

Semi-Reference Sonar Image Quality Assessment Based on Task and Visual Perception

Weiling Chen, *Member, IEEE*, Ke Gu, *Member, IEEE*, Tiesong Zhao, *Senior Member, IEEE*, Gangyi Jiang, *Senior Member, IEEE*, and Patrick Le Callet, *Fellow, IEEE*

Abstract—In submarine and underwater detection tasks, conventional optical imaging and analysis methods are not universally applicable due to the limited penetration depth of visible light. Instead, sonar imaging has become a preferred alternative. However, the capture and transmission conditions in complicated and dynamic underwater environments inevitably lead to visual quality degradation of sonar images, which might also impede further recognition, analysis and understanding. To measure this quality decrease and provide a solid quality indicator for sonar image enhancement, we propose a task- and perception-oriented sonar image quality assessment (*TPSIQA*) method, in which a semi-reference (SR) approach is applied to adapt to the limited bandwidth of underwater communication channels. In particular, we exploit reduced visual features that are critical for both human perception of and object recognition in sonar images. The final quality indicator is obtained through ensemble learning, which aggregates an optimal subset of multiple base learners to achieve both high accuracy and a high generalization ability. In this way, we are able to develop a compact but generalized quality metric using a small database of sonar images. Experimental results demonstrate competitive performance, high efficiency, and strong robustness of our method compared to the latest available image quality metrics.

Index Terms—Sonar image, semi-reference, image quality assessment (IQA), task-aware quality assessment

I. INTRODUCTION

IN muddy and dark deep marine environments, object detection is generally achieved via sonar devices because their imaging principle does not require natural light. Due to this unique imaging process, sonar images are visually different from conventional natural scene images (NSIs), as depicted in the first row of Fig. 1. The main differences are that sonar images are grayscale images with relatively little detail and contrast, whereas NSIs exhibit rich color variations, thick lines and complex texture content. In addition, the typical

Manuscript received December 23, 2019; revised April 6, 2020; accepted April 27, 2020.

This work was supported by the National Natural Science Foundation of China (61901119, 61671152). (*Corresponding author: Tiesong Zhao*.)

W. Chen and T. Zhao are with Fujian Key Lab for Intelligent Processing and Wireless Transmission of Media Information, Fuzhou University, Fuzhou 350108, China (E-mail: weiling.chen, t.zhao@fzu.edu.cn).

K. Gu is with the Faculty of Information Technology, Beijing University of Technology, Engineering Research Center of Intelligent Perception and Autonomous Control, Ministry of Education, Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing Artificial Intelligence Institute, Beijing 100124, China (E-mail: guke.doctor@gmail.com).

G. Jiang is with the Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China (E-mail: jianggangyi@nbu.edu.cn).

Patrick Le Callet is with Équipe Image, Perception et Interaction, Laboratoire des Sciences du Numérique de Nantes, Université de Nantes, France (E-mail: patrick.lecallet@univ-nantes.fr).

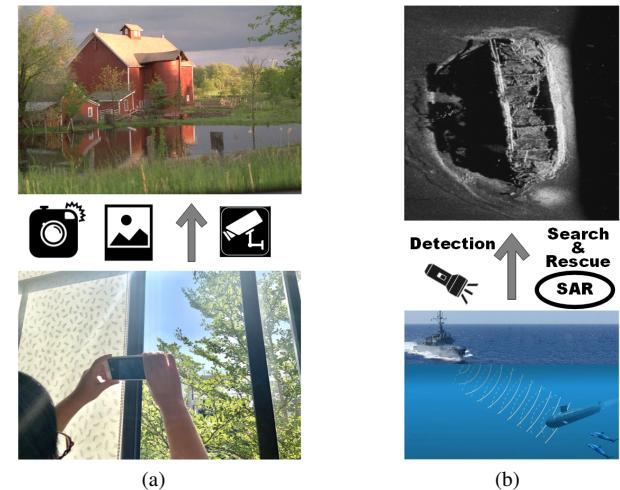


Fig. 1. Differences between an NSI and a sonar image. (a) An NSI captured by a camera, which is suitable for either aesthetic appreciation or monitoring tasks. (b) A side-scan sonar image for tasks such as detection or search and rescue.

acquisition approaches and primary application scenarios for these two types of images are also different, as illustrated in the second and third rows of Fig. 1. In consideration of the above discrepancies between sonar images and NSIs, new image processing techniques, *e.g.* image quality assessment (IQA) methods, are required to deal with sonar images. During capture and transmission, the deterioration in the underwater acoustic channels inevitably causes distortions in sonar images, which degrade the image quality and negatively impact oceanic information analysis. To evaluate and eliminate these distortions, it is essential to develop sonar IQA (SIQA) approaches, which must be specific to the capture equipment used because images obtained by different sonar devices have different characteristics. In this paper, we are mainly concerned with the quality assessment of images captured by advanced acoustic lenses and side-scan sonar devices, which have been widely used in underwater detection due to their sufficient precision and resolution [1].

Although IQA for sonar images is still a largely underexplored field, optical IQA has received considerable attention. Optical IQA methods can be classified based on the following two taxonomies, which also apply to SIQA. Depending on the accessibility of the reference information, IQA methods can be divided into three categories. In full-reference (FR) IQA methods, complete reference information is obtainable

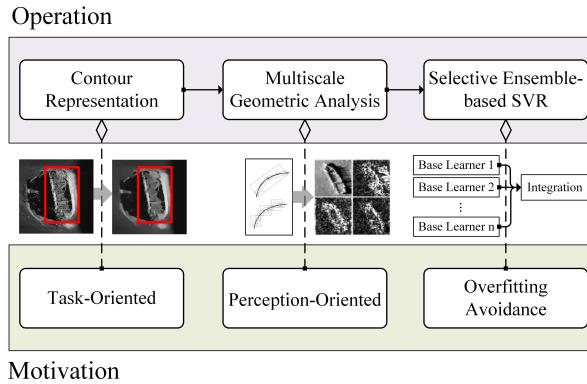


Fig. 2. The pipeline of the *TPSIQA* method.

for comparison with the test image [2]-[4]. In contrast, no-reference (NR) methods do not consider any reference information in their algorithms [5]-[9]. If an IQA method takes advantage of only a subset of reference information, it is called a semi-reference (SR) method [10]-[12]. Based on the philosophy applied when modeling features, IQA methods can also be classified into three categories. The first type in this classification is statistics-based modeling, in which a statistical model is typically employed to summarize intrinsic image characteristics from many perspectives. This can include the modeling of natural scene statistics (NSS) [13]-[14], structure and information [15]-[16], etc. The second type is perception-based modeling, which considers various perceptual characteristics of the human visual system (HVS). In essence, specific features are extracted to quantify visual attributes influencing image understanding and appreciation [17]-[20]. Deep-learning-based modeling is the third type. This type of method focuses on creating a relationship between the high-level features and perceived quality of images [21]-[25].

Despite the previously mentioned efforts, even state-of-the-art NSI quality methods are not suitable for sonar images [26]. The reasons are threefold. First, the statistical characteristics of sonar images are different from those of NSIs; thus, NSS-based quality indicators cannot be used for SIQA. Second, the size of typical sonar image databases is limited because sonar image acquisition is much more difficult and expensive than NSI acquisition. Hence, popular deep-learning-based IQA methods may lead to serious overfitting on sonar images. Finally, quality evaluations of NSIs focus on perceptual quality, whereas we tend to assess utility quality for images with specific applications, such as sonar images. According to both the theoretical analysis and the experimental results in [27]-[28], perceptual quality is not a good substitute for utility quality.

Task-oriented IQA methods have been explored for medical images, for which device-, operation- or application-relevant parameters are usually obtained [29]-[31]. Some task-oriented IQA methods have been proposed for synthetic aperture sonar and forward-looking sonar images with low precision and resolution. Measurements of sonar platform motion, environmental characteristics, and the degree of navigation errors have been employed to represent the quality of synthetic aperture

sonar images [32]-[33]. In [34], a forward-looking sonar image quality method was proposed in which various low-level features are extracted to determine whether the image quality is sufficient for reliable obstacle recognition. However, new sonar imaging technologies have been emerging worldwide for complex detection tasks. An increasing number of sonar images require human decisions in practical applications. Hence, we have proposed a series of perception-oriented SIQA methods in [35]-[37]. To achieve better estimation of utility quality for high-resolution sonar images, an SIQA method should be not only perception-oriented but also task-oriented. Unfortunately, too little work has been devoted to this topic, to our knowledge.

To fill this void, this paper proposes an SR task- and perception-oriented SIQA (*TPSIQA*) method, whose pipeline is shown in Fig. 2. We summarize the major contributions of this work as follows.

- 1) Proposing an SR SIQA method that requires less reference information than state-of-the-art SIQA methods. This is of great practical significance because of the restricted bandwidth of underwater acoustic transmission.
- 2) Proposing a task- and perception-oriented strategy that considers both the practical tasks and the visual characteristics of sonar images. By contrast, most NSI quality metrics consider only perception and show inferior performance on sonar images.
- 3) Building multiple base learners with different subsets of features and then selectively pruning some of them before aggregation. In this way, the overfitting problem caused by limited data can be effectively mitigated.
- 4) Conducting a robustness test of the proposed method with distorted reference information. The results indicate the worst channel conditions permissible for the proposed method.

The overall structure of this study is divided into four sections, including this introductory section. Section II presents details about the proposed method, including feature extraction, learner generation, exclusion, and stacking. In Section III, we report performance comparisons of the *TPSIQA* method with state-of-the-art models to validate the superiority of the proposed method. The paper is concluded in Section IV.

II. METHODOLOGY

To explore the features of sonar images, we collected typical images acquired with advanced acoustic lenses and side-scan sonar devices as important subjects for IQA. The objects in the selected images include the most common types of scenes that are the subjects of detection, such as swimmers, shipwrecks, underwater creatures, and the seabed. Further investigations and the algorithm design are elaborated as follows.

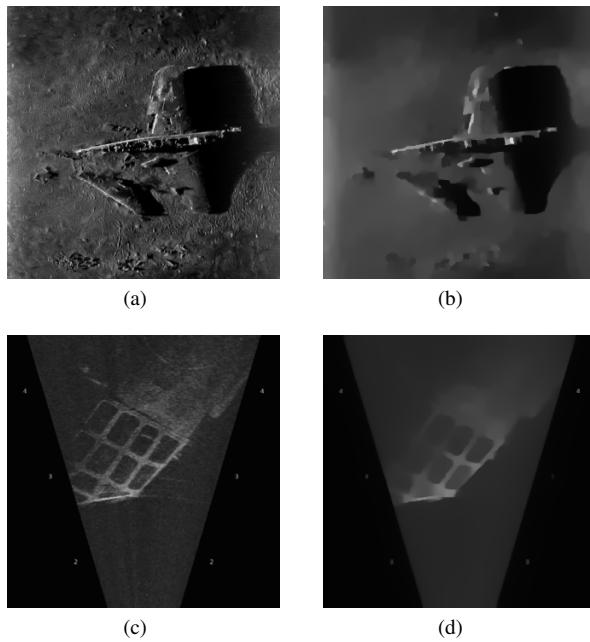


Fig. 3. Two pairs consisting of an original sonar image and its global structure. (a) Original sonar image “Aircraft Wreckage”. (b) The global structure of “Aircraft Wreckage”. (c) Original sonar image “Windscreens”. (d) The global structure of “Windscreens”.

A. Design of the SR SIQA Method

As mentioned before, the applications of FR IQA are limited for scenarios in which unimpaired reference images are not available. Although most widely used IQA methods follow the FR approach because of its good performance, complete information is generally unavailable in underwater transmission scenarios. Since NR methods lose some prior knowledge of the content, they are universally inferior to IQA methods that consider reference information. Thus, the SR approach is selected in this paper as a trade-off between evaluation accuracy and the amount of reference data required. In addition, the performance of the current work on SR IQA is usually discussed without considering the degradation of the reference information, *i.e.*, all evaluations adopt the lossless reference assumption. In fact, however, the reference information cannot always be reliably transmitted due to the poor conditions of underwater acoustic channels. Consequently, the robustness of an SR SIQA method will significantly affect its performance. In this paper, we reduce the amount of reference information needed to ease the transmission burden. Moreover, the performance with lossless reference information and the robustness to distorted reference information of the proposed method are both tested.

Considering the significant features for SR SIQA, we incorporate both practical task characteristics and visual characteristics. On the one hand, the evaluated quality of a sonar image should consider its utility for a specific application. Images containing clear and visually recognizable objects are considered high in quality for corresponding applications. On the other hand, human intervention is still necessary in some challenging tasks for which current image processing

capabilities are insufficient. As a result, a combination of task and perception characteristics is finally utilized in our architecture. Since the global structure captures the main morphology of objects in a sonar image, we evaluate sonar image quality by measuring the variances of the global structure due to transmission. Given that the optimal scale for object recognition depends on the viewing conditions [38], the global structure is described using multiscale geometric analysis in this paper. The algorithm is summarized as follows.

Step 1 Decomposing the pristine sonar image using the contourlet transform, whose coefficients formulate a concise representation of origin.

Step 2 Extracting the statistical characteristics of those coefficients as reference features to reduce the amount of reference data needed.

Step 3 Repeating the feature extraction process on the distorted sonar image.

Step 4 Mapping the feature variance to subjective quality by means of selective ensemble-based support vector regression (SVR).

The employed learning machine is characterized by a random subspace of features and learner pruning against overfitting. It is readily proven that employing multiple learners can improve the performance of a quality evaluation module in the case of a small training set [39].

B. Task- and Perception-Aware Feature Extraction

Considering the requirements of object recognition, the extracted features should be relevant to the global structure. The importance of structural information to the HVS for utility quality has been verified by previous studies. The conclusion that image structure is important for real applications such as target recognition has been validated in [27] and [28]. A significant number of studies have been published on structure-aware IQA approaches [51], [40]-[41]. These methods evaluate image quality according to both global (*i.e.*, image contours) and local (*i.e.*, image textures) structures. According to [42], most sonar image applications are related to object detection. It has been demonstrated that the global structure dominates the representation of objects, while the local structure can affect the visual effect of an image [37]. Thus, the global structure is preferred for SIQA.

We have designed a subjective test to examine whether the above conclusion (*i.e.*, the importance of structure to object detection) is still true for sonar images. Moreover, this conclusion is restricted to the claim that the global structure contains the most important information for target recognition. Fifty sonar images were selected as test samples, whose global structures were extracted [43] to form image pairs. We show two of the pairs as examples in Fig. 3. We invited 25 viewers with basic knowledge of underwater target recognition to provide their personal opinions regarding whether the target could be recognized in the original image and the corresponding global structure. Then, the evaluation of each pair of images was

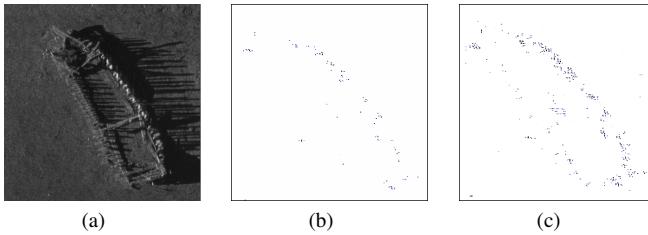


Fig. 4. Nonlinear approximations of a sonar image using the M most significant coefficients in a subspace. (a) Original sonar image. (b) $M = 516$. (c) $M = 1024$.

considered as an event. A cost was assigned to each event under the minimal error probability criterion. In this subjective test, the final average cost of all events was approximately zero. It is evident that this result supports our conclusion. For the reasons discussed above, the contour information, which is one of the primary representations of global structure, is extracted as a quality indicator.

Visually, a contour is composed of discontinuity points and smooth curves. The discontinuities at edge points can be successfully isolated by means of 2-D wavelets. However, the smoothness along contours cannot be represented in this way. Thus, only limited directional information can be captured by 2-D wavelets. To address this situation, contourlet construction has been proposed to describe both geometry and directionality in images [44]. Contourlets are more effective than wavelets in capturing the smooth contours and geometric structures in images since they provide richer sets of directions and shapes [45]. Fig. 4 shows images reconstructed using the M most significant contourlet coefficients in a subspace. It is obvious that the sketch of the ship in the test sonar image quickly becomes more refined as the value of M increases. Moreover, the refinement of the reconstructed image mainly occurs in the vicinity of the contours. This observation yields the important information that contourlet construction can quickly and efficiently offer a good description of contours. The most significant coefficients tend to be distributed at points where the contourlets adapt to both the locations and directions of contours.

The contourlet transform characterizes a discrete-domain construction whose convergence can then be studied for expansion to the continuous domain. Specifically, a multiresolution and multidirectional expansion of a discrete domain is constructed using nonseparable filter banks in the same way as the filter banks from which the wavelets are derived. Such a construction leads to a flexible multiresolution, local, and directional image expansion using contour segments. In detail, the contourlet transform consists of two major steps. A Laplacian pyramid is first used to capture point discontinuities, followed by a directional filter bank to link the point discontinuities to linear structures. The output of each stage of pyramidal and directional decomposition is defined as a subband in this paper. The number of subbands is determined by the pyramid level and the level of directional decomposition at each pyramid level. If the number of levels is zero, a critically sampled 2-D wavelet decomposition step

is performed.

Based on the above analysis, we employ the contourlet transform to represent contours in this paper. The general framework of the feature extraction process is shown in Fig. 5. As the first column of Fig. 5 shows, there will be 10 subbands in total for each image. The reference information about the contours is contained in these subbands. To compress the amount of reference information, statistical characteristics reflecting the amount of information, the energy fluctuation, and the amplitude magnitude of the contourlet transform coefficients are calculated for each subband. For the i th ($i \in [1, 10]$) subband, the coefficient at position (x, y) is denoted by $\mathcal{C}_i(x, y)$, and the amount of information is expressed as shown in Eq. (1):

$$\rho_i = \sum_{x=1}^M \sum_{y=1}^N p(\mathcal{C}_i(x, y)) \log p(\mathcal{C}_i(x, y)), \quad (1)$$

where $p(\mathcal{C}_i(x, y))$ represents the probability distribution of the coefficients in this subband and $M \times N$ is the size of this subband. Then, the energy fluctuation and amplitude magnitude are measured in the logarithmic domain as follows:

$$E_i = \frac{1}{MN} \left(\sum_{x=1}^M \sum_{y=1}^N \log |\mathcal{C}_i(x, y)| - \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N |\mathcal{C}_i(x, y)| \right), \quad (2)$$

$$\mu_i = \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N \log |\mathcal{C}_i(x, y)|. \quad (3)$$

The reference information extracted from the reference image can be stacked as shown in Eq. (4). Compared with existing SR SIQA methods, the amount of reference data needed is reduced in this paper.

$$\mathbf{F}_r = \{\rho_1, \rho_2, \dots, \rho_{10}; E_1, E_2, \dots, E_{10}; \mu_1, \mu_2, \dots, \mu_{10}\}. \quad (4)$$

The information obtained from the test image in the same fashion as \mathbf{F}_r is denoted by \mathbf{F}_d . We define the feature vector from which the contour integrity can be inferred in Eq. (5):

$$\Delta \mathbf{F} = |\mathbf{F}_r - \mathbf{F}_d|. \quad (5)$$

When distortion is added to the test sonar image, $\Delta \mathbf{F}$ is related to the quality of the distorted sonar image.

C. Selective Ensemble Learning

The feature vector $\Delta \mathbf{F}$ includes 30 elements. Since not all features share the same importance for contour representation, only l elements ($l < 30$ in this case) are randomly selected for model training. An experiment was conducted to verify that using a subset of the features is superior to using all features. First, we randomly chose l elements from $\Delta \mathbf{F}$ 1000 times. Then, SVR was applied to relate the extracted elements to the corresponding mean opinion score (MOS) values. The

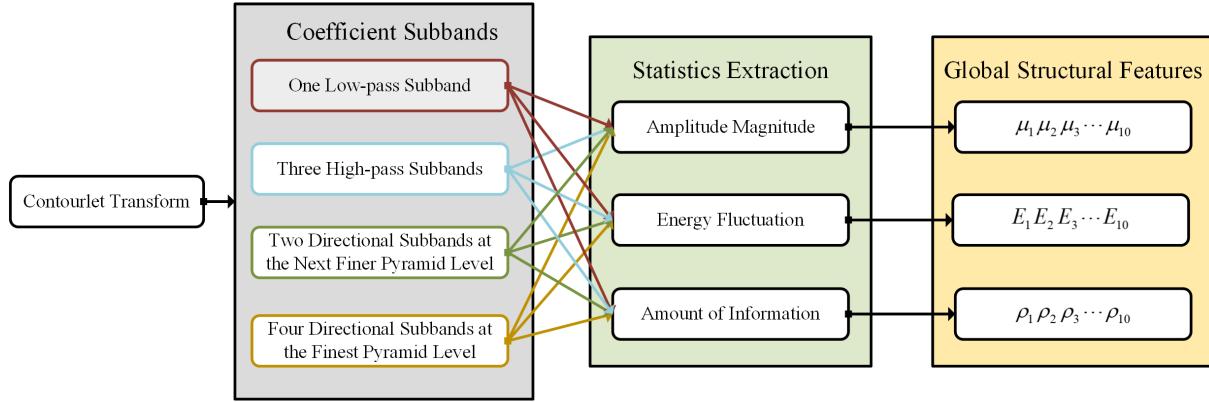


Fig. 5. A general framework for feature extraction.

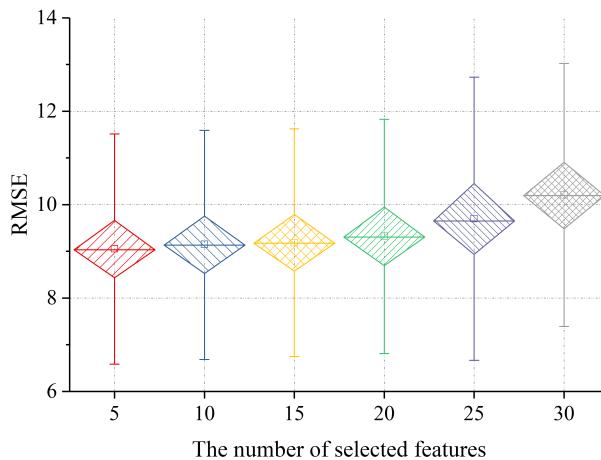


Fig. 6. Box plots of the RMSE distributions with different numbers of selected features.

prediction accuracy of SVR was measured using the root mean square error (RMSE). Finally, for each selected value of l , 1000 RMSE values were obtained, the distribution of which was depicted as a box plot, as shown in Fig. 6. The values of l were chosen from the sequence $5, 10, \dots, 30$. From Fig. 6, we can conclude that directly employing the complete feature set for quality evaluation may be inferior to using only some of the features, since overfitting might be introduced when the entire feature set is used. In addition, a decrease in the length of the feature vector can lead to better performance. Hence, l is set to five in this paper based on Fig. 6. The process of feature selection is repeated 50 times. As a result of feature selection, only partial subsets of features are generated for the base learner ensemble. Using each aforementioned partial feature subset, a base learner can be obtained. We repeat the feature selection and base learner training process m times.

Accordingly, m base learners can be obtained for the ensemble. However, not all obtainable learners will offer the same contribution to the performance of the ensemble due to their performance differences [46]. In this paper, a selection technique is utilized to exclude base learners with insufficient

performance. We employ the generalization error as a performance indicator and derive a corresponding performance threshold. A base learner with a generalization error larger than this threshold is considered a negative base learner to be excluded. Accordingly, the positive base learners constitute the remaining base learners for the ensemble.

The output of the i th base learner is $L_i(s)$, where $i \in [1, m]$. s follows a distribution $\Psi(s)$, the expected and actual outputs of which are \tilde{s} and $L_i(s)$, respectively. Then, the output of the ensemble on s can be defined as

$$\hat{L}(s) = \sum_{i=1}^m \omega_i L_i(s), \quad (6)$$

where $0 \leq \omega_i \leq 1$ and $\sum_{i=1}^m \omega_i = 1$. $\text{Er}_i(s)$ and $\widehat{\text{Er}}(s)$ are used to denote the generalization error of the i th base learner on s and the generalization error of the ensemble on s , respectively. The definitions of $\text{Er}_i(s)$ and $\widehat{\text{Er}}(s)$ are shown in Eqs. (7) and (8):

$$\text{Er}_i(s) = (L_i(s) - \tilde{s})^2, \quad (7)$$

$$\widehat{\text{Er}}(s) = (\hat{L}(s) - \tilde{s})^2. \quad (8)$$

Then, the generalization errors Er_i and $\widehat{\text{Er}}$ of the i th base learner and the ensemble on the distribution $\Psi(s)$ can be expressed as

$$\text{Er}_i = \int \text{Er}_i(s) \Psi(s) ds, \quad (9)$$

$$\widehat{\text{Er}} = \int \widehat{\text{Er}}(s) \Psi(s) ds. \quad (10)$$

The correlation between the i th and j th base learners is

$$\text{corr}_{ij} = \int \Psi(s) (L_i(s) - \tilde{s})(L_j(s) - \tilde{s}) ds. \quad (11)$$

It is easy to prove that $\text{corr}_{ij} = \text{corr}_{ji}$ and $\text{corr}_{ii} = \text{Er}_i$. According to Eqs. (6)-(8), $\widehat{\text{Er}}(s)$ becomes

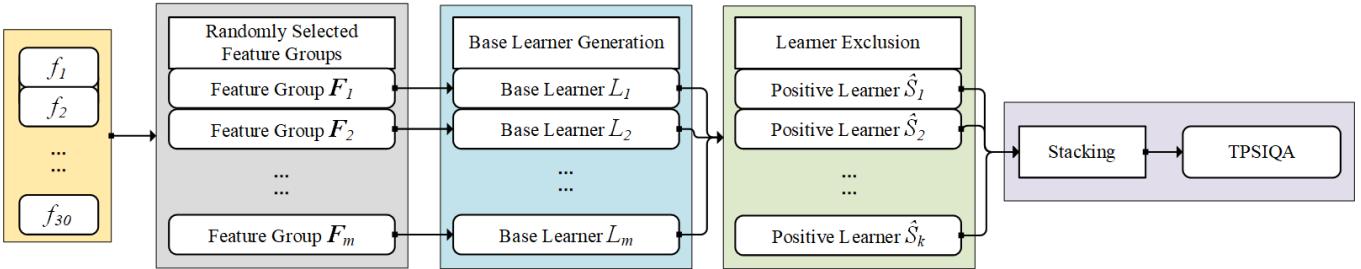


Fig. 7. The general framework of the selective ensemble learning technique.

$$\widehat{Er}(s) = \left(\sum_{i=1}^m \omega_i L_i(s) - \tilde{s} \right) \left(\sum_{j=1}^m \omega_j L_j(s) - \tilde{s} \right). \quad (12)$$

Combining Eqs. (10)-(12), we obtain

$$\begin{aligned} \widehat{Er} &= \sum_{i=1}^m \sum_{j=1}^m \omega_i \omega_j \int L_i(s) L_j(s) ds \\ &\quad - \tilde{s} \sum_{i=1}^m \omega_i \int L_i(s) ds - \tilde{s} \sum_{j=1}^m \omega_j \int L_j(s) ds + \tilde{s}^2 \quad (13) \\ &= \sum_{i=1}^m \sum_{j=1}^m \omega_i \omega_j \text{corr}_{ij}. \end{aligned}$$

In this paper, we initialize identical weights for all base learners, that is, $\omega_i = \frac{1}{m}$ ($i = 1, 2, \dots, m$). Hence, we rewrite Eq. (13) as

$$\widehat{Er} = \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m \text{corr}_{ij}. \quad (14)$$

Then, each learner is examined to determine whether it should be excluded. When the κ th learner is examined, the generalization error after excluding this learner is

$$\widehat{Er}^+ = \frac{1}{(m-1)^2} \sum_{\substack{i=1 \\ i \neq \kappa}}^m \sum_{\substack{j=1 \\ j \neq \kappa}}^m \text{corr}_{ij}. \quad (15)$$

From Eqs. (14) and (15), we obtain

$$(m-1)^2 \widehat{Er}^+ = m^2 \widehat{Er} - 2 \sum_{\substack{i=1 \\ i \neq \kappa}}^m \text{corr}_{i\kappa} - Er_\kappa. \quad (16)$$

If the exclusion of the κ th base learner will make the final ensemble better, \widehat{Er} should not be smaller than \widehat{Er}^+ , that is,

$$\widehat{Er} \geq \widehat{Er}^+. \quad (17)$$

Considering Eqs. (14)-(16), it can be derived from Eq. (17) that

$$\widehat{Er} \leq \frac{1}{2m-1} \left(2 \sum_{\substack{i=1 \\ i \neq \kappa}}^m \text{corr}_{i\kappa} + Er_\kappa \right). \quad (18)$$

The above equation can be further simplified by substituting in Eq. (14):

$$Er_\kappa \geq \frac{2m-1}{m^2} \sum_{i=1}^m \sum_{j=1}^m \text{corr}_{ij} - 2 \sum_{\substack{i=1 \\ i \neq \kappa}}^m \text{corr}_{i\kappa}, \quad (19)$$

based on which the threshold Thr_κ is defined as $\frac{2m-1}{m^2} \sum_{i=1}^m \sum_{j=1}^m \text{corr}_{ij} - 2 \sum_{\substack{i=1 \\ i \neq \kappa}}^m \text{corr}_{i\kappa}$. If the generalization error of the κ th base learner is larger than Thr_κ , that is, $Er_\kappa \geq Thr_\kappa$, then this base learner should be excluded. After the pruning of the negative base learners, the ensemble is implemented by averaging the outputs of the remaining positive base learners. Fig. 7 depicts the general framework of the selective ensemble learning technique, where f_i denotes one component of ΔF extracted from all training images. Since there are 30 components in ΔF , $i \in [1, 30]$. F_i ($i \in [1, m]$) represents the selected feature group, where each group includes five kinds of features. L_i ($i \in [1, m]$) denotes the base learner trained using the i th selected feature group, and \hat{S}_i ($i \in [1, k]$) is the i th selected positive base learner.

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. Evaluation Protocols and Methods for Comparison

1) *Database description:* The performance of the IQA methods considered in this paper is validated on the Sonar Image Quality Database (SIQD) [8]. This database consists of 40 reference sonar images obtained with acoustic lens sonar and side-scan sonar devices in an underwater scenario. The resolution of the selected sonar images is 320×320 . The number of distorted sonar images included in the SIQD database is 800. The distortions contained in the SIQD database fall into four categories, each with four to five distortion levels. The first category includes distortions caused by a compression coding scheme called ComGBR [47], which manifest as a kind of blur. The noncentricity distortions and other blur-like distortions caused by set partitioning in hierarchical trees (SPIHT) compression coding [48] are mainly contained in the second category. Noise induced by manmade bit errors in the bitstreams of ComGBR coding comprises the third category. The main expressions of distortion contained in the fourth category are messy, unnatural artifacts introduced by manmade bit errors in the bitstreams of SPIHT coding. The MOS value of each sonar image is also included in the SIQD database. Subjective testing was implemented among

viewers with expertise in underwater target recognition. For subjective score collection, the viewers were asked to rate the quality of each test sonar image considering the clarity and recognizability of the objects in the image for the detection task.

2) *Cross-validation*: Based on the SIQD database, all sonar images were divided into 40 groups, four of which were reserved for testing, while the rest were used for training. Each group comprised one reference sonar image and 20 distorted sonar images with the same content. As described in Sec. II-C, features were randomly selected for modeling. In this study, the base learner training process and feature selection process were repeated 50 times to obtain positive base learners. Finally, the selected positive base learners were integrated to obtain the *TPSIQA* score. To consider the generalization ability, e -times- \hat{k} -fold cross-validation was employed in the performance comparison, where \hat{k} was set to 10 and e was set to 10000. The average performance estimates from e -times- \hat{k} -fold cross-validation are reported for the final comparisons.

3) *Methods for comparison*: We compared the performance of the *TPSIQA* method with the performance of five state-of-the-art SR IQA methods, one of which was proposed for sonar images, while the other four are IQA methods for NSIs or screen content images (SCIs) with high citation scores. In addition, the performance of our method was compared with the performance of ten FR IQA methods, among which nine are either classical or newly proposed methods designed for NSIs and one is a state-of-the-art SIQA method. The SR IQA methods selected for comparison include the reduced-reference image quality method for contrast change (*RIQMC*) [10], the quality assessment method of contrast (*QMC*) [11], the reduced-reference wavelet-domain quality measure for SCIs (*RWQMS*) [12], orientation-selectivity-based visual patterns (*OSVP*) [49], and the partial-reference sonar image quality predictor (*PSIQP*) [36]. The considered FR IQA methods comprise the structural similarity index (*SSIM*) [2], the peak signal-to-noise ratio (*PSNR*), the gradient similarity (*GSM*) [3], the feature similarity index (*FSIM*) [50], the perceptual similarity model (*PSIM*) [51], the analysis of distortion distribution structural similarity index (*ADD-SSIM*) [52], the gradient magnitude similarity deviation (*GMSD*) [53], the local-tuned-global model (*LTG*) [54], the most apparent distortion (*MAD*) [55], and the sonar image quality predictor (*SIQP*) [35]. Only FR and SR algorithms were selected for comparison with our method since NR IQA methods usually show worse performance than reference-based IQA methods.

4) *Performance criteria*: Six commonly used criteria were chosen to report the performance of the selected IQA methods. We first determined whether the proposed method can distinguish significantly different and similar pairs according to the performance criterion proposed in [56]. In this test, the z-score defined in Eq. (20) was first calculated for each image pair (i, j) :

$$z(i, j) = \frac{|\text{MOS}(i) - \text{MOS}(j)|}{\sqrt{\frac{\text{var}(i)}{\text{Num}(i)} + \frac{\text{var}(j)}{\text{Num}(j)}}}, \quad (20)$$

where $\text{var}(\cdot)$ and $\text{Num}(\cdot)$ denote the variance of the subjective opinions and the number of subjective observers, respectively, for the corresponding image. Then, we measured the probability that the images in a pair (i, j) are different by means of the cumulative distribution function (cdf) of the normal distribution, as shown in Eq. 21:

$$p_s = \text{cdf}(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp\left(-\frac{z^2}{2}\right) dz. \quad (21)$$

The images in the pair (i, j) are considered to be significantly different when $p_s(i, j) > 0.95$. The qualities predicted by the selected IQA methods were processed as follows:

$$\Delta_{\text{method}}(i, j) = q_{\text{method}}(i) - q_{\text{method}}(j), \quad (22)$$

where $q_{\text{method}}(\cdot)$ denotes the objective quality predicted by a particular *method*. We define 95% of the values in $|\Delta_{\text{method}}|$ as the threshold for identifying significantly different and similar image pairs. Finally, the area under the receiver operating characteristic curve (AUC) was employed to determine whether a particular IQA method can identify significantly different and similar image pairs.

Five additional criteria were adopted to assess the performance of the IQA methods in terms of prediction monotonicity, accuracy and consistency. The prediction monotonicity is measured by the Spearman rank-order correlation coefficient (SROCC), Kendall's rank-order correlation coefficient (KROCC), and the monotonic correlation coefficient (MC) [57]. The prediction accuracy is assessed using the Pearson linear correlation coefficient (PLCC), while the prediction consistency is evaluated in accordance with the RMSE. Among them, the PLCC and RMSE were calculated between the MOS values and the mapped predicted qualities. The mapped predicted qualities were obtained by performing a five-parameter logistic regression between the MOS values and predicted qualities. In addition, the MC was estimated by performing a monotonic regression between the MOS values and predicted qualities q and was defined as the PLCC between the MOS values and Q_{mono} . Eq. (23) expresses the definition of monotonic regression:

$$Q_{\text{mono}} = f \uparrow \bigcup f \downarrow, \quad (23)$$

where $f \uparrow = \{h(\cdot) : h(w) \geq h(z), \forall w > z\}$ and $f \downarrow = \{h(\cdot) : h(w) \leq h(z), \forall w > z\}$, with w and $z \in q$. The SROCC and KROCC are independent of any monotonic score mapping and were measured between the MOS values and the qualities predicted by the selected IQA methods. Higher values of the SROCC, KROCC, PLCC and MC are expected for better IQA methods, while higher RMSE values indicate worse performance.

B. Performance Evaluation

1) *Feature visualization and selection*: As stated in Sec. II-B, 10 subbands are selected for feature extraction. In effect, more subbands will yield a finer description of the image contours, while fewer subbands will result in a coarser representation. Too many subbands can have negative effects

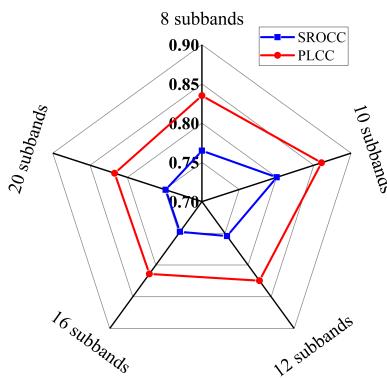


Fig. 8. Performance comparison of the proposed framework with different numbers of subbands.

on performance due to the increased amount of reference data required as well as the introduction of undesired information on the local structure. By contrast, too few subbands cannot completely capture the contour information. We conducted an experiment comparing the performance of the proposed framework with different numbers of subbands. The results are reported in terms of the performance metrics for prediction monotonicity (SROCC) and prediction accuracy (PLCC) in the radar chart shown in Fig. 8. There are five axes in the radar chart, corresponding to 8 subbands, 10 subbands, 12 subbands, 16 subbands and 20 subbands. We mark the scale for both metrics on the center axis. The red dot and blue square on each axis represent the values of the SROCC and PLCC, respectively. It can be clearly observed that 10 subbands can provide the best performance among the selections. Therefore, the sonar images were decomposed into 10 subbands for feature extraction in this study.

In addition, to show the relation between the extracted features and the sonar image quality, we attempted to characterize the effect of the image quality on the extracted features. First, the images in the SIQD database were divided into three groups according to their MOS values. The first group comprised sonar images with MOS values less than 40, the second group consisted of sonar images with MOS values between 40 and 60, and the sonar images with MOS values larger than 60 were assigned to the third group. A box plot was drawn for each of the 30 elements of ΔF in each group, where the box plots for the same element were plotted on the same coordinates. In total, 30 figures were obtained in this way to illustrate the impact of image quality on the feature values. Half of them are shown in Fig. 9; the selected plots are sufficiently representative and can illustrate the most typical impacts of distortions on the feature values. The values of the extracted features exhibit an obvious trend with respect to the MOS values. Specifically, sonar images of worse quality tend to have higher values of the corresponding elements of ΔF . The relations between the feature values and image quality scores were quantitatively determined with the help of SVR based on the selective ensemble learning technique.

After feature extraction from the contourlet coefficients, we utilized the selective ensemble approach to integrate optimal learners trained on subsets of features. We compared the

TABLE I
PERFORMANCE COMPARISON WITH AND WITHOUT SELECTIVE ENSEMBLE LEARNING ON TEST SET

Selective Ensemble	SROCC	KROCC	MC	PLCC	RMSE
w/o	0.730	0.535	0.772	0.744	9.345
w/	0.800	0.609	0.861	0.815	8.336

TABLE II
NUMBERING OF SELECTED SR METHODS

1	2	3	4	5	6
<i>TPSIQA</i>	<i>OSVP</i>	<i>RIQMC</i>	<i>QMC</i>	<i>RWQMS</i>	<i>PSIQP</i>

performance of our proposed framework with and without selective ensemble learning. The results are summarized in Tab. I, in which the better performance results are highlighted in boldface. Compared to the condition without selective ensemble learning, our proposed technique achieves noticeable performance gains of approximately 9.5% for the SROCC, 13.8% for the KROCC, 11.5% for the MC, 9.5% for the PLCC and 10.8% for the RMSE. These findings demonstrate the superiority of the selective ensemble learning method used in *TPSIQA*.

2) Performance comparison with existing IQA methods:

The SR approach is employed in the method proposed in this paper to achieve a trade-off between evaluation accuracy and the amount of data required. To ensure fair comparisons, the performance of the proposed method will be compared with the performance of other reference-based IQA methods, *i.e.*, FR and SR IQA methods, in this section.

For the SR IQA methods, we first evaluated their discrimination ability for pairs of images with significantly different qualities in the SIQD database. For convenience, the selected SR IQA methods are numbered as shown in Tab. II. Fig. 10 shows the AUCs (left) and the significance plot (right) for our method and the selected SR IQA methods. The error bars in the left panel of Fig. 10 represent 95% confidence intervals. A black box in the significance plot indicates that the method in the column significantly outperforms the method in the row. When the method in the column has significantly lower performance, the corresponding box is white. Otherwise, the box is gray. Fig. 10 clearly shows the superior performance of our method in terms of discrimination ability.

In Tab. III, prediction monotonicity, accuracy and consistency comparisons are presented between the *TPSIQA* method and the other five SR IQA methods. We highlight the best performance in bold font, and the second-best performance is underlined. As shown in Tab. III, only the proposed method and the *PSIQP* method achieve SROCC and PLCC values above 0.8. In addition, only the IQA methods designed for sonar images can achieve values greater than 0.6 for the KROCC but lower than 9 for the RMSE. In terms of the MC, the *TPSIQA* method is among the best, with a performance gain greater than 6 % relative to the second-ranking IQA

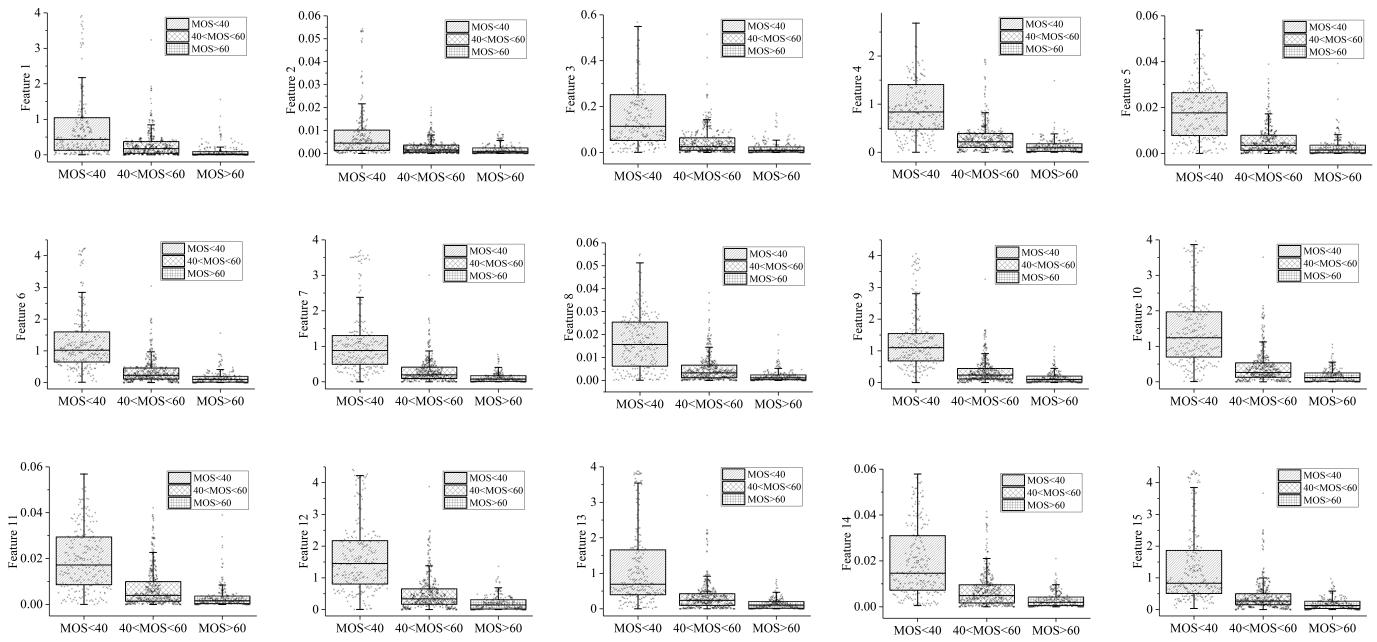


Fig. 9. Distributions of feature values for different image qualities.

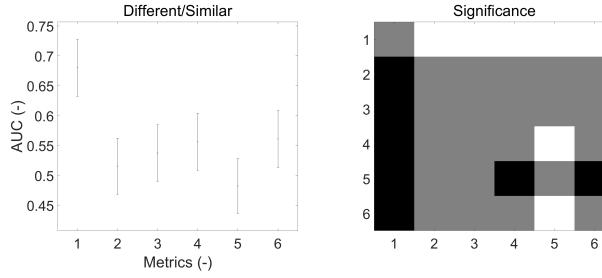


Fig. 10. AUCs and the statistical significance of the differences.

method (the *OSVP* method).

Moreover, due to the importance of statistical significance tests, we implemented the F-test between the *TPSIQA* method and the selected IQA methods. First, the confidence level was taken to be 95 %. Then, a critical threshold F_c was defined in accordance with the values of the residuals of the compared methods and the confidence level. Finally, the ratio of the prediction residual variances of the objective qualities to the MOS values was calculated, denoted by F_t , for which the objective qualities were regressed using a five-parameter logistic nonlinear regression function. When F_t is larger than F_c , the two compared IQA methods can be considered significantly different.

Tab. IV tabulates the results of the significance comparison, where “+1” represents that the *TPSIQA* method is significantly better than the compared method, “0” indicates that the performances of the *TPSIQA* method and the compared method are statistically indistinguishable, and “-1” indicates that the *TPSIQA* method is significantly worse than the compared method. As seen from Tab. IV, the performance of our method is statistically indistinguishable from that of the *PSIQP* method. The *PSIQP* method is an SR SIQA method proposed

TABLE III
PERFORMANCE COMPARISON WITH OTHER SR IQA METHODS ON THE SIQD DATABASE

IQA	SROCC	KROCC	MC	PLCC	RMSE
<i>OSVP</i>	0.744	0.553	0.812	0.770	9.157
<i>RIQMC</i>	0.405	0.285	0.564	0.611	11.326
<i>QMC</i>	0.535	0.387	0.648	0.673	10.549
<i>RWQMS</i>	0.552	0.404	0.671	0.671	10.623
<i>PSIQP</i>	0.804	0.614	0.777	0.821	8.162
<i>TPSIQA</i>	0.800	0.609	0.861	0.815	8.336

TABLE IV
COMPARISON OF THE STATISTICAL SIGNIFICANCE OF THE *TPSIQA* METHOD AND THE OTHER FIVE SR IQA METHODS ON THE SIQD DATABASE

IQA	<i>OSVP</i>	<i>RIQMC</i>	<i>QMC</i>	<i>RWQMS</i>	<i>PSIQP</i>
Index	+1	+1	+1	+1	0

in 2018 that needs 100 features as reference information. In practice, nearly 1300 bits are needed for the representation of reference information. By comparison, in the *TPSIQA* method, only 600 bits are required to represent the reference information. In other words, the *TPSIQA* method achieves statistically equivalent performance to the *PSIQP* method while requiring less reference information. This is a very important improvement for underwater acoustic transmission due to the limited bandwidth of underwater acoustic channels.

To provide a comprehensive comparison, Tab. V tabulates the performance metrics of the selected FR IQA methods

TABLE V
PERFORMANCE COMPARISON WITH FR IQA METHODS ON THE SIQD DATABASE

IQA	SROCC	KROCC	MC	PLCC	RMSE
<i>SSIM</i>	0.694	0.511	0.801	0.727	9.815
<i>PSNR</i>	0.699	0.508	0.777	0.729	9.517
<i>FSIM</i>	0.754	0.564	0.832	0.782	8.882
<i>GSM</i>	0.695	0.511	0.797	0.726	9.795
<i>ADD-SSIM</i>	0.762	0.568	0.846	0.792	8.739
<i>GMSD</i>	0.774	0.579	0.826	0.791	8.733
<i>PSIM</i>	0.776	0.584	0.816	0.797	8.633
<i>LTG</i>	0.724	0.534	0.781	0.748	9.480
<i>MAD</i>	0.763	0.577	0.829	0.785	8.874
<i>SIQP</i>	0.831	0.649	0.894	0.849	7.545
<i>TPSIQA</i>	<u>0.800</u>	<u>0.609</u>	<u>0.861</u>	<u>0.815</u>	<u>8.336</u>

TABLE VI
COMPARISON OF THE STATISTICAL SIGNIFICANCE OF THE *TPSIQA* METHOD AND THE 10 FR IQA METHODS ON THE SIQD DATABASE

IQA	<i>SSIM</i>	<i>PSNR</i>	<i>FSIM</i>	<i>GSM</i>	<i>ADD-SSIM</i>
Index	+1	+1	+1	+1	+1
IQA	<i>GMSD</i>	<i>PSIM</i>	<i>LTG</i>	<i>MAD</i>	<i>SIQP</i>
Index	+1	+1	+1	+1	-1

as well as the *TPSIQA* method. As shown in Tab. V, the IQA methods designed for sonar images are competitive with all of the other selected FR IQA methods. To be exact, the *TPSIQA* method is the second-ranking method, while *SIQP* shows superior performance compared to the rest. Notably, however, the *SIQP* method requires the original image as reference information, whereas the *TPSIQA* method needs only partial information on the original image as a reference. The performance gains of the *TPSIQA* method with respect to the *GSM* method, which is also a structure-aware feature-based IQA method, are approximately 15 % in terms of the SROCC, 19 % in terms of the KROCC, 8 % in terms of the MC, 12 % in terms of the PLCC, and 14 % in terms of the RMSE. Since the *GSM* method evaluates image quality by measuring the similarity of both the global and local structures, the introduction of local structure information may lead to a negative impact on performance.

We also tabulate the significance comparison results in Tab. VI. It can be seen from Tab. VI that the *TPSIQA* method is significantly better than the selected IQA methods designed for NSIs. The *SIQP* method, which is an FR sonar IQA method, is significantly better than the proposed *TPSIQA* method. However, the amount of reference data required by the *SIQP* method is markedly greater than that of our method.

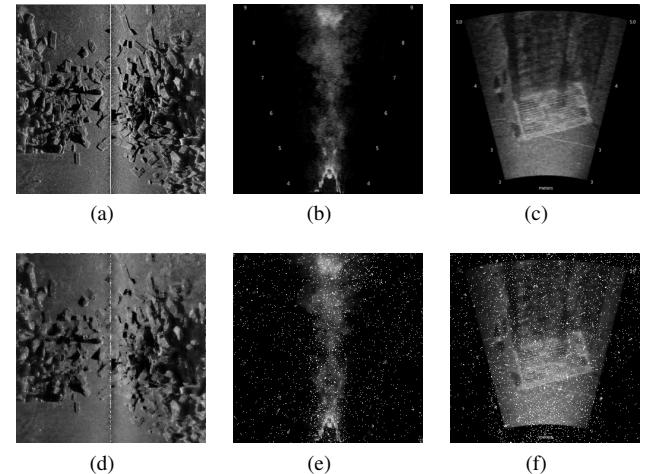


Fig. 11. *TPSIQA* scores for several examples that illustrate the good performance of the proposed method. (a)-(c) Reference sonar images. (d) Distorted version of (a) with MOS=52.72 and *TPSIQA* score=53.06. (e) Distorted version of (b) with MOS=33.85 and *TPSIQA* score=34.44. (f) Distorted version of (c) with MOS=37.23 and *TPSIQA* score=37.33.

TABLE VII
TIME CONSUMPTION (SECONDS/IMAGE) OF THE *TPSIQA* METHOD AND THE OTHER FIVE SR IQA METHODS ON THE SIQD DATABASE

Method	<i>OSVP</i>	<i>RIQMC</i>	<i>QMC</i>
Cost (s)	1.50×10^{-2}	2.01×10^{-1}	1.70×10^{-2}
Method	<i>RWQMS</i>	<i>PSIOP</i>	<i>TPSIQA</i>
Cost (s)	1.61×10^{-1}	2.80×10^{-2}	3.80×10^{-2}

3) *Intuitive Comparison and Computation Time Comparison:* To visually illustrate the performance of the proposed *TPSIQA* method, Fig. 11 shows the predicted qualities provided by our method. Fig. 11(a) to Fig. 11(c) are the reference sonar images, where Fig. 11(a) depicts underwater wreckage, Fig. 11(b) presents a swimmer and Fig. 11(c) shows a lobster defending its territory. Fig. 11(d) to Fig. 11(f) are the corresponding distorted sonar images. The quality predictions offered by the *TPSIQA* method are very close to the MOS values provided in the SIQD database.

To enable a comprehensive comparison, the *TPSIQA* method along with the other five SR IQA methods were run on the SIQD database, and the average time consumption of each method is tabulated in Tab. VII. These computational cost tests were conducted on the MATLAB R2014a software platform on a computer with a 3.20 GHz CPU and 8.00 GB of RAM. *RIQMC* and *RWQMS* require considerably more processing time for one sonar image on average than the other IQA methods. The *TPSIQA* method can process at least 20 more sonar images per second than the *RWQMS* method or the *RIQMC* method. In addition, our method is competitive with the other selected methods.

TABLE VIII
PERFORMANCE COMPARISON BETWEEN THE *NRCDM* METHOD AND THE *TPSIQA* METHOD WITH BIT ERRORS (THE SELECTED BER IS 2×10^{-3})

Method	SROCC	KROCC	MC	PLCC	RMSE
<i>NRCDM</i>	0.752	0.564	0.817	0.758	9.257
<i>TPSIQA</i>	0.760	0.575	0.827	0.772	9.097

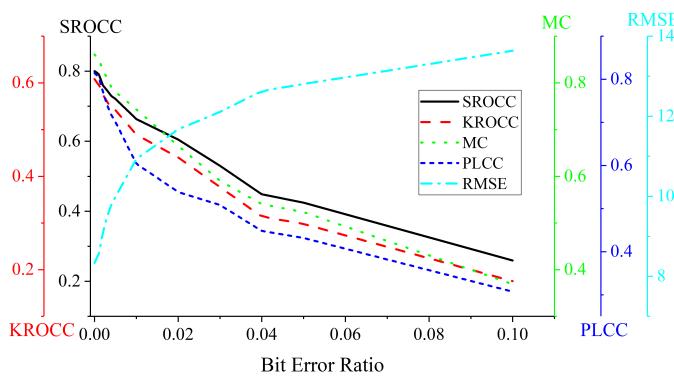


Fig. 12. Performance of the *TPSIQA* method with distorted reference information.

C. Robustness Analysis

The abovementioned experimental results for the *TPSIQA* method are all based on the assumption that there is no error in the reference information. However, although the reference information can be protected using strong error protection schemes, sacrificing transmission efficiency for reliability is not advisable in underwater acoustic channels. Therefore, the performance of SR IQA methods should also be discussed under conditions in which the reference information is distorted to varying degrees. Moreover, the switching threshold, which indicates the poorest allowable transmission conditions for the reliable transmission of reference information, is preliminarily discussed. In this study, the performance of the proposed method with distorted reference information was tested. The reference information was distorted by bit errors, and the degree of distortion was controlled by the bit error ratio (BER). As shown in Fig. 12, the performance of the *TPSIQA* method decreases as the BER increases. When the BER is lower than 2×10^{-3} , the performance of the *TPSIQA* method is nearly unaffected. When the BER reaches 3×10^{-3} , the performance of the proposed method is lower than that of the no-reference contour degradation measurement (*NRCDM*) metric [37], which was proposed using the NR approach based on the same database.

For a more intuitive comparison, a performance comparison between the *TPSIQA* method with bit errors and the *NRCDM* metric is tabulated in Tab. VIII. Considering both Fig. 12 and Tab. VIII, it can be concluded that when the BER is lower than 2×10^{-3} , *TPSIQA* is a good choice for assessing the quality of a received image. It is worth mentioning that the software-defined acoustic modem (SDAM) proposed by

Demirors has achieved real-time data rates of 104 kbit/s with a BER of 2×10^{-5} at a distance of 200 m for underwater acoustic channels [58]. When the BER is higher than 3×10^{-3} , the *NRCDM* metric can provide better performance than the *TPSIQA* method. That is, $\text{BER} = 3 \times 10^{-3}$ is the switching threshold for the *TPSIQA* method. Notably, the switching threshold discussed in this paper is based on the premise of no error protection for the reference information. In practice, the switching threshold can be represented in terms of not only the BER but also other transmission or channel parameters.

IV. CONCLUSION

Based on detection-related applications, this paper proposes a task- and perception-oriented SR SIQA method. In this method, sonar image quality is evaluated by exploiting the changes in global structure before and after distortion. The global structure is represented in terms of the amplitude, energy and entropy of contourlet transform coefficients extracted at different scales and in different directions. Then, several optimal SVR-based base learners trained on various subsets of features are selected to form an ensemble to avoid overfitting. Extensive experiments conducted on the SIQD database have confirmed that the *TPSIQA* method is comprehensively superior to the existing state-of-the-art FR IQA methods and widely cited SR IQA methods for NSIs and SCIs. Moreover, it is also competitive with FR and SR SIQA methods. Finally, the robustness of the proposed method is discussed, which sheds new light on task- and perception-driven underwater applications.

REFERENCES

- [1] M. Zhu, T. Zhang, B. Yang, Y. Liu & J. Tang, "Sonar system of Jiaolong human-occupied vehicle," *Chinese Science Bulletin*, vol. 59, no. 35, pp. 3462-3470, 2014.
- [2] Z. Wang, A.C. Bovik, H. Sheikh & E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, 2004.
- [3] A. Liu, W. Lin & M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1500-1512, 2012.
- [4] K. Gu, S. Wang, H. Yang, W. Lin, G. Zhai, X. Yang & W. Zhang, "Saliency-guided quality assessment of screen content images," *IEEE Transactions on Multimedia*, vol. 18, no. 6, pp. 1098-1110, 2016.
- [5] Y. Zhang & R. Zhang, "No-reference image sharpness assessment based on maximum gradient and variability of gradients," *IEEE Transactions on Multimedia*, vol. 20, no. 7, pp. 1796-1808, 2018.
- [6] P. Freitas, W. Akamine & M. Farias, "No-reference image quality assessment using orthogonal color planes patterns," *IEEE Transactions on Multimedia*, vol. 20, no. 12, pp. 3353-3360, 2018.
- [7] K. Gu, D. Tao, J. Qiao & W. Lin, "Learning a no-reference quality assessment model of enhanced images with big data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 4, pp. 1301-1313, 2018.
- [8] K. Gu, J. Zhou, J. Qiao, G. Zhai, W. Lin & A. C. Bovik, "No-reference quality assessment of screen content pictures," *IEEE Transactions on Image Processing*, vol. 26, no. 8, pp. 4005-4018, 2017.
- [9] K. Gu, V. Jakhetiya, J. Qiao, X. Li, W. Lin & D. Thalmann, "Model-based referenceless quality metric of 3D synthesized images using local image description," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 394-405, 2018.
- [10] K. Gu, G. Zhai, W. Lin & M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Transactions on Cybernetics*, vol. 46, no. 1, pp. 284-297, 2016.

- [11] K. Gu, G. Zhai, X. Yang, W. Zhang & C. Chen, "Automatic contrast enhancement technology with saliency preservation," *IEEE Transactions on Circuits Systems for Video Technology*, vol. 25, no. 9, pp. 1480-1494, 2015.
- [12] S. Wang, K. Gu, X. Zhang, W. Lin, L. Zhang, S. Ma & W. Gao, "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, 2016.
- [13] A. K. Moorthy & A. C. Bovik, "Blind image quality assessment: from natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350-3364, 2011.
- [14] M. A. Saad, A. C. Bovik & C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339-3352, 2012.
- [15] A. Rehman & Z. Wang, "Reduced-reference image quality assessment by structural similarity estimation," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3378-3389, 2012.
- [16] R. Soundararajan & A. C. Bovik, "RRED indeces: Reduced reference entropic differencing for image quality assessment," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 517-526, 2012.
- [17] Z. Tang, Y. Zheng, K. Gu, K. Liao, W. Wang & M. Yu, "Full-reference image quality assessment by combining features in spatial and frequency domains," *IEEE Transactions on Broadcasting*, vol. 65, no. 1, pp. 138-151, 2019.
- [18] W. Sun, Q. Liao, J. Xue & F. Zhou, "SPSIM: A superpixel-based dimilarity index for full-reference image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4232-4244, 2018.
- [19] K. Rahul & A. K. Tiwari, "FQI: Feature-based reduced-reference image quality assessment method for screen content images," *IET Image Processing*, vol. 13, no. 7, pp. 1170-1180, 2019.
- [20] W. Zhu, G. Zhai, X. Min, M. Hu, J. Liu, G. Guo & X. Yang, "Multi-channel decomposition in tandem with free-energy principle for reduced-reference image quality assessment," *IEEE Transactions on Multimedia*, vol. 21, no. 9, pp. 2334-2346, 2019.
- [21] S. Bosse, D. Maniry, K. R. Müller, T. Wiegand & W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 206-219, 2018.
- [22] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang & W. Zuo, "End-to-end blind image quality assessment using deep neural networks," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1202-1213, 2018.
- [23] E. Prashnani, H. Cai, Y. Mostofi & P. Sen, "PieAPP: Perceptual image-error assessment through pairwise preference," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018.
- [24] D. Pan, P. Shi, M. Hou, Z. Ying, S. Fu & Y. Zhang, "Blind prediction similarity quality map for image quality assessment," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018.
- [25] B. Yan, B. Bare & W. Tan, "Naturalness-aware deep no-reference image quality assessment," *IEEE Transactions on Multimedia*, vol. 21, no. 10, pp. 2603-2615, 2019.
- [26] W. Chen, F. Yuan, E. Cheng & W. Lin, "Subjective and objective quality evaluation of sonar images for underwater acoustic transmission," *2017 IEEE International Conference on Image Processing*, Beijing, China, 2017, pp. 176-180.
- [27] D. M. Rouse, S. S. Hemami & P. Le Callet, "Estimating the usefulness of distorted natural images using an image contour degradation measure," *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, vol. 28, no. 2, pp. 157-188, 2011.
- [28] D. M. Rouse, R. Pepion, S. S. Hemami & P. Le Callet, "Image utility assessment and a relationship with image quality assessment," *Human Vision and Electronic Imaging XIV*, 2009.
- [29] G. Tarroni, O. Oktay, W. Bai, A. Schuh, H. Suzuki, J. Passerat-Palmbach, A. de Marvao, D. O'Regan, S. Cook, B. Glockner, P. M. Matthews & D. Rueckert, "Learning-based quality control for cardiac MR images," *IEEE Transactions on Medical Imaging*, vol. 38, no. 5, pp. 1127-1138, 2019.
- [30] S. R. Dolly, Y. Lou, M. A. Anastasio & H. Li, "Task-based image quality assessment in radiation therapy: initial characterization and demonstration with computer-simulation study," *Physics in Medicine & Biology*, vol. 64, no. 14, pp. 1-19, 2019.
- [31] R. M. Winter, S. Leibfarth, H. Schmidt, K. Zwirner, D. Mönnich, S. Welz, N. F. Schwenzer, C. la Fougère, K. Nikolaou, S. Gatidis, D. Zips, D. Thorwarth, "Assessment of image quality of a radiotherapy-specific hardware solution for PET/MRI in head and neck cancer patients," *Radiotherapy & Oncology*, vol. 128, no. 3, pp. 485-491, 2018.
- [32] D. P. Williams, "Image-quality prediction of synthetic aperture sonar imagery," *IEEE International Conference on Acoustics Speech and Signal Processing*, Dallas, TX, USA, pp. 2114-2117, 2010.
- [33] C. Debes, R. Engel, A. Zoubir & A. Kraft, "Quality assessment of synthetic aperture sonar images," *OCEANS 2009-EUROPE*, Bremen, Germany, 2009.
- [34] J. Kalwa & A. Madsen, "Sonar image quality assessment for an autonomous underwater vehicle," *Proceedings World Automation Congress*, Seville, Spain, pp. 33-38, 2004.
- [35] W. Chen, K. Gu, W. Lin, F. Yuan & E. Cheng, "Statistical and structural information backed full-reference quality measure of compressed sonar images," *IEEE Transactions on Circuits and Systems for Video Technology*, 2019.
- [36] W. Chen, K. Gu, X. Min, F. Yuan, E. Cheng, & W. Zhang, "Partial-reference sonar image quality assessment for underwater transmission," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 6, pp. 2776 - 2787, 2018.
- [37] W. Chen, K. Gu, W. Lin, Z. Xia, P. Le Callet, & E. Cheng, "Reference-free quality assessment of sonar images via contour degradation measurement," *IEEE Transactions on Image Processing*, vol. 28, no. 11, pp. 5336 - 5351, 2019.
- [38] Z. Wang, E. P. Simoncelli & A. C. Bovik, "Multi-scal structural similarity for image quality assessment," *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, Pacific Grove, CA, USA, 2003.
- [39] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123-140, 1996.
- [40] Z. Shi, J. Zhang, Q. Cao, K. Pang & T. Luo, "Full-reference image quality assessment based on image segmentation with edge feature," *Signal Processing*, vol. 145, pp. 99-105, 2018.
- [41] Q. Li, W. Lin & Y. Fang, "BSD: Blind image quality assessment based on structural degradation," *Neurocomputing*, vol. 236, pp. 93-103, 2016.
- [42] W. Chen, F. Yuan, E. Cheng & C. Gao, "Sonar image quality assessment based on degradation measurement," *2018 OCEANS - MTS/IEEE Kobe Techno-Oceans (OTOC)*, Kobe, Japan, 2018.
- [43] T. M. Murphy & L. H. Finkel, "Shape representation by a network of V4-like cells," *Neural Networks*, vol. 20, no. 8, pp. 851-867, 2007.
- [44] M. N. Do, & M. Vetterli, "The contourlet transform: an efficient directional multiresolution image Representation," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2091-2106, 2006.
- [45] W. Dong, H. Bie, L. Lu & Y. Li, "Image quality assessment by considering multiscale and multidirectional visibility differences in shearlet domain," *IEEE Access*, vol. 7, pp. 78715-78728, 2019.
- [46] K. Gu, Z. Xia & J. Qiao, "Stacked selective ensemble for PM2.5 forecast," *IEEE Transactions on Instrumentation and Measurement*, 2019.
- [47] W. Chen, F. Yuan & E. Cheng, "Adaptive underwater image compression with high robust based on compressed sensing," *IEEE International Conference on Signal Processing, Communications and Computing*, Hong Kong, China, 2016.
- [48] A. Said & W. Pearlman, "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, 1996.
- [49] J. Wu, W. Lin, G. Shi, L. Li, & Y. Fang, "Orientation selectivity based visual pattern for reduced-reference image quality assessment," *Information Sciences*, vol. 351, pp. 18-29, 2016.
- [50] L. Zhang, L. Zhang, X. Zhang & D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 13, no. 8, pp. 2378-2386, 2011.
- [51] K. Gu, L. Li, H. Lu, X. Min & W. Lin, "A fast reliable image quality predictor by fusing micro- and macro-structures," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 5, pp. 3903-3912, 2017.
- [52] K. Gu, S. Wang, G. Zhai, W. Lin, X. Yang & W. Zhang, "Analysis of distortion distribution for pooling in image quality prediction," *IEEE Transactions on Broadcasting*, vol. 6, no. 2, pp. 446-456, 2016.
- [53] W. Xue, L. Zhang, X. Mou & A.C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684-695, 2014.

- [54] K. Gu, G. Zhai, X. Yang & W. Zhang, "An efficient color image quality metric with local-tuned-global model," *2014 IEEE International Conference on Image Processing*, pp. 506-510, 2014.
- [55] E. Larson, & D.M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 01-06, 2010.
- [56] L. Krasula, K. Fliegel, P. Le Callet & K. Milos, "On the accuracy of objective image and video quality models: New methodology for performance evaluation," *2016 Eighth International Conference on Quality of Multimedia Experience*, Lisbon, Portugal, 2016.
- [57] L.M. Kaplan, S.D. Burks, R.S. Blum, R.K. Moore & N. Quang, "Analysis of image quality for image fusion via monotonic correlation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 222-235, 2009.
- [58] E. Demirors, G. Sklivanitis, G.E. Santagati, T. Melodia & S. N. Batalama, "A high-rate software-defined underwater acoustic modem with real-time adaptation capabilities," *IEEE Access*, vol. 6, pp. 18602-18615, 2018.



Weiling Chen (Member, IEEE) received the B.S. and Ph.D. degrees in communication engineering from Xiamen University, Xiamen, China, in 2013 and 2018, respectively. She is currently a Lecturer with the College of Physics and Information Engineering, Fuzhou University, China. From Sep. 2016 to Dec. 2016, she was visiting at the School of Computer Science and Engineering, Nanyang Technological University, Singapore. Her current research interests include image quality assessment, image compression, and underwater acoustic communication.



Ke Gu (Member, IEEE) is a Professor with 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 (T-MM), the Best Student Paper Award at the IEEE International Conference on Multimedia and Expo (ICME) in 2016, and the Excellent Ph.D. Thesis Award from the Chinese Institute of Electronics in 2016. He was the Leading Special Session Organizer in 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 for the IEEE ACCESS and IET Image Processing (IET-IPR), and an Area Editor for the Signal Processing Image Communication (SPIC). He is a Reviewer for 20 top SCI journals.



Tiesong Zhao (Senior Member, IEEE) received the B.S. degree in electrical engineering from the University of Science and Technology of China, Hefei, China, in 2006, and the Ph.D. degree in computer science from the City University of Hong Kong, Hong Kong, in 2011. He served as a Research Associate with the Department of Computer Science, City University of Hong Kong (2011-2012), a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo (2012-2013) and a Research Scientist with the Ubiquitous Multimedia Laboratory, The State University of New York at Buffalo (2014-2015). He is currently a Minjiang Distinguished Professor in the College of Physics and Information Engineering, Fuzhou University, China. His research interests include multi-media signal processing, coding, quality assessment and transmission. Dr. Zhao has around 50 publications in these fields. Due to his contributions in video coding and transmission, he received the Fujian Science and Technology Award for Young Scholars in 2017. He has also been serving as an Associate Editor of IET Electronics Letters since 2019.



Gangyi Jiang (Senior Member, IEEE) received the M.S. degree from Hangzhou University, Hangzhou, China, in 1992, and the Ph.D. degree from Ajou University, South Korea, in 2000. He is currently a Professor with the Faculty of Information Science and Engineering, Ningbo University, Ningbo, China. He has published over 100 referred papers in international journals and conferences. His research interests include video compression, multi-view video coding, and visual perception.



Patrick Le Callet (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in image processing from the École Polytechnique de l'Université de Nantes. He was an Assistant Professor and a full-time Lecturer with the Department of Electrical Engineering, Technical Institute, University of Nantes, from 1997 to 1999 and from 1999 to 2003. He led the Image and Video Communication Laboratory, CNRS IRCCyN, from 2006 to 2016. He was one of the five members of the Steering Board of CNRS from 2013 to 2016. Since 2015, he has been the Scientific Director of the Cluster Ouest Industries Creatives, a five-year program gathering over ten institutions (including three universities). Since 2017, he has been one of the seven members of the Steering Board of the CNRS LS2N Laboratory (450 researchers), a Representative of Polytech Nantes. He is mostly involved in research dealing with applications of human vision modeling in image and video processing. His current research interests include quality of experience assessment, visual attention modeling and applications, perceptual video coding, and immersive media processing.