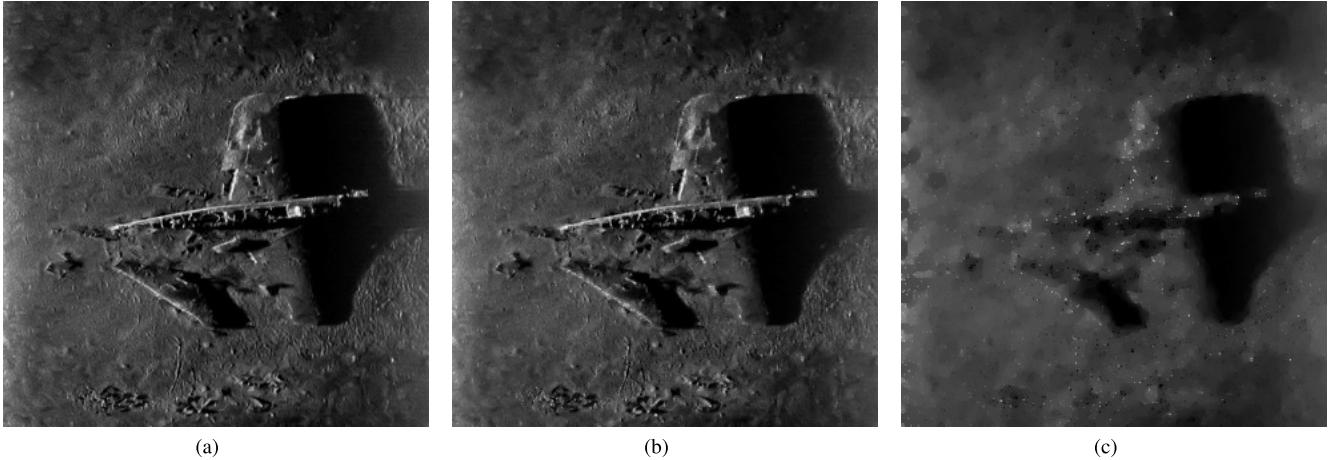


Fig. 4. Comparison of sonar images with distinct quality levels. (a) Reference image; (b) high-quality sonar image, MOS = 69.32; (c) low-quality sonar image, MOS=34.88.

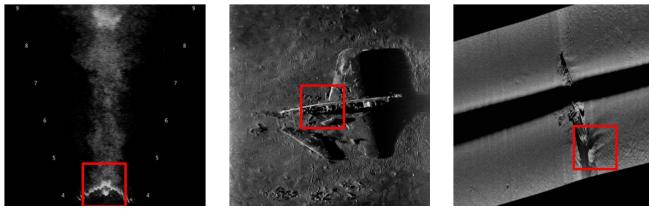


Fig. 5. Illustration of the most active areas of different sonar images.

The first step of the proposed edge-based method is to extract the salient region. In most cases, the important information is located at active areas that are more salient than inactive areas. We detect the active areas of a sonar image by image activity measurement (IAM) [40]. The activity ( $IAM_0$ ) of a  $m \times n$  image block  $I$  is defined as:

$$IAM_0 = \frac{1}{m \times n} [A + B] \quad (6)$$

where  $A$  and  $B$  are defined as:

$$\begin{aligned} A &= \sum_{i=1}^{m-1} \sum_{j=1}^n |I(i, j) - I(i+1, j)| \\ B &= \sum_{i=1}^m \sum_{j=1}^{n-1} |I(i, j) - I(i, j+1)|. \end{aligned} \quad (7)$$

We cut the sonar images into  $k_1 \times k_1$  blocks, then the  $IAM_0$  of each reference block  $\{IAM(b_{l1}), IAM(b_{l2}), \dots, IAM(b_{ln})\}$  is calculated according to Eq. (6), where  $b_{li}$  denotes the image block;  $i = 1, 2, \dots, n$  with  $n$  being the number of image blocks for one sonar image;  $IAM(\cdot)$  is the activity operator to calculate the  $IAM_0$ . Fig. 5 presents examples of the most active areas of different sonar images. Considering the content and the resolution of selected sonar images,  $k_1$  is defined as 64, which has been proved to provide the best performance when compared with other values.

As shown in Fig. 5, the most active area in a sonar image contains the important information like the swimmer, the broken part of plane wreckage, and the ship wreckage.

as highlighted in Fig. 5. The most active image block  $b_{lm}$  is selected, where  $m = \operatorname{argmax}_i IAM(b_{li})$ . The Canny edge detector is employed to extract the edge map for the most active image block. The edge maps of a pristine sonar image and its contaminated version are denoted as  $E_r = e_r^m$  and  $E_d = e_d^m$ , respectively. The structural information similarity is obtained as:

$$\hat{e}(x, y) = \frac{E_r(x, y) \& E_d(x, y) + c_2}{E_r(x, y) \parallel E_d(x, y) + c_2} \quad (8)$$

where  $c_2$  is again a small constant to avoid instability when  $E_d(x, y) \parallel E_r(x, y)$  is equal to zero. Since both  $E_r$  and  $E_d$  are logical map, ‘&’ indicates logical AND, and ‘||’ represents logical OR. The differences of performance caused by varying  $c_2$  values will be discussed in Section III-C.

#### D. Feature Integration

In this section, the saliency-based pooling is first introduced to get two feature parameters, then the quadratic polynomial model is established to integrate the extracted features.

Since the HVS is easier to be attracted by salient features, more importance should be assigned to the associated salient regions. In this paper, the saliency of a sonar image is reflected using image activity theory mentioned above. According to [28], the central vision likes an ‘uncrowded window’. When glancing over an image, viewers can recognize the objects inside of this window (in the central vision) while the objects outside of this window cannot be recognized. It can be logically infer that saliency levels of the pixels in the same ‘uncrowded window’ are very close to each other. So the activity operator works over each image block whose size is  $k_2 \times k_2$  and the  $IAM_0$  value of this block is assigned to each pixel in this block as its activity. Then the activity map  $IAM_{map}$  of an image is obtained. Besides, we have tested the performance of abovementioned block-based activity map and sliding-window-based activity map (the activity operator moves pixel-by-pixel over the entire image). The performances are very close to each other while the sliding-window-based activity map takes much more running time than block-based

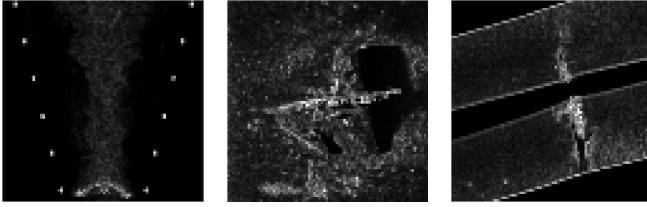


Fig. 6. Illustration of activity maps for different sonar images.

activity map. In this paper,  $k_2$  is selected as 4. The normalized activity map  $\overline{IAM}_{map}$  is achieved as a weighting function for feature pooling as:

$$\overline{IAM}_{map}(x, y) = \frac{IAM_{map}(x, y)}{\sum_x \sum_y IAM_{map}(x, y)}. \quad (9)$$

We display the activity maps of Fig. 5 in Fig. 6, where the brighter pixels indicate higher activities. The statistical information feature  $s$  and the structural information feature  $e$  are calculated using the normalized activity map as Eq. (10) and Eq. (11):

$$s = \sum_x \sum_y \hat{s}(x, y) \overline{IAM}_{map}(x, y) \quad (10)$$

$$e = \sum_x \sum_y \hat{e}(x, y) \overline{IAM'}_{map}(x, y) \quad (11)$$

where  $\overline{IAM}_{map}(x, y)$  and  $\overline{IAM'}_{map}(x, y)$  are the normalized activity maps of corresponding sonar image and the most active block of this sonar image, respectively.

Finally, as compared with different parameter models, the quadratic polynomial model which achieves the best performance as shown in Eq. (12) below is used to integrate the extracted features and generate the SIQP metric:

$$SIQP = \sum_{i=1}^2 (r_{1i}s^i + r_{2i}e^i) + r_3se \quad (12)$$

where the  $r_{1i}$ ,  $r_{2i}$  ( $i = 1, 2$ ) and  $r_3$  are the parameters of quadratic polynomial model. The details about the parameter selection and the parameter sensitivity in Eq. (12) will be discussed in Section III-C.

### III. EXPERIMENTS AND ANALYSES

#### A. Sonar Image Quality Database

All the performance comparisons are conducted in the SIQD database [39] which contains 40 reference sonar images within which some typical scenes, such as swimmer, shipwrecks, underwater creatures and seabed, were involved. These sonar images are all captured by acoustic lens sonar or side-scan sonar. In real applications, the resolution of sonar images depends on the precise model and the relevant display monitor. Since the underwater communication rate is not as good as the wireless environment, sonar images with a resolution of  $320 \times 320$  are selected. The reference sonar images are presented in Fig. 7. As mentioned in Section I, in order to collect and process underwater information with higher efficiency, we hope to achieve real-time transmission of acquired

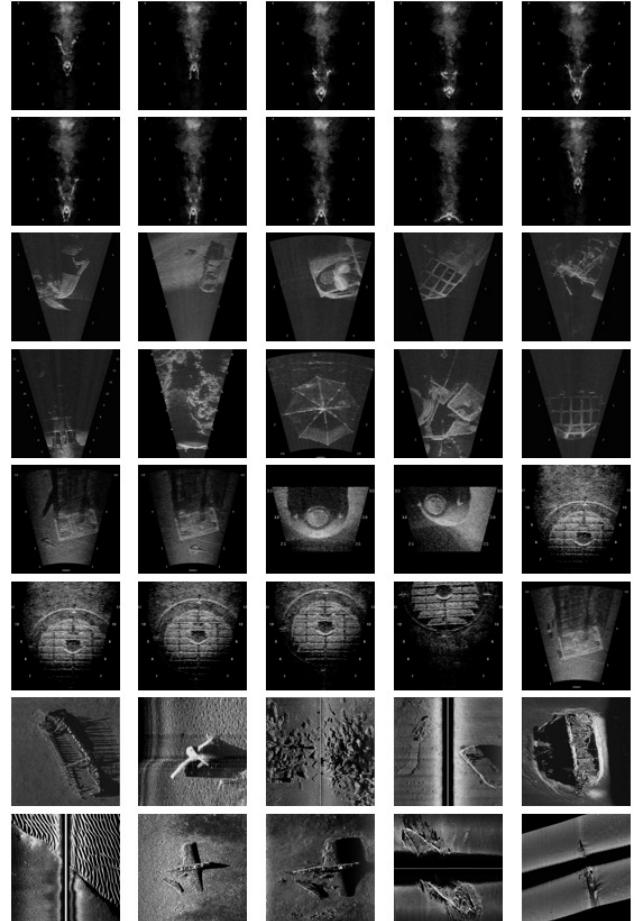


Fig. 7. Reference sonar images.

sonar images. Because of the poor condition of underwater acoustic channel, such as limited bandwidth and multipath effect, sonar images are often required to be compressed before transmission, and easily to be distorted via transmission. In this paper, we mainly discussed about the quality of transmitted sonar images. Therefore the distorted images in the SIQD database are afflicted with compression and packet loss.

The compressed sonar images in the SIQD database are obtained using Compression based on Gradient-Based Recovery (ComGBR) [41] and Set Partitioning in Hierarchical Trees (SPIHT) coding [42] which provide good robustness and high compressibility respectively. We use SPIHT and ComGBR in the SIQD database for the following reasons. First, underwater acoustic channel is a fast time-varying channel. On account of a bad channel condition, the need of a kind of compression method with high robustness is greater than wireless channel. Second, under this background, ComGBR was proposed based on the compressed sensing (CS) theory. Since each random CS measurement is with nearly equal importance [43], each bit in compressed stream shares the same importance. So the robustness of CS-based compression method is better than conventional progressive compression like SPIHT and JPEG. Due to its high robustness, ComGBR is especially suitable for underwater acoustic transmission. For more details, readers can refer to [41]. Third, when the channel

condition is good, progressive compression can be used. JPEG is a codec standard, which was created considering more about general applicability. Since the proportion of important bits of JPEG is higher than SPIHT, SPIHT is more suitable for underwater acoustic transmission. Besides, SPIHT has been used in underwater acoustic multimedia transmission [44], [45]. The compression ratio for ComGBR compression, which is defined as the ratio of amount of data after compression and before compression, is selected as 0.1, 0.2, 0.3, 0.4, 0.5, and a value depends on the content of image; for SPIHT, parameter rate (shows how many bits will be used to represent one pixel after encoding) and level (wavelet decomposition level) can be adjusted to obtain compressed images with different compression ratio. In this paper, parameter level is set as 6, and the rate is set as 0.01, 0.03, 0.1, 0.3, and 3 considering the perceived quality and compression ratio.

The distorted images due to packet loss are collected by making man-made bit error in the bit stream after SPIHT coding and ComGBR coding, respectively. Five levels of bit error ratio (BER) is used to simulate different conditions of underwater acoustic channel. The BERs are as  $10^{-1} \sim 10^{-5}$  in accordance to the recent achievements in underwater acoustic communication [46]–[48]. We have set a stochastic variable  $c$  whose value is either 1 with probability equals to BER or 0 with probability 1-BER. Each package contains 8 bits in this paper. When  $c$  equals to 0, current packet will be transmitted correctly, otherwise, it will be lost. Overall, 40 reference images and 800 distorted images were collected as the SIQD database in this study.

All the images are divided into 20 groups, and each group is tested by 25 viewers. All of them are experienced in underwater acoustic communication related works. In conformity to the suggestion of ITU-R BT.500-13 [49], the subjective test utilizes the single stimulus with multiple repetition (SSMR) method with a 5-category discrete scale for scoring. The opinion score for each test sonar image is provided by subjective viewers considering whether it will perform well in practical applications. Then, the subjective agreements among subjective viewers should be examined. The Euclidean distance (*EUD*) values and normalized cross correlation (*NCC*) values between every two subjective rating vectors for each image group are calculated. The average *EUD* values for 20 image groups are between 0.12 to 0.18 while the average *NCC* values for 20 image groups are between 0.93 to 0.95. The values of *EUD* and *NCC* respectively close to 0 and 1 confirm that the viewers have lead to an agreement on the quality scores of the test sonar images. The outlier coefficient (*OC*) with  $OC = 2.5\%$  indicates that 97.5% of the sonar images in the SIQD database had agreement among viewers. Finally, the MOS is calculated for each sonar image [49], Fig. 8 presents the histogram of MOSs for test images.

#### B. Testing Metric and Evaluation Protocols

The proposed SIQP metric will be compared with eleven IQA metrics. The selected metrics fall into two categories. The first category is composed of three classical quality metrics: (1) Structural Similarity Index (SSIM) [9] which

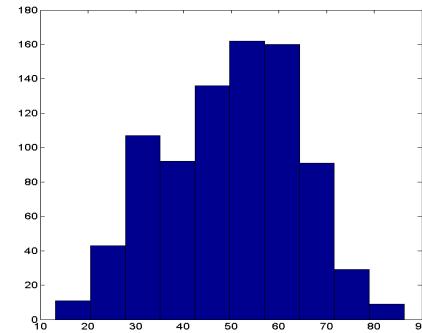


Fig. 8. Histogram of MOSs for test images.

estimates the image quality by comparing the luminance, contrast and structural information of lossless and lossy images; 2) Visual signal-to-noise ratio (VSNR) [50] that measures the visual fidelity on basis of nearthreshold and suprathreshold properties of human vision; 3) Peak signal-to-noise ratio (PSNR), assuming that distortion is only caused by additive signal-independent noise. The second category includes eight state-of-the-art metrics which have achieved good performances for CC-NSIs. These works are included in this category: 1) Feature similarity index (FSIM) [10], considering that gradient magnitude and phase congruency complement each other in quantifying the local quality of images; (2) Most apparent distortion (MAD) [21], using different quality estimation schemes for images of high- and low-quality; (3) Visual saliency-induced index (VSI) [51]; (4) Gradient magnitude similarity deviation (GMSD) [37]; (5) Perceptual similarity model (PSIM) [38]; (6) Gradient similarity (GSM) [11] based on the low-level features like gradient; (7) Analysis of distortion distribution structural similarity index (ADD-SSIM) [15] that based on a new pooling model by analysing the distortion distribution affected by image content and distortion; (8) Local-tuned-global model (LTG) [52] which under the supposition that the HVS perceives the image quality according to global quality degradation and salient local distortion. To make comparison fairly, important parameters in these selected IQA metrics have been optimized using sonar images.

In order to remove the nonlinearity caused by the subjective rating process [53], a logistic mapping based on five parameters is employed between the subjective scores (MOSs) and the objective scores  $x$ :

$$f(x) = \beta_1 \left( \frac{1}{2} - \frac{1}{1 + \exp(\beta_2(x - \beta_3))} \right) + \beta_4 x + \beta_5. \quad (13)$$

The performances of the selected IQA metrics and the proposed SIQP metric are measured using five representative correlation performance criteria: 1) Kendall rank order correlation coefficient (KROCC) and Spearman rank order correlation coefficient (SROCC), which are not influenced by any monotonic linear or nonlinear mappings applied on the objective quality scores; 2) mean absolute error (MAE), root mean square error (RMSE) and Pearson linear correlation coefficient (CC), between the objective predictions converted with a non-linear mapping and the subjective quality ratings. Among the above five evaluation criteria, KROCC and SROCC

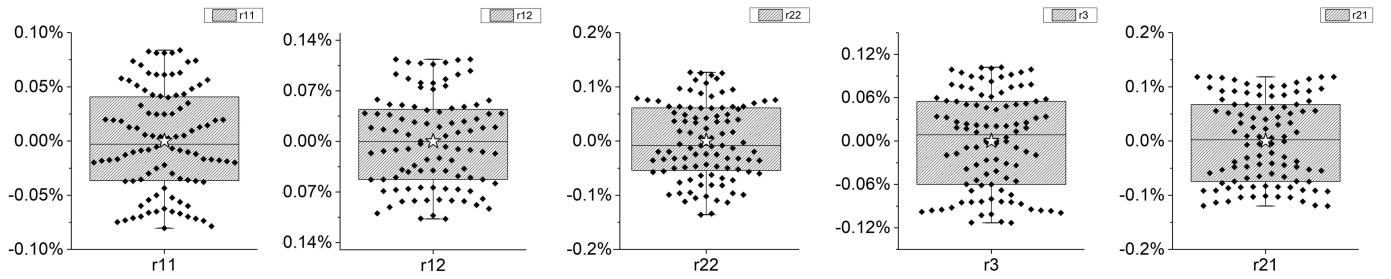


Fig. 9. Parameter distributions over 100 iterations (the horizontal axes correspond to the weighting parameters, and the vertical axes correspond to the ratios of 100 fitting results to the selected values).

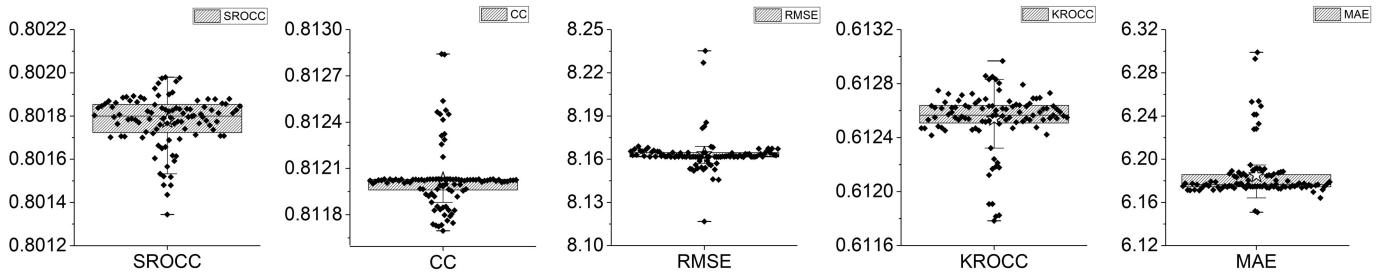


Fig. 10. Performance criterion distributions over 100 iterations (the horizontal axis corresponds to the weighting parameters, and the vertical axis corresponds to the five performance criteria).

measure the prediction monotonicity, while, MAE, RMSE and CC evaluate the prediction accuracy. A better IQA algorithm should have greater SROCC, KROCC and CC values and smaller RMSE and MAE values [54], [55].

### C. Parameter Selection and Parameter Sensitivity

A good IQA method must be insensitive to different visual contents; conversely, a poor IQA method is only valid on a few specific images, that is, it is very dependent on a training set. So, in this section, the sensitivities of the parameters of quadratic polynomial model are tested. Since the varying  $K$  and  $c_2$  values will influence the performance of the proposed SIQP metric, the choices of  $K$  and  $c_2$  will also be discussed in this section.

To integrate the statistical and the structural information features, a quadratic polynomial model which is capable of combining two features into one numerical value is built. In the proposed SIQP metric, the important parameters of quadratic polynomial model are determined as  $r_{11} = -2.28 \times 10^4$ ,  $r_{21} = 2.07 \times 10^4$ ,  $r_{12} = 0.35 \times 10^4$ ,  $r_3 = -1.88 \times 10^4$ , and  $r_{22} = 1.68 \times 10^4$ . These parameters are determined after 100 experiments with different content grouping. There are totally 40 kinds of different contents in the proposed SIQD database. The SIQD database is subdivided into different sets for training and testing (completely non-overlapping), in which four fifth of the SIQD database contents are deployed for training, and the rest one fifth are used for testing. After getting 100 randomly chosen sets for training and testing, the parameter fitting process is run over the 100 iterations on 100 training groups, and 100 groups of model parameters are obtained. These 100 groups of parameters and their box

plots are shown in Fig. 9. The five-pointed star, asterisks, horizontal line in the box indicate the mean value, the maximum and minimum value, and median value, respectively. The bottom and top of the box represent the upper and lower quartile of these 100 fitting results, respectively. As illustrated in Fig. 9, the five model parameters gather on a small scale, specifically, the biggest difference of parameters trained with different iterations are not bigger than 0.15% of the average value of each parameter (the selected value). Since the differences between parameters trained with different iterations are not so big when compared to the values of parameters, the model parameters are decided by averaging the fitting results of 100 iterations. In this paper, we hold the opinion that both the differences of parameters trained in different iterations and the differences of performances in difference iterations will be small if the model is not sensitive to difference content. To make the experimental results more convincing, we test the performances of 100 groups of fitted parameters on 100 testing groups, the distribution of five performance criteria are shown on Fig. 10. It is obvious that the performances of parameters trained in 100 iterations are similar to each other and in a relatively high standard. Thus Fig. 9 and Fig. 10 show the insensitivity of the model parameters.

To verify the contribution of the combination of edge-based method with local-entropy-based method, we tested the performance of statistical information feature and structural information feature, respectively. Then the performance of the quadratic polynomial model that integrates aforementioned two features is tested. We list the performance comparison results in Table I, and indicate in bold the best-performance of the corresponding feature. As can be observed, the combination enhances the performance of the selected features.

TABLE I  
PERFORMANCE COMPARISON OF STATISTICAL INFORMATION FEATURE,  
STRUCTURAL INFORMATION FEATURE AND COMBINATION  
OF THESE TWO FEATURES

Features	SROCC	CC	RMSE	KROCC	MAE
Statistical information feature	0.791	0.803	8.334	0.602	6.321
Structural information feature	0.762	0.785	8.668	0.569	6.728
Combination	<b>0.802</b>	<b>0.812</b>	<b>8.154</b>	<b>0.613</b>	<b>6.179</b>

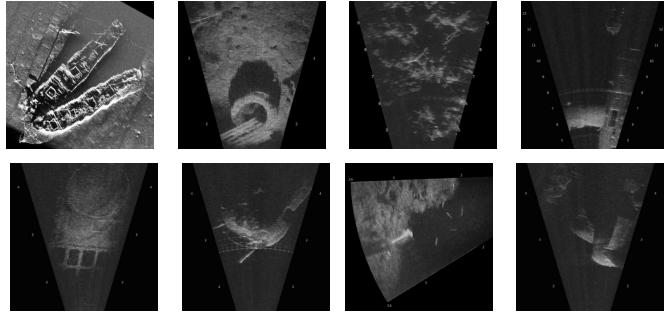


Fig. 11. Examples of sonar images for derivation of  $K$  and  $c_2$  values.

The performance gain is about 1% for SROCC and CC, 2% for RMSE, KROCC and MAE when comparing with statistical information feature. And the performance improvements over structural information feature are about 5% for SROCC, 3% for CC, 6% for RMSE, 3% for KROCC and 8% for MAE. As illustrated in Table I, the performance of the model combining the statistical and the structural information features is superior to separate performances. And the multiple strategies of the HVS to determine image quality does work for sonar images. To confirm the choices of  $K$  and  $c_2$  are reasonable, the performance of our proposed SIQP method is compared under different  $K$  and  $c_2$  values. In order to remove the impact of another parameter, the other parameter is fixed while the testing parameter is enumerated twenty numbers in an appropriate interval around assigned value. We derive values of  $K$  and  $c_2$  by analysing the characteristics of sonar images which are not included in the SIQD database. Fig. 11 shows the examples of these images. The contents of these sonar images are different from images in the SIQD database. But they have the same characteristics, and were captured in similar underwater environments by the same types of sonars (acoustic lens sonar and side-scan sonar) as images in the SIQD database.

Considering the masking effect,  $K$  is constrained into the range between 40 and 90. To confirm that the selected value for  $K$  is reasonable, the plots of selected performance criteria as a function of  $K$  are displayed in Fig. 12. Furthermore, the performance nearly remains the same when  $K$  is larger than 40 and smaller than 90, and the performance reduces with the decrease in  $K$  when it is smaller than 40 and with the increase in  $K$  when it is larger than 90, which are valid for different criteria. The decrease of performance is on account of fact that the small/large value of  $K$  (i.e.,  $K < 40$  or  $K > 90$ ) will introduce the overestimation of image distortion

in some cases. The choice of  $K = 50$  makes the performance approaches or even reaches the highest possible level.

To determine the value of  $c_2$ , we simulate various values of  $c_2$  on the proposed SIQD database. The selected five performance criteria between objective qualities and MOSs are shown in Fig. 13 to illustrate the decision of  $c_2$ . For easy comparison, only the first eleven data points are shown in Fig. 13. As can be observed in Fig. 13, when the value of  $c_2$  is close to zero, the performance drops fast. The performance criteria are nearly constant when the value of  $c_2$  is larger than 14, and the difference of criteria for different  $c_2$  is smaller than 0.1. We can derive the conclusions: 1) when  $c_2$  is close to zero, the weights of statistical information feature and structural information feature are out of balance which leads to the decrease of performance; 2) when  $c_2$  is too large, the structural information feature tends to be a constant which results in the uselessness of it. So, according to Fig. 13, a proper value of  $c_2$  is chosen to provide with the best performance.

The edge detector in Eq. (2) and Eq. (8) is Canny edge detector. For Eq. (2), the low threshold and high threshold are set to 0.08 and 0.13, respectively. The deviation of the Gaussian filter is set to 3.6. For Eq. (8), the deviation of the Gaussian filter is set to  $\sqrt{2}$ , while the low and high thresholds are chosen automatically according to the highest value of the gradient magnitude of the image. All these configurations have been optimized in the SIQD database.

#### D. Performance Test and Statistical Significance

The performance comparison of twelve IQA metrics are tabulated in Table II. The two best-performing algorithms are highlighted in bold font. To test the performances of IQA metrics for different distortions belong to the SIQD database, we designed the test image classes by binning the images such that images distorted by ComGBR compression coding and SPIHT compression coding as ‘Class 1’ and ‘Class 2’, respectively, those distorted by bit error in the ComGBR and SPIHT bit streams as ‘Class 3’ and ‘Class 4’, respectively. Sonar images in ‘Class 1’ are mainly afflicted with a kind of blur caused by ComGBR, which is a compression coding scheme based on CS, while sonar images in ‘Class 2’ contain non-eccentricity distortion and another blur-like distortion caused by SPIHT. Noise is the main distortion contained in sonar images of ‘Class 3’. There is structural degradation in sonar images from ‘Class 4’.

As listed in Table II, the proposed SIQP metric shows the best performance in all distortion types. For sonar images which belong to ‘Class 1’, both the classical and the state-of-the-art IQA metrics exhibit poor performances. Among these, VSI achieves comparable performances which are better than the other selected IQA metrics. The performances of these IQA metrics for images belonging to ‘Class 1’ show that all selected IQA metrics are not very good at evaluating the degree of blur. When considering the ‘Class 2’ and ‘Class 3’, the PSIM shows the second best performance which is worse than the SIQP but better than the other selected IQA metrics. Almost all the selected IQA metrics (including classical

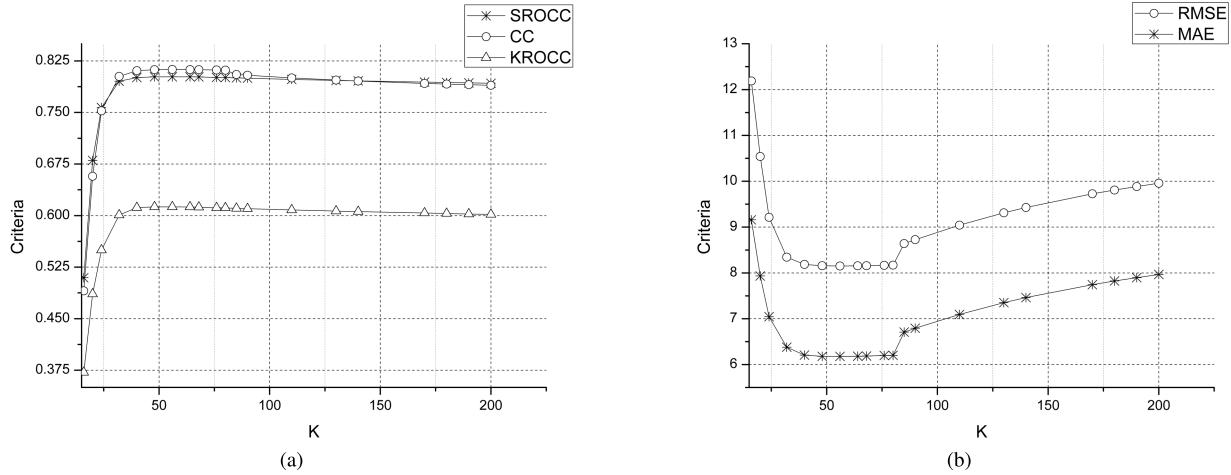


Fig. 12. Plot of (a) SROCC, CC and KROCC, (b) RMSE and MAE as a function of  $K$ .

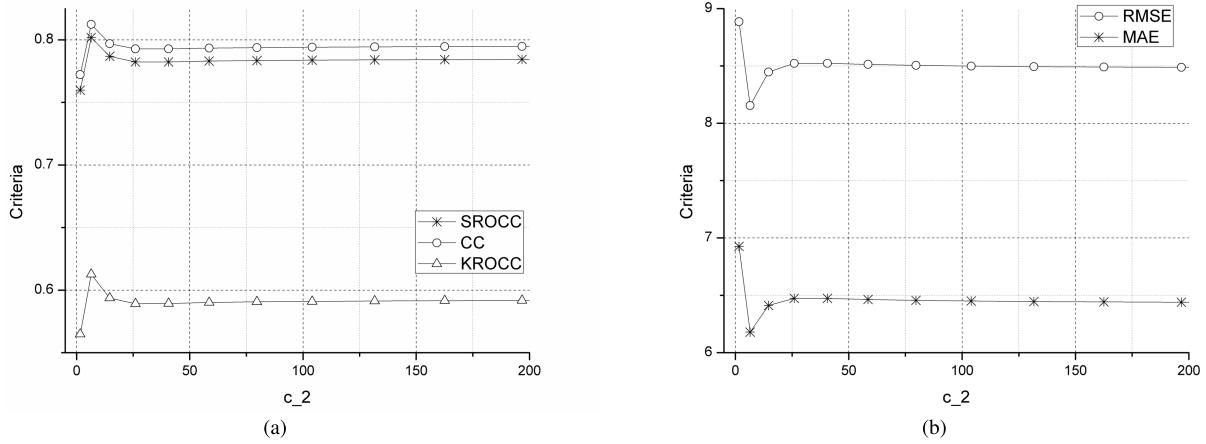


Fig. 13. Plot of (a) SROCC, CC and KROCC, (b) RMSE and MAE as a function of  $c_2$ .

methods and state-of-the-art methods) fail on the images from ‘Class 2’, which indicates that the non-eccentricity distortion may confuse the prediction of subjective quality. Nearly all the selected IQA metrics are good noise predictors for sonar images because of the good performances for ‘Class 3’ as shown in Table II. For sonar images which belong to ‘Class 4’, the majority of the selected modern IQA models show reasonable performances and the SIQP metric performs even better, and this indicates its ability to evaluate structural degradation. In general, the proposed SIQP metric has better performance for sonar images with different kinds of distortion, while, the state-of-the-art IQA metrics deliver poor performance.

To conduct an overall comparison on the entire SIQD database, we report the performances of all the selected IQA metrics and the proposed SIQP metric as tabulated in the last 5 rows of Table II. It can be seen that our SIQP metric delivers better performance than other metrics, i.e., the SIQP metric has acquired the highest SROCC, CC and KROCC, and lowest RMSE and MAE. Only the SIQP metric achieves performances greater than 0.8 for SROCC and CC and greater than 0.6 for KROCC, but lower than 9 for RMSE and lower than 7 for MAE. Relative to the second-ranking IQA metrics,

the performance gain of the proposed SIQP metric is around 5% in terms of SROCC, 5% in terms of CC and over 8%, 7%, and 9% in terms of RMSE, KROCC and MAE respectively, which indicate the promotion both in performance accuracy and monotonicity. Compared to the classical IQA metrics, the proposed SIQP metric yields a higher than 14% gain for SROCC, 13% gain for CC, 16% gain for RMSE 20% gain for KROCC and 18% gain for MAE. By comparison with those selected state-of-the-art quality metrics which show good performances in CC-NSIs, our SIQP metric also presents noticeably better performance no matter in performance accuracy or monotonicity. And as can be seen from Table II, the state-of-the-art IQA metrics show worse performances than our SIQP metric but perform better than classical IQA metrics of sonar images.

The statistical significance shows the performance statically by comparing the prediction residual. In Table III, the residuals of the SIQP metric and each of selected IQA metrics are calculated using F-test with the assumption that the significance level is set as 0.05. A value of  $H = +1$  indicates that the SIQP metric performs statistically better than the test metric, a value of  $H = 0$  shows that two metrics have statistically

TABLE II  
PERFORMANCE COMPARISON AMONG TWELVE IQA MODELS ON THE SIQD DATABASE

Distortion	Criteria	SSIM	PSNR	VSNR	FSIM	VSI	GMSD	ADD-SSIM	MAD	GSM	PSIM	LTG	SIQP
Class 1	SROCC	0.471	0.470	0.197	0.577	<b>0.697</b>	0.656	0.679	0.627	0.459	0.634	0.597	<b>0.741</b>
	CC	0.494	0.467	0.259	0.590	<b>0.734</b>	0.680	0.731	0.720	0.501	0.656	0.627	<b>0.776</b>
	RMSE	11.208	11.404	12.452	10.411	<b>8.760</b>	9.458	8.796	8.975	11.159	9.732	10.048	<b>8.140</b>
	KROCC	0.318	0.322	0.142	0.392	<b>0.510</b>	0.462	0.485	0.446	0.308	0.446	0.410	<b>0.555</b>
	MAE	8.812	9.018	9.865	8.362	<b>6.791</b>	7.434	6.830	7.050	8.796	7.620	8.032	<b>6.241</b>
Class 2	SROCC	0.561	0.473	0.318	0.585	0.574	0.617	0.586	0.524	0.496	<b>0.664</b>	0.568	<b>0.769</b>
	CC	0.603	0.500	0.370	0.645	0.602	0.634	0.598	0.536	0.569	<b>0.698</b>	0.614	<b>0.783</b>
	RMSE	9.062	9.833	10.553	8.681	9.070	8.783	9.100	9.627	9.337	<b>8.133</b>	8.966	<b>7.069</b>
	KROCC	0.385	0.328	0.210	0.394	0.401	0.425	0.406	0.358	0.331	<b>0.465</b>	0.383	<b>0.565</b>
	MAE	7.271	8.046	8.389	7.048	7.257	7.101	7.242	7.616	7.510	<b>6.630</b>	7.192	<b>5.582</b>
Class 3	SROCC	0.773	0.802	0.797	0.808	0.814	0.753	<b>0.833</b>	0.766	0.787	0.828	0.804	<b>0.848</b>
	CC	0.826	0.826	0.807	0.827	0.832	0.795	<b>0.837</b>	0.786	0.826	0.824	0.800	<b>0.849</b>
	RMSE	8.810	<b>8.806</b>	9.227	8.782	8.665	9.494	8.563	9.668	8.822	8.862	9.374	<b>8.267</b>
	KROCC	0.590	0.629	0.613	0.640	0.641	0.547	<b>0.665</b>	0.583	0.610	0.651	0.626	<b>0.677</b>
	MAE	6.147	6.050	6.973	6.000	5.896	7.081	<b>5.866</b>	7.021	6.125	6.344	6.839	<b>5.757</b>
Class 4	SROCC	0.675	0.627	0.690	0.757	0.775	0.760	0.735	<b>0.782</b>	0.665	0.771	0.722	<b>0.807</b>
	CC	0.698	0.627	0.706	0.785	0.784	0.783	0.754	<b>0.789</b>	0.697	0.788	0.755	<b>0.825</b>
	RMSE	9.907	10.760	9.789	8.567	8.581	8.593	9.083	<b>8.487</b>	9.915	8.513	9.065	<b>7.811</b>
	KROCC	0.472	0.436	0.483	0.548	0.568	0.548	0.522	<b>0.581</b>	0.465	0.562	0.505	<b>0.601</b>
	MAE	8.072	8.656	7.822	6.970	6.738	6.855	7.151	<b>6.499</b>	8.078	6.672	7.426	<b>6.222</b>
Overall	SROCC	0.654	0.622	0.451	0.706	<b>0.760</b>	0.725	0.723	0.717	0.642	0.728	0.707	<b>0.802</b>
	CC	0.673	0.639	0.493	0.721	<b>0.769</b>	0.730	0.742	0.737	0.658	0.739	0.718	<b>0.812</b>
	RMSE	10.345	10.760	12.166	9.687	<b>8.944</b>	9.551	9.376	9.450	10.533	9.423	9.740	<b>8.154</b>
	KROCC	0.469	0.443	0.312	0.510	<b>0.569</b>	0.522	0.524	0.527	0.455	0.529	0.506	<b>0.613</b>
	MAE	8.066	8.540	9.921	7.589	<b>6.850</b>	7.636	7.397	7.346	8.289	7.478	7.748	<b>6.179</b>

TABLE III

COMPARISON OF THE STATISTICAL SIGNIFICANCES OF THE SIQP METRIC AND 10 IQA METRICS ON THE SIQD DATABASE

IQA Index	SSIM +1	GSM +1	VSNR +1	FSIM +1	VSI +1
IQA Index	GMSD +1	ADD-SSIM +1	MAD +1	PSIM +1	LTG +1

TABLE IV

TIME COMPLEXITY COMPARISON (SECONDS/IMAGE) OF THE SIQP METRIC AND NINE NR IQA METRICS ON THE SIQD DATABASE

Metrics	SSIM	PSNR	VSNR	FSIM	VSI	GMSD
Cost (s)	0.023	0.008	0.017	0.236	0.097	0.014
Metrics	ADD-SSIM	MAD	GSM	PSIM	LTG	SIQP
Cost (s)	0.043	0.320	0.021	0.033	0.019	0.381

equivalent performance, while a value  $H = -1$  presents that the SIQP is statistically worse than the test metric. As listed in Table III, the results are all ‘+1’, and this indicates that our SIQP metric performs statistically better than all the selected IQA metrics in the SIQD database.

Finally, we will analyze and compare the complexity of proposed SIQD and selected FR IQA methods. Suppose that an image has  $N$  pixel. The most classical PSNR, which requires  $2N$  additions and  $N$  multiplications, has the lowest complexity. The proposed SIQP metric, which includes the calculation of local entropy map, global information map, structural information map, and activity map, requires  $268N$  additions and  $260N$  multiplications. Additionally, we also

tabulate the average time consumption of one sonar image on the SIQD database of SIQP metric and other compared methods in Table IV, and comes to the conclusions as follows. First, since the SIQP metric includes calculating a local entropy map, its time complexity is higher when compared to the other selected IQA methods. It is kind of a weakness of SIQP metric and will be improved in the further work. Second, the imaging time for sonar images is longer than CC-NSIs and the transmission rate for underwater acoustic channel is much slower than terrestrial channel, so the time consumption in quality assessment is not significant when compared to imaging time and transmitted time. Third, most of the applications of FR sonar IQA methods such as compression monitoring

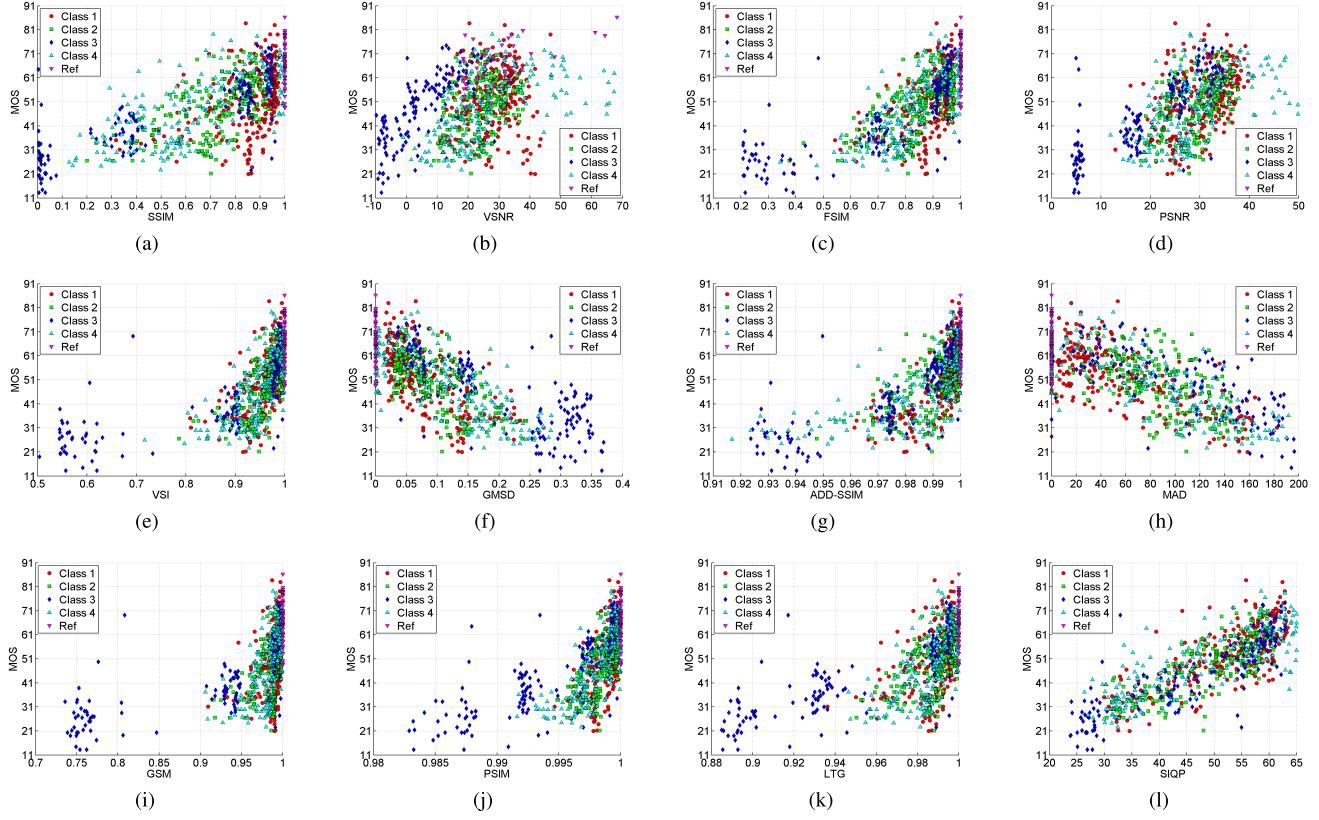


Fig. 14. The scatter plots of MOSs versus qualities assessed by different image quality assessment methods on the SIQD database. Each sample point stands for a test image. (a) SSIM; (b) VSNR; (c) FSIM; (d) PSNR; (e) VSI; (f) GMSD; (g) ADD-SSIM; (h) MAD; (i) GSM; (j) PSIM; (k) LTG; (l) SIQP.

have no strict requirement for real-time, so we care more about performance rather than complexity in this case.

#### E. Visualized Comparison

For visual comparison, the scatter plots between MOS and quality predictions yielded by the selected eleven metrics and the SIQP metric are presented in Fig. 14. The distinct symbols are used to label the sample points related to different distortion classes as mentioned in Section III-D: red circle for ‘Class 1’, green square for ‘Class 2’, deep-blue diamonds for ‘Class 3’, blue triangle for ‘Class 4’, and purple inverted triangle for reference images without any distortion. Two kinds of information can be concluded from Fig. 14. First, the distribution of sample points corresponds to the convergence and monotonicity of the quality assessment method intuitively, because there must be high correlation between quality predicted by a good image quality assessment method and MOS for an image. It is apparent that the proposed SIQP metric provides the best dependency between MOSs and predicted qualities when compared with the scatter plots of the other IQA metrics. Second, the distribution of symbols represents the robustness of relevant quality assessment methods to different distortion types. For a quality metric, when points represent the same type of distortion uniformly distributed along horizontal axis, this quality metric should be robust across different distortions. As can be seen from

Fig. 14, the proposed SIQP metric is robust across the four types of distortions contained in the SIQD database.

## IV. CONCLUSION

Sonar images are important carriers of underwater information which calls for more research attention. They are different from CC-NSIs in the aspects of imaging mechanism, image characteristics and utilities. In this work, we have first built the SIQD database that consists of 840 sonar images. Since the MOSs contained in the SIQD database are obtained according to the application of sonar images, the proposed SIQP metric is task-based. Considering the utility of sonar images and the HVS characteristics, we have proposed a local entropy and edge based IQA metric (SIQP). The local entropy is extracted for the representation of statistical feature that is more important when a sonar image has high-quality, and the edge is employed to capture the structural feature that is more important when a sonar image has low-quality. Then a quadratic polynomial model is used to integrate two features. Results of experiments present that, in comparison to the relevant classical and state-of-the-art IQA models, the SIQP metric shows better performance. And the proposed metric is also robust across different distortion types. The source code of the new model and database will be released to the public. In the future work, the weighting strategy between two features will be considered for obtaining better performance, and the improved integration model will be investigated, too.

## REFERENCES

- [1] M. Stojanovic and L. Freitag, "Recent trends in underwater acoustic communications," *Mar. Technol. Soc. J.*, vol. 47, no. 5, pp. 45–50, Sep. 2013.
- [2] K. Gu, D. Tao, J.-F. Qiao, and W. Lin, "Learning a no-reference quality assessment model of enhanced images with big data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 1301–1313, Apr. 2018.
- [3] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Using free energy principle for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 50–63, Jan. 2015.
- [4] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
- [5] Q. Li, W. Lin, J. Xu, Y. Fang, and D. Thalmann, "No-reference image quality assessment based on structural and luminance information," in *Proc. Int. Conf. Multimedia Modeling*, vol. 9516. Cham, Switzerland: Springer, 2016, pp. 301–312.
- [6] K. Gu, G. Zhai, W. Lin, and M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 284–297, Jan. 2016.
- [7] K. Gu, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "Automatic contrast enhancement technology with saliency preservation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 9, pp. 1480–1494, Sep. 2015.
- [8] T.-J. Liu, Y.-C. Lin, W. Lin, and C.-C. J. Kuo, "Visual quality assessment: Recent developments, coding applications and future trends," *APSIPA Trans. Signal Inf. Process.*, vol. 2, pp. 1–20, Jul. 2013.
- [9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [10] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
- [11] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500–1512, Apr. 2012.
- [12] K. Gu, J. Qiao, X. Min, G. Yue, W. Lin, and D. Thalmann, "Evaluating quality of screen content images via structural variation analysis," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 10, pp. 2689–2701, Oct. 2018.
- [13] 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.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.
- [14] J. Wu, W. Lin, G. Shi, and A. Liu, "Perceptual quality metric with internal generative mechanism," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 43–54, Jan. 2013.
- [15] K. Gu, S. Wang, G. Zhai, W. Lin, X. Yang, and W. Zhang, "Analysis of distortion distribution for pooling in image quality prediction," *IEEE Trans. Broadcast.*, vol. 62, no. 2, pp. 446–456, Jun. 2016.
- [16] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Communun. Image Represent.*, vol. 22, no. 4, pp. 297–312, May 2011.
- [17] D. P. Williams, "Image-quality prediction of synthetic aperture sonar imagery," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Mar. 2010, pp. 2114–2117.
- [18] C. Debes, R. Engel, A. M. Zoubir, and A. Kraft, "Quality assessment of synthetic aperture sonar images," in *Proc. IEEE Oceans-Bremen*, May 2009, pp. 1–4.
- [19] J. Kalwa and A. L. Madsen, "Sonar image quality assessment for an autonomous underwater vehicle," in *Proc. IEEE World Autom. Congr.*, Jun./Jul. 2004, pp. 33–38.
- [20] K. Gu *et al.*, "Saliency-guided quality assessment of screen content images," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1098–1110, Jun. 2016.
- [21] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, p. 011006, Jan. 2010.
- [22] K. Gu, G. Zhai, X. Yang, and W. Zhang, "A new psychovisual paradigm for image quality assessment: From differentiating distortion types to discriminating quality conditions," *Signal, Image, Video Process.*, vol. 7, no. 3, pp. 423–436, May 2013.
- [23] M. Zhao, Q. Tu, Y. Lu, Y. Chang, B. Yang, and A. Men, "No-reference image quality assessment based on phase congruency and spectral entropies," in *Proc. Picture Coding Symp.*, May/Jun. 2015, pp. 302–306.
- [24] Y. Zhang, J. Wu, G. Shi, and X. Xie, "Reduced-reference image quality assessment based on entropy differences in DCT domain," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2015, pp. 2796–2799.
- [25] W. Shi, F. Jiang, and D. Zhao, "Image entropy of primitive and visual quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2016, pp. 2087–2091.
- [26] Y. Xu, J. Li, J. Wang, R. Deng, and Y. Dong, "Quality evaluation of adaptive optical image based on DCT and Rényi entropy," in *Proc. Sel. Papers Conf. Photoelectron. Technol. Committee Chin. Soc. Astronaut.*, II, vol. 9522, 2015.
- [27] S. A. Fezza, M.-C. Larabi, and K. M. Faraoun, "Stereoscopic image quality metric based on local entropy and binocular just noticeable difference," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 2002–2006.
- [28] D. G. Pelli and K. A. Tillman, "The uncrowded window of object recognition," *Nature Neurosci.*, vol. 11, no. 10, pp. 1129–1135, Oct. 2008.
- [29] Q. Li, W. Lin, and Y. Fang, "BSD: Blind image quality assessment based on structural degradation," *Neurocomputing*, vol. 236, pp. 93–103, May 2016.
- [30] Q. Li, W. Lin, Y. Fang, X. Zhang, and Y. Zhang, "No-reference image quality assessment based on local region statistics," in *Proc. 30th IEEE Conf. Vis. Commun. Image Process. (VCIP)*, Nov. 2016, pp. 1–4.
- [31] W. Hou and X. Gao, "Saliency-guided deep framework for image quality assessment," *IEEE Multimedia*, vol. 22, no. 2, pp. 46–55, Apr./Jun. 2015.
- [32] F. Lu, Q. Zhao, and G. Yang, "A no-reference image quality assessment approach based on steerable pyramid decomposition using natural scene statistics," *Neural Comput. Appl.*, vol. 26, no. 1, pp. 77–90, Jan. 2015.
- [33] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Oct. 1948.
- [34] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Process., Image Commun.*, vol. 29, no. 8, pp. 856–863, Sep. 2014.
- [35] S. Hochstein and M. Ahissar, "View from the top: Hierarchies and reverse hierarchies in the visual system," *Neuron*, vol. 36, no. 5, pp. 791–804, 2002.
- [36] L. Zhang, D. Zhang, and X. Mou, "RFSIM: A feature based image quality assessment metric using Riesz transforms," in *Proc. Int. Conf. Image Process. (ICIP)*, Sep. 2010, pp. 321–324.
- [37] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684–695, Feb. 2014.
- [38] K. Gu, L. Li, H. Lu, X. Min, and W. Lin, "A fast reliable image quality predictor by fusing micro- and macro-structures," *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 3903–3912, May 2017.
- [39] W. Chen, Y. Fang, K. Gu, F. Yuan, and E. Cheng, "Subjective quality assessment of sonar images based on underwater acoustic transmission," *Signal Image Video Process.*, 2018.
- [40] S. Saha and R. Vemuri, "An analysis on the effect of image activity on lossy coding performance," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, vol. 3, May 2000, pp. 295–298.
- [41] W. Chen, F. Yuan, and E. Cheng, "Adaptive underwater image compression with high robust based on compressed sensing," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput.*, Aug. 2016, pp. 1–6.
- [42] 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.*, vol. 6, no. 3, pp. 243–250, Jun. 1996.
- [43] C. Deng, W. Lin, B.-S. Lee, and C. T. Lau, "Robust image coding based upon compressive sensing," *IEEE Trans. Multimedia*, vol. 14, no. 2, pp. 278–290, Apr. 2012.
- [44] Q. Li, Z. Wang, and C. Li, "Efficient underwater image coding algorithm suitable for underwater acoustic channel transmission," *Comput. Eng. Appl.*, vol. 48, no. 6, pp. 209–213, 2012.
- [45] Z. Liu, "Research of image coding technology based on wavelet transform," M.S. thesis, Dept. Underwater Acoustic Eng., Harbin Eng. Univ., Harbin, China, 2015.
- [46] Y. Yin, F. Zhou, G. Qiao, and S. Liu, "Orthogonal multicarrier M-ary cycle shift keying spread spectrum underwater acoustic communication," *Acta Phys. Sinica*, vol. 62, no. 22, pp. 224–302, 2013.
- [47] Y. Yu, F. Zhou, and G. Qiao, "Orthogonal M-ary code shift keying spread spectrum underwater acoustic communication," *Chin. J. Acoust.*, vol. 62, no. 3, pp. 279–288, 2014.
- [48] C. He, J. Huang, J. Han, and Q. Zhang, "Cyclic shift keying spread spectrum underwater acoustic communication," *Acta Phys. Sinica*, vol. 134, no. 12, pp. 8379–8385, 2009.
- [49] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document Rec. ITUR BT.500-13, International Telecommunication Union, 2012.

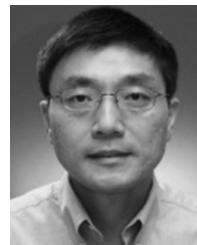
- [50] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.
- [51] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency-induced index for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4270–4281, Aug. 2014.
- [52] K. Gu, G. Zhai, X. Yang, and W. Zhang, "An efficient color image quality metric with local-tuned-global model," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2014, pp. 506–510.
- [53] Video Quality Expert Group (VQEG). (2003). *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment II*. [Online]. Available: <http://www.vqeg.org/>
- [54] J. Wu, W. Lin, G. Shi, and A. Liu, "Reduced-reference image quality assessment with visual information fidelity," *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1700–1705, Nov. 2013.
- [55] Q. Li, W. Lin, J. Xu, and Y. Fang, "Blind image quality assessment using statistical structural and luminance features," *IEEE Trans. Multimedia*, vol. 18, no. 12, pp. 2457–2469, Dec. 2016.



**Weiling Chen** received the B.S. and Ph.D. degrees in communication engineering from Xiamen University, Xiamen, China, in 2013 and 2018, respectively. In 2016, she was visiting at the School of Computer Science and Engineering, Nanyang Technological University, Singapore. She is currently a Lecturer with Fuzhou University, Fuzhou, China. Her current research interests include image quality assessment, image compression, and underwater acoustic communication. She is the Reviewer for ICIP 2016, ICIP 2017, IEEE ACCESS, DSP, T-CSVT, and T-IP.



**Ke Gu** received the B.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015, respectively. He is currently a Professor with the Beijing University of Technology, Beijing, China. His research interests include environmental perception, image processing, quality assessment, and machine learning. He received the Best Paper Award from the IEEE TRANSACTIONS ON MULTIMEDIA, the Best Student Paper Award at the IEEE International Conference on Multimedia and Expo in 2016, and the Excellent Ph.D. Thesis Award from the Chinese Institute of Electronics in 2016. He was the Leading Special Session Organizer of the VCIP 2016 and the ICIP 2017, and serves as a Guest Editor for the DIGITAL SIGNAL PROCESSING journal. He is currently an Associate Editor of the IEEE ACCESS and the IET Image Processing. He is a Reviewer for 20 top SCI journals.



**Weisi Lin** (F'16) received the Ph.D. degree from Kings College London. He is currently an Associate Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests include image processing, visual quality evaluation, and perception-inspired signal modeling, with more than 340 refereed papers published in international journals and conferences. He is a fellow of the Institution of Engineering Technology and an Honorary Fellow of the Singapore Institute of Engineering Technologists. He has been elected as an APSIPA Distinguished Lecturer from 2012 to 2013. He served as a Technical-Program Chair for the Pacific-Rim Conference on Multimedia 2012, the IEEE International Conference on Multimedia and Expo 2013, and the International Workshop on Quality of Multimedia Experience 2014. He has been on the Editorial Boards of the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON MULTIMEDIA from 2011 to 2013, the IEEE SIGNAL PROCESSING LETTERS, and the *Journal of Visual Communication and Image Representation*.



**Fei Yuan** received the B.S., M.S., and Ph.D. degrees in electronic engineering from Xiamen University, Xiamen, China, in 2002, 2005, and 2008, respectively. He is currently an Associate Professor and a Supervisor for Master's with Xiamen University. His main research interests include underwater acoustic communication, marine information, image processing, and embedded system. He is a member of the China Institute of Communications, the Acoustical Society of China, IEEE, and IET. He is the Reviewer of the *Journal of Xiamen University*, the *Chinese Journal of Acoustics*, the *Chinese Journal of Scientific Instrument*, *Applied Acoustic*, *IJNAOE*, and so on. He is also the Reviewer of provincial and state projects, such as the NSFC Natural Science Foundation of China, scientific planning subjects of Xiamen, and so on.



**En Cheng** received the B.S., M.S., and Ph.D. degrees in electronic engineering from Xiamen University, Xiamen, China, in 1985, 1988, and 2006, respectively. Since 2002, he has been a Senior Member of the Chinese Institute of Electronics. From 2003 to 2012, he was the Department Head of the Department of Communication Engineering, Xiamen University. He is the Standing Deputy Director of the Key Laboratory of Underwater Acoustic Communication and Marine Information Technology (Xiamen University), Ministry of Education, Xiamen. He is currently a Professor and a Supervisor for Ph.D. candidates at Xiamen University. He published more than 50 papers in international journals and conferences. His main research interests include underwater acoustic communication and networks.