IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 29, 2019, accepted June 16, 2019, date of publication June 20, 2019, date of current version July 10, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2923987

Subset-Guided Consistency Enhancement Assessment Criterion for an Imageset Without Reference

HAO LIU, (Member, IEEE), DONG WEI, AND DEMIN LI
College of Information Science and Technology, Donghua University, Shanghai 201620, China
Engineering Research Center of Digitized Textile & Fashion Technology, Ministry of Education, Donghua University, Shanghai 201620, China
Corresponding author: Hao Liu (liuhao@dhu.edu.cn)

This work was supported by the Natural Science Foundation of Shanghai under Grant 18ZR1400300.

ABSTRACT In a new era for machine vision, some image enhancement algorithms can be used to improve the quality of an imageset without reference. To assess the performance of imageset enhancement, the existing average criterion utilizes a no-reference image quality metric to calculate the quality score of each enhanced image, and quantifies the performance of an enhancement algorithm by the mean value of all scores. If the quality scores of some images fluctuate greatly, their mean value difficultly reflects the degradation or worst cases during imageset enhancement. Therefore, this paper analyzes and illustrates the need and significance of consistency enhancement assessment, and then proposes a subset-guided consistency enhancement assessment criterion for an imageset without reference. By measuring the subset of an imageset, the proposed criterion firstly calculates the difference of quality scores of each image before and after enhancement and then filters the outlier data outside confidence interval, and finally quantifies the consistency enhancement performance of an enhancement algorithm according to its consistency enhancement degree. When a small subset is used to guide its large imageset, the average criterion judges a consistency or non-consistency enhancement algorithm with a 16.7% false identification ratio, and also makes one misjudgment about the optimal-consistency algorithm, while the proposed criterion always correctly judges the non-consistency or optimal-consistency enhancement algorithm. This paper can help the scientific community to select a robust enhancement algorithm in the degradation or worst cases. As compared with the average criterion, the proposed criterion is more robust in terms of subset-guided consistency enhancement assessment, which may effectively find an optimal-consistency or non-consistency enhancement algorithm for the rest of an imageset.

INDEX TERMS Imageset enhancement, no-reference image quality, consistency enhancement assessment, subset guided.

I. INTRODUCTION

With the progress of intelligent image processing and machine vision, there have been some image enhancement algorithms for unmanned surveillance systems, where the performance of an image enhancement algorithm may be measured by a no-reference image quality metric [1], [2]. After obtaining an enhanced image, the enhancement performance for the image is quantified by its quality score. When assessing the quality enhancement of an imageset which contains many images, the existing criterion is the average criterion, where a no-reference image quality metric is used to respectively measure each image in the enhanced imageset, and then calculate the mean value of their quality scores on all images of the imageset. The enhancement algorithm with a higher mean value can be selected in future applications. However, under some special circumstances, the image enhancement algorithm with a higher mean value can't improve all images in an imageset. When the quality score of an initial image is greater than that of its enhanced image in an imageset, it is impossible to prevent this degradation case only by the average criterion. In such applications as unmanned driving and biomedicine, any degradation or worst case can't be ignored, because a fatal accident may occur.

The associate editor coordinating the review of this manuscript and approving it for publication was Ikramullah Lali.

Therefore, when assessing the quality enhancement of an imageset, the quality score of an enhanced image may be fewer than that of its initial image, and this problem has attracted increasing attention. Although the average criterion is good in most situations, it still has its own limitation when choosing an enhancement algorithm for a large imageset.

In the era for machine vision, the amount of image data is increasing quickly, and the enhancement task is gradually oriented to an imageset. For an increasing number of images, an assessment criterion is urgently needed to find an image enhancement algorithm with consistency performance for pattern recognition, intelligent monitoring or other machine vision applications. Based on the basic motivation, this paper proposes a subset-guided consistency enhancement assessment criterion for an imageset without reference. The proposed criterion can provide a performance assessment mechanism for different image enhancement algorithms, and then find an optimal-consistency or non-consistency enhancement algorithm for an imageset.

In the past few years, underwater image enhancement has drawn considerable attention in both image processing and computer vision [3]. Without loss of generality, this paper takes an underwater imageset as an example to expound the proposed assessment criterion. When assessing the imageset enhancement under a prior model-based metric, a lightweight data-driven approach should be added especially for underwater imaging in unknown scenes, where the difference of quality scores of each image before and after enhancement need be fully considered. Theoretically, this difference represents the gain ability of each image enhancement algorithm under the same metric.

The remaining part of this paper is organized as follows: Section II introduces various underwater image enhancement algorithms, and Section III analyzes typical metrics for no-reference image quality evaluation, and Section IV describes the subset-guided consistency enhancement assessment criterion. Through extensive experiments, Section V illustrates the relative advantages of the proposed criterion. Section VI concludes the paper.

II. IMAGE ENHANCEMENT ALGORITHMS

The image enhancement refers to highlighting useful information and removing or weakening useless information according to a specific requirement, which aims to make the enhanced image more suitable for the vision characteristics of human eyes or machine [4]. There are some classical underwater image enhancement algorithms for challenging situations such as dynamic or unknown scenes. For instance, the contrast enhancement is a fundamental enhancement mechanism, which makes the probability density function of image gray level meet the form of approximate uniform distribution so as to increase the dynamic range of an image and improve the image contrast [5]. Typically, the contrast limited adaptive histogram equalization (CLAHE) [6] algorithm is an improvement of the traditional histogram equalization, which can overcome noises by limiting the image contrast. When the underwater imaging environment is assumed to have low-backscatter, shallow-water conditions, the assumptions make it possible to design a simple frequency-domain inverse filter (IF) for image restoration [7]. In the field of color constancy, the dynamic threshold white balance (DTWB) [8] algorithm is also a common image enhancement algorithm. For underwater image enhancement, Voronin *et al.* [9] proposed a hybrid equalization (HE) algorithm by combining multiple physical model-based enhancement methods. Güraksin *et al.* [10] proposed a novel contrast stretching (CS) algorithm especially for underwater image enhancement. Iqbal *et al.* [11] proposed an unsupervised color model (UCM) algorithm for improving low-quality underwater images. Each of these algorithms has its own advantages and disadvantages in a certain application scenario, and it is beyond the scope of this work to comprehensively assess the state-of-the-art underwater image enhancement algorithms. Some assumptions in an enhancement algorithm may fail for a different imageset, and no algorithm has constant superiority due to different properties of underwater imaging and lighting conditions [12]. By introducing the definition of consistency enhancement degree, this paper aims to provide an effective assessment criterion by measuring the subset of an imageset without reference, and find an optimal-consistency or non-consistency enhancement algorithm for the rest of the imageset.

III. NO-REFERENCE IMAGE QUALITY METRICS

In the process of image acquisition, transmission and storage, it is inevitable that image quality will be affected. The image quality metric has been concerned by many researchers. A variety of image quality metrics have also been proposed, and these metrics can be divided into subjective metric and objective metric. The subjective metric mainly relies on the observers to evaluate the images and obtain their mean opinion score. For a large imageset, the subjective metric has some shortcomings such as heavy workload and low efficiency. The objective metric is used to obtain the quantitative quality score of an image. According to whether reference images are needed during the quality evaluation, the objective metrics are divided into three types: full-reference, reduced-reference and no-reference. In this work, the used image quality metrics belong to no-reference type. The no-reference image quality metric is challenging due to the absence of reference image and the ever-changing content of different images to be evaluated.

Based on a prior model, a no-reference image quality metric can measure the image quality without any reference. For a gray image, the common quality metrics utilize its contrast and edge sharpness. For the quality evaluation of a color image, Wang *et al.* [13] put forward a universal mechanism by transforming a color image into gray images and then measuring the image quality, where the process of transforming color image into gray image will cause a certain quality loss. The existing color quality metrics are focused on saturation, brightness, sharpness, contrast, chrominance or other feature representations [14]. Panetta *et al.* [15] proposed a human-visual-system-inspired underwater image quality metric which reflects the human visual perception. For unmanned surveillance systems, a no-reference image quality metric is needed in prior model-based manner.

Due to the complicated underwater environment and lighting conditions, the underwater image enhancement is a challenging task, where an underwater image is degraded by wavelength-dependent absorption, forward scattering and backward scattering [16]. Blasinski *et al.* [17] provided an open-source underwater image simulation tool. However, there still exists a gap between the synthetic underwater images with reference and real-world underwater images without reference. According to Shannon's information theory, the Entropy metric is widely used to represent the information uncertainty and complexity of an image as a no-reference image quality metric [18]. In addition, the underwater color image quality evaluation (UCIQE) is served as a no-reference metric [19], where the measurements of chrominance variation, average saturation, and luminance contrast of an underwater image are linearly combined. The UCIQE metric is widely used to evaluate underwater image quality. In our experiments, the model coefficients are set to the values given in [19]. As long as the non-consistency or consistency performance can be fairly judged under the Entropy or UCIQE metric, the proposed criterion will be valuable for an imageset without reference.

In this paper, "algorithm", "metric" and "criterion" have different meanings. For various imagesets, each algorithm or metric has its strength and limitation. In fact, the proposed criterion doesn't depend on any state-of-the-art enhancement algorithm or no-reference image quality metric. To the best of our knowledge, the existing algorithms or metrics still depend on the average criterion for imageset enhancement assessment [20], [21], but they don't consider consistency enhancement performance. Therefore, the proposed criterion will be mainly compared with the average criterion.

(a) Initial (b) IF (c) DTWB
(d) HE (e) CLAHE (f) CS
(g) UCM

FIGURE 1. Subjective comparison of image enhancement algorithms.

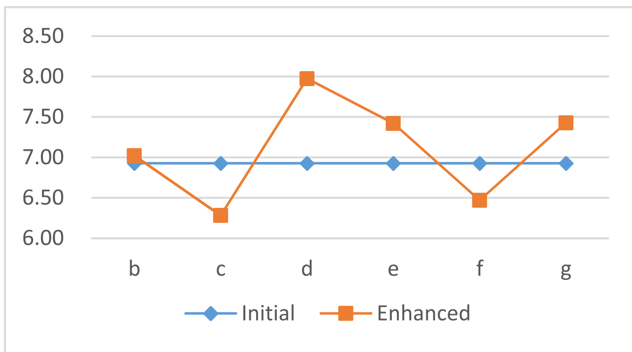


FIGURE 2. Quality scores of the initial image and its enhanced images.

IV. CONSISTENCY ENHANCEMENT ASSESSMENT

A. ANALYSIS OF CONSISTENCY ENHANCEMENT

This section will utilize six simple but effective enhancement algorithms, i.e., the IF algorithm [7], the DTWB algorithm [8], the HE algorithm [9], the CLAHE algorithm [6], the CS algorithm [10] and the UCM algorithm [11] to improve a single underwater image. Based on these image enhancement algorithms, Fig. 1 illustrates the subjective results of an initial image and its enhanced images, whose quality evaluation depends on the human visual perception system. However, some machine vision applications have to depend on no-reference image quality metrics such as Entropy or UCIQE.

Under the Entropy metric, the initial image and its enhanced images are evaluated respectively, and the quality scores are drawn into a line graph, as shown in Fig. 2. It can be seen that the quality scores of two enhanced images are reduced, where the IF, HE, CLAHE and UCM algorithms can enhance the quality of the initial image, but the DTWB and CS algorithms reduce the quality of the initial image.

As shown in Fig. 3, the imageset **A** is composed of six underwater images, and the six images are numbered as: IMG1, IMG2, IMG3, IMG4, IMG5 and IMG6. Here, the imageset **A** is used to illustrate what the consistency enhancement assessment is. Firstly, the Entropy metric is applied to obtain the initial quality score of each initial image in the imageset **A**. Then, the IF, DTWB, HE, CLAHE, CS and UCM algorithms are respectively implemented to enhance the imageset **A**. Finally, the Entropy metric is applied to obtain the new quality score of each enhanced image in the imageset **A**. The evaluation results are shown in Fig. 4 to Fig. 9.

As can be seen from Fig. 4, Fig. 6, Fig. 7 and Fig. 9, the new quality score of each enhanced image is higher than that of

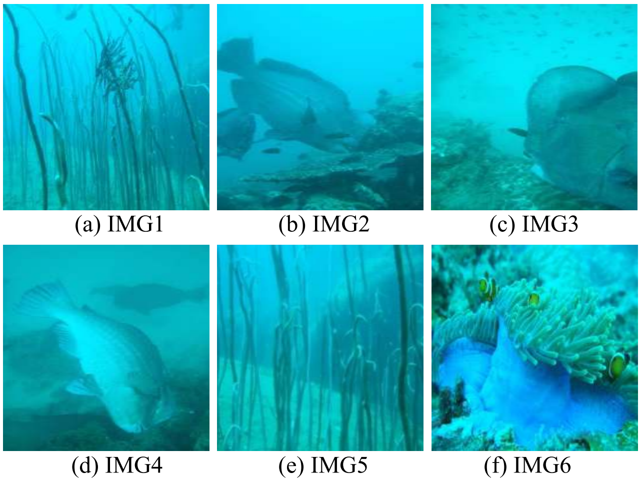


FIGURE 3. The imageset A.

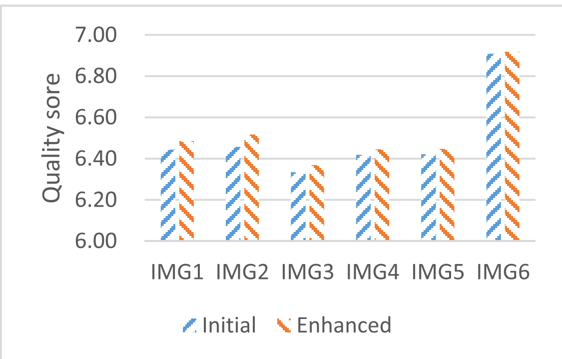


FIGURE 4. Performance of the IF algorithm.

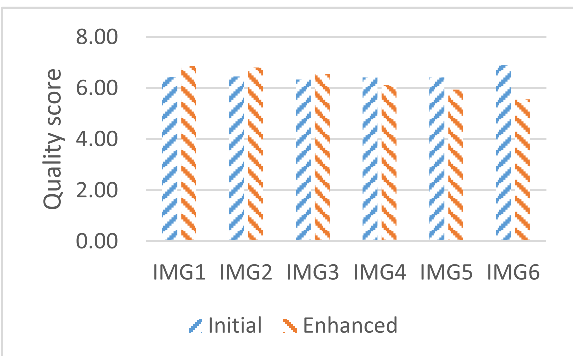


FIGURE 5. Performance of the DTWB algorithm.

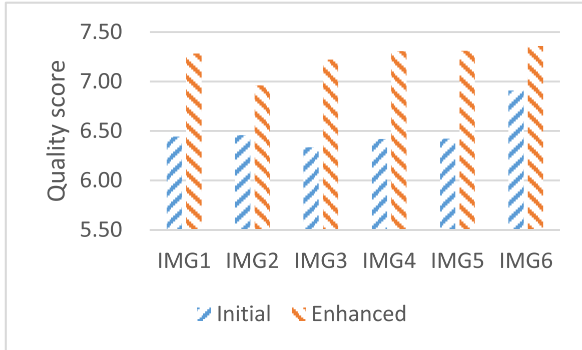


FIGURE 6. Performance of the HE algorithm.

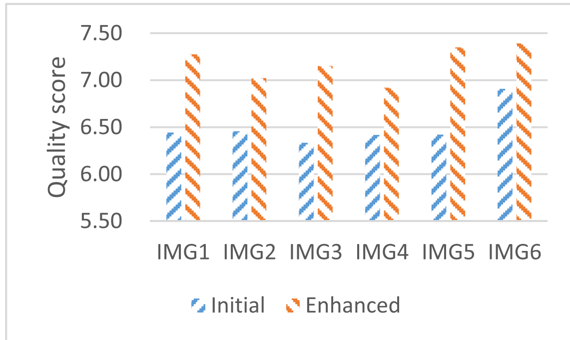


FIGURE 7. Performance of the CLAHE algorithm.

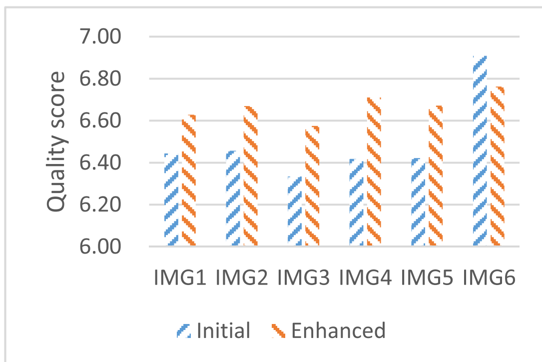


FIGURE 8. Performance of the CS algorithm.

its initial image in the imageset **A**. Therefore, the IF, HE, CLAHE and UCM algorithms are regarded as consistency enhancement. On the contrary, according to the Fig. 5 and Fig. 8, after the imageset **A** is enhanced by the DTWB algorithm or CS algorithm, there are a few cases where the new quality score of an enhanced image is lower than that of the initial image, and thus the algorithm is regarded as non-consistency enhancement (N/CE) for an imageset. Among the four consistency enhancement algorithms, the HE algorithm has relatively significant enhancement effect on the small subset. In a similar application scenario, HE is usually regarded as the optimal-consistency enhancement algorithm for a large imageset. In the following, the consistency performance will be judged and quantified.

B. THE PROPOSED CRITERION

After defining the consistency enhancement algorithm through experimental data, a subset-guided consistency enhancement assessment (SCEA) criterion is proposed for an imageset without reference. The general process of the proposed SCEA criterion is outlined. Firstly, the proposed criterion calculates the difference of quality scores of

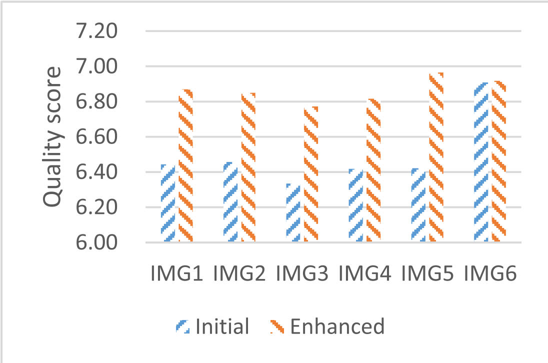


FIGURE 9. Performance of the UCM algorithm.

each image before and after enhancement for an imageset. Secondly, the criterion parameters are determined according to a certain application scenario and the confidence interval is used to filter the outlier data [22]. Then, the consistency or non-consistency of an enhancement algorithm is judged according to the valid data. Finally, the proposed criterion utilizes the consistency enhancement degree (D_{ce}) to quantify the consistency performance of a consistency enhancement algorithm. The flow chart of the proposed SCEA criterion is shown in Fig. 10, and the detailed steps of SCEA are explained as follows:

Step 1: The proposed criterion uses a no-reference image quality metric Q to evaluate the quality scores on all images ($I_1, I_2, \ldots, I_n$) in an initial imageset which usually is a small subset of an imageset without reference, and obtains the initial quality score α_i of an initial image I_i as one parameter in the SCEA equations, where $i(i = 1, 2, \ldots n)$ is the image number, and n is the amount of all images in the initial imageset.

Step 2: The image quality enhancement is performed on all images in the initial imageset by using an image enhancement algorithm E, so as to obtain an enhanced imageset ($I_1^{'}, I_2^{'}, \ldots, I_n^{'}$).

Step 3: The image quality is respectively evaluated on each enhanced images $I_i^{'}$ by using the image quality metric Q of Step 1, and thus the new quality score β_i is obtained. If β_i is greater than the initial quality score α_i, it indicates that under the image quality metric Q, the image enhancement algorithm E improves the quality of initial image I_i; on the contrary, If β_i is fewer than α_i, it indicates that under the image quality metric Q, the enhancement algorithm E reduces the quality of initial image I_i.

Step 4: The quality score difference (QSD_i) of initial image I_i and enhanced images $I_i^{'}$ is obtained from the difference between the initial quality score α_i and new quality score β_i. If the value of QSD_i is positive, it means that under the image quality metric Q, the image enhancement algorithm E improves the quality of initial image I_i.

$$QSD_i = \beta_i - \alpha_i \tag{1}$$

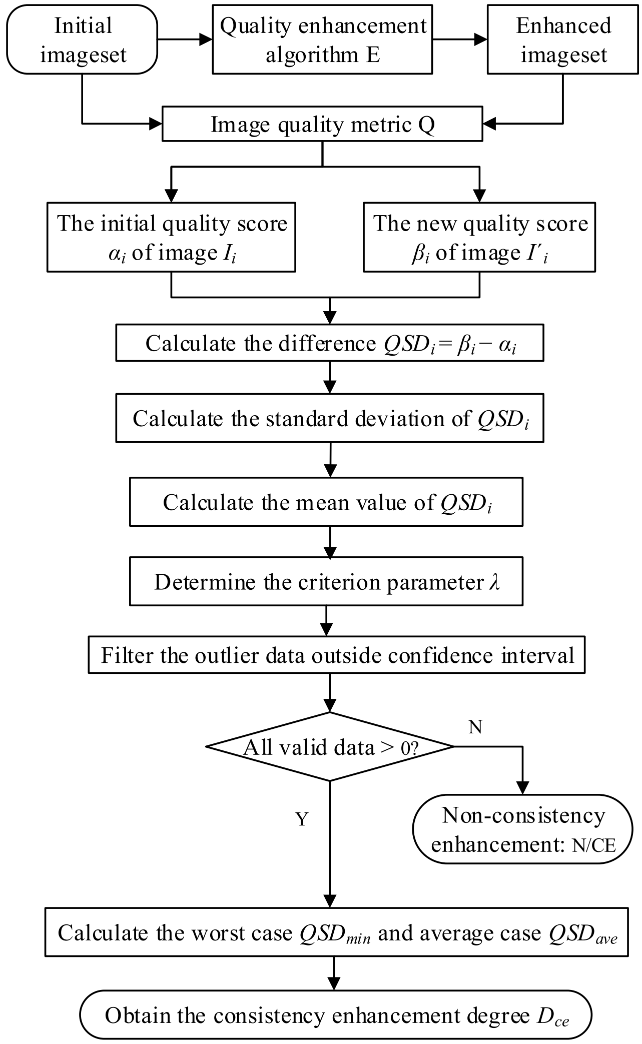


FIGURE 10. Flow chart of the proposed SCEA criterion.

Step 5: The proposed criterion calculates the mean value U of the quality score differences of all images. The equation is as follows:

$$U = \frac{1}{n} \sum\nolimits_{i=1}^{n} QSD_i \tag{2}$$

Then, it calculates the standard deviation S of QSD_i:

$$S = \frac{1}{n} \sqrt{\sum\nolimits_{i=1}^{n} (QSD_i - U)^2} \tag{3}$$

Step 6: The proposed criterion determines the parameter λ according to the requirements of an application scenario, and substitutes the parameter into the following equation:

$$\varepsilon = \lambda \cdot S \tag{4}$$

The proposed criterion calculates the parameter ε, and gets the confidence interval $[U - \varepsilon, U + \varepsilon]$, and then uses the confidence interval to filter the outlier data. The QSD_i values outside confidence interval are regarded as outlier data for the imageset. The proposed criterion will discard them, and keep the QSD_i values in confidence interval as valid data.

For example, with $\lambda = 1.96$, $[U-\varepsilon, U+\varepsilon]$ is a 95% confidence interval so as to remove few error data.

Step 7: Sort the valid data in ascending order by the QSD_i subscript number, and sort it into $QSD_1, QSD_2, \ldots, QSD_m$, where $j(j = 1, 2, \ldots, m)$ is the new image subscript, and $m(m \leq n)$ is the amount of valid data.

Step 8: Compare the valid data with zero. If all valid data are greater than zero, it indicates that under the image quality metric Q, the image enhancement algorithm E is the consistency enhancement algorithm, then the flow continues to Step 9; Otherwise, the image enhancement algorithm E is N/CE, where the degradation case occurs.

Step 9: Based on the above valid data, the minimum quality score (QSD_{min}) of QSD_j is obtained to represent the worst case:

$$QSD_{min} = min\left\{QSD_1, QSD_2, \ldots, QSD_m\right\} \tag{5}$$

Then, the mean quality score (QSD_{ave}) of QSD_j is also calculated as the average case of valid data:

$$QSD_{ave} = ave\left\{QSD_1, QSD_2, \ldots, QSD_m\right\} \tag{6}$$

In general, the consistency performance with high QSD_{ave} but low QSD_{min}, or low QSD_{ave} but high QSD_{min} is unsatisfactory in such applications as marine automation and aquatic robots. For risk-sensitive assessment, the mean value or standard deviation difficultly reflects the degradation or worst cases during imageset enhancement. To ensure both gain fairness and worst-case resilience, QSD_{ave} and QSD_{min} have the similar importance in the mathematical foundation.

Step 10: To leverage both worst case and average case, Equation (7) provides a case-combined solution as a weighted contribution of minimum and average differences. As a linear combination function of both QSD_{ave} and QSD_{min}, the consistency enhancement degree D_{ce} is expressed as follows:

$$D_{ce} = (1-\mu) \cdot QSD_{ave} + \mu \cdot QSD_{min} \tag{7}$$

where a weighting factor $\mu \in [0, 1]$ indicates the relative importance of the QSD_{min} component which depends on a specific application scenario. Although the adaptive selection of μ is very challenging, it is practical that μ is set to a fixed value [23], [24]. Further, since QSD_{ave} and QSD_{min} have a similar order of magnitude in quality score difference, "$\mu = 0.5$" is regarded as a simple but reasonable tradeoff point between QSD_{ave} and QSD_{min}. The reason behind (7) is that the index D_{ce} expects to strike a balance between gain fairness and worst-case resilience, which helps the scientific community to adaptively select a robust enhancement algorithm. In the following, one typical scenario is needed to prove that the average criterion is possibly biased when choosing an enhancement algorithm, while the proposed criterion can avoid the extreme cases. According to (7), the proposed criterion can obtain the consistency enhancement degree D_{ce} of the consistency enhancement algorithm E under the image quality metric Q.

Based on the subset of an imageset, these consistency enhancement algorithms can be screened out from various image enhancement algorithms through above SCEA steps. A higher D_{ce} value means that its enhancement algorithm is more robust in terms of consistency enhancement performance. With the highest D_{ce} value, the optimal-consistency enhancement algorithm can be selected for the rest of an imageset.

TABLE 1. The mean quality score of the initial imageset A and enhanced imageset A (under the Entropy metric).

Imageset **A**	Mean quality score
Initial	6.4150
IF Enhanced	6.5304
DTWB Enhanced	6.3056
HE Enhanced	7.2401
CLAHE Enhanced	7.1856
CS Enhanced	6.6696
UCM Enhanced	6.8651

V. EXPERIMENTAL RESULTS

A. COMPARISON BETWEEN SCEA CRITERION AND AVERAGE CRITERION

When an image quality metric is always giving better or worse scores for an enhancement algorithm, the existing average criterion can provide an assessment result similar to that of the proposed SCEA criterion. In contrast with other data-driven approaches, the SCEA criterion is still computationally lightweight [25], [26]. When N/CE is unbearable in some occasional cases, the SCEA criterion has its own unique advantages by assessing the consistency performance of an enhancement algorithm. The SCEA criterion is not dependent on any enhancement algorithm. In the following, the IF, DTWB, HE, CLAHE, CS and UCM algorithms are respectively applied to enhance different imagesets.

Under the Entropy metric, the mean quality scores of the initial imageset **A** and enhanced imageset **A** are respectively calculated, and the experimental results are given in Table 1. The enhanced imageset **A** is successively obtained by six enhancement algorithms. As can be seen from Table 1, the mean quality score of the enhanced imageset **A** by the IF, HE, CLAHE, CS or UCM algorithm is higher than that of the initial imageset **A**, where the CS algorithm is relatively better than the IF algorithm for the imageset **A**. However, in Fig. 8, the CS algorithm reduces the quality score of IMG6, which is N/CE; IF is a consistency enhancement algorithm. Therefore, the enhancement algorithm with a higher mean quality score is not necessarily a consistency enhancement algorithm. There is a potential problem to rank the enhancement algorithm only by the average criterion. If the CS algorithm is applied to the rest of an imageset, it may generate the serious degradation for individual images.

Under the Entropy metric, the SCEA criterion is respectively applied to calculate the D_{ce} value of the imageset **A** by six enhancement algorithms. The experiment adopts the criterion parameters $\mu = 0.5$ and $\lambda = 2.0$ with the

TABLE 2. The D_{ce} value of imageset A (under the Entropy metric).

Enhancement algorithms	D_{ce} value
IF	0.0220
DTWB	N/CE
HE	0.5964
CLAHE	0.5862
CS	N/CE
UCM	0.4163

97% confidence interval. The results are shown in Table 2. It can be found from Table 2 that IF, HE, CLAHE and UCM are the consistency enhancement algorithms. Among them, the D_{ce} value of the HE algorithm is the highest, which is regarded as the optimal-consistency enhancement algorithm for the imageset **A**. On the contrary, the DTWB and CS algorithms are N/CE. The experimental results are consistent with the conclusion in Section IV. In contrast with the average criterion, the SCEA criterion can more robustly assess the consistency performance of each image enhancement algorithm in the application scenario.

B. THE ROBUSTNESS OF SCEA

When providing an enhancement algorithm for an imageset, due to the huge image quantity and abundant enhancement algorithms, a small subset may be extracted from a large imageset in order to reduce cost. Some enhancement algorithms are applied to enhance the quality of the subset. The enhancement algorithm with strong consistency performance is judged and selected, which will be applied to the rest of the imageset.

Although being available in most situations, the average criterion may lead to potential problems in a few cases, and the following experiment indicates that such problems may happen. To validate that the proposed SCEA criterion is more competitive rather than the average criterion in a certain application scenario, we randomly select 200 images as the imageset **D** from a large-scale underwater imageset [27]. Then, we randomly select 50 images as the imageset **C** from the imageset **D**, and randomly select 5 images as the subset **B** from the imageset **C**. To verify the subset-guided enhancement, experiments are respectively performed on subset **B**, imageset **C** and imageset **D**.

Under the UCIQE metric, the mean quality scores of the initial subset **B** and enhanced subset **B** are given in Table 3. According to the quantitative results, different enhancement algorithms can be ranked. As can be seen from Table 3, the mean quality scores of the enhanced subset **B** by the DTWB, HE, CLAHE and UCM algorithms are higher than that of the initial subset **B**. According to the average criterion on subset **B**, the DTWB, HE, CLAHE and UCM algorithms have consistency enhancement performance, where CLAHE is the optimal-consistency enhancement algorithm among the six algorithms. In a subset-guided framework, DTWB, HE,

TABLE 3. The mean quality score of subset B, imageset C and imageset D (under the UCIQE metric).

	Mean quality score		
	Subset **B**	Imageset **C**	Imageset **D**
Initial	0.4550	0.4697	0.4640
IF Enhanced	0.4530	0.4670	0.4615
DTWB Enhanced	0.4566	0.4404	0.4147
HE Enhanced	0.5553	**0.5596**	**0.5591**
CLAHE Enhanced	**0.5570**	0.5545	0.5545
CS Enhanced	0.4472	0.4612	0.4554
UCM Enhanced	0.5288	0.5331	0.5312

CLALE and UCM are considered as consistency enhancement algorithms on imageset **C** and imageset **D**, where CLAHE is optimal.

In the verification experiment, the mean quality scores of the initial imageset **C**, initial imageset **D** and enhanced imageset **C**, enhanced imageset **D** are also given in Table 3. With the expansion of the imageset, when the DTWB algorithm is applied to enhance the imageset **C** and imageset **D**, the mean quality score of the enhanced imageset **C** and enhanced imageset **D** is fewer than the initial imageset **C** and initial imageset **D**. It means that the DTWB algorithm is N/CE on imageset **C** and imageset **D**, which is different from the assessment result on subset **B**. Therefore, the average criterion gives one false identification among the six enhancement algorithms. When using the subset **B** to guide the imageset **C** and imageset **D**, the average criterion judges a consistency or non-consistency enhancement algorithm with a 16.7% false identification ratio among the six algorithms. The reason for this phenomenon is that the DTWB algorithm is N/CE. The N/CE algorithm may produce serious degradation when applying to a large imageset. In addition, HE is the optimal-consistency enhancement algorithm on imageset **C** and imageset **D**, which is inconsistent with the assessment result on subset **B**. The average criterion also makes one misjudgment about the optimal-consistency enhancement algorithm on the three imagesets. In summary, when a subset-guided algorithm is used for a large imageset, the existing average criterion may lead to the false identification or misjudgment.

The following will verify whether the SCEA criterion is more robust than the average criterion when a subset is used to guide its imageset. Under the UCIQE metric, with $\lambda = 2.0$ and $\mu = 0.5$, the SCEA criterion successively calculates the D_{ce} value on subset **B** by each enhancement algorithm, and the results are shown in Table 4. It can be seen that IF, DTWB and CS are N/CE, while HE, CLAHE and UCM are consistency enhancement algorithms where HE is optimal. In a subset-guided framework, the IF, DTWB and CS algorithms are considered as N/CE on imageset **C** and imageset **D**; HE, CLAHE and UCM are considered as consistency enhancement algorithms on imageset **C** and imageset **D**, where HE is optimal. In the verification experiment, the SCEA criterion

TABLE 4. The D_{ce} value of subset B, imageset C and imageset D (under the UCIQE metric).

Enhancement algorithms	D_{ce} value Subset **B**	Imageset **C**	Imageset **D**
IF	N/CE	N/CE	N/CE
DTWB	N/CE	N/CE	N/CE
HE	**0.0690**	**0.0596**	**0.0643**
CLAHE	0.0678	0.0576	0.0633
CS	N/CE	N/CE	N/CE
UCM	0.0454	0.0374	0.0401

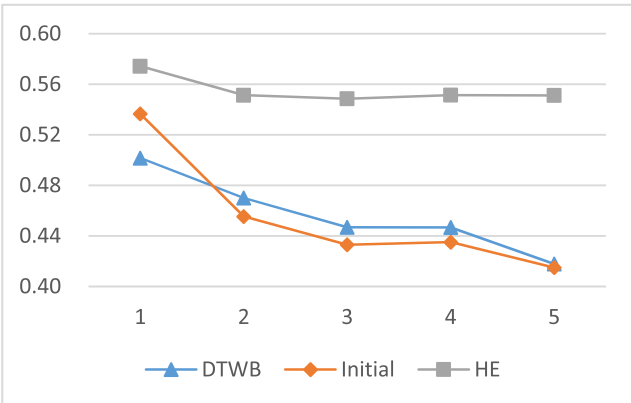


FIGURE 11. The quality score of subset B (under the UCIQE metric).

successively calculates the D_{ce} value on the imageset **C** and imageset **D** by each enhancement algorithm. The actual results are also shown in Table 4, where the IF, DTWB and CS algorithms are always N/CE, while HE, CLAHE and UCM are always consistency enhancement algorithms, and HE is always optimal among the six algorithms. The assessment results on imageset **C** and imageset **D** are consistent with the assessment result on subset **B**. It means that when performing the subset-guided enhancement for a large imageset, the proposed SCEA criterion can correctly judge each consistency or non-consistency enhancement algorithm, and the selection of an optimal-consistency enhancement algorithm is also enough robust.

As shown in Fig. 11 to Fig. 13, we take the DTWB algorithm and HE algorithm as typical examples, and draw the quality score graph of one-by-one images in initial subset **B**, initial imageset **C**, initial imageset **D** and enhanced subset **B**, enhanced imageset **C**, enhanced imageset **D** respectively by the two algorithms. According to Fig. 11 to Fig. 13, the DTWB algorithm almost enhances the subset **B**. However, because the DTWB algorithm is N/CE, for larger imageset **C** and imageset **D**, the performance of the DTWB algorithm is poor. This is consistent with the results in Table 3. In addition, with the expansion of the imageset, the DTWB algorithm is always N/CE, while HE is always the consistency enhancement algorithm whose consistency enhancement performance is better than that of the DTWB algorithm.

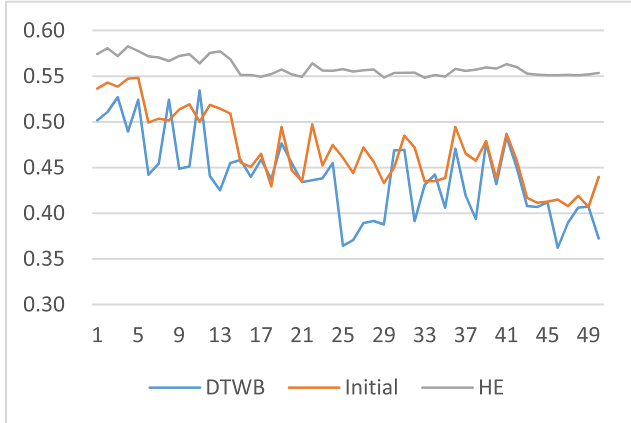


FIGURE 12. The quality score of imageset C (under the UCIQE metric).

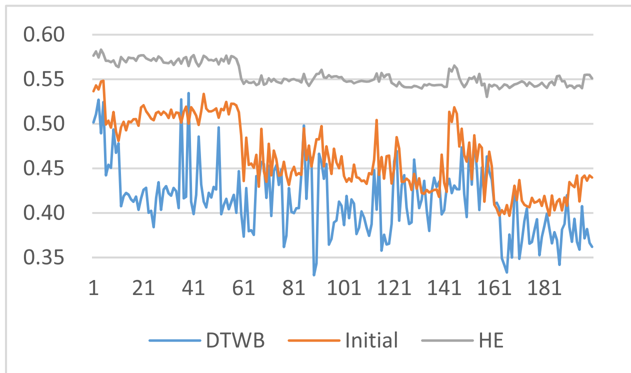


FIGURE 13. The quality score of imageset D (under the UCIQE metric).

This is also consistent with the results in Table 4. With a highest D_{ce} value on subset **B**, the optimal-consistency enhancement algorithm can be correctly selected for imageset **C** or imageset **D**, which indicates during imageset enhancement assessment, the proposed SCEA criterion is more robust than the existing average criterion when using a subset to guide its large imageset.

C. COMPLEXITY ANALYSIS

The computational complexity is analyzed through comparing the execution time between the proposed SCEA criterion and existing average criterion. Based on Matlab R2016a, the complexity test is conducted on a 64-bit PC with Intel Core i3-8100 CPU @3.60 GHz and 4 GB RAM. The initial subset **B** is sequentially enhanced by the IF, DTWB, HE, CLAHE, CS and UCM algorithms so as to obtain the enhanced subset **B**. The average criterion and SCEA criterion are respectively used to assess the consistency enhancement performance of each algorithm. The computational results are presented in Table 5. As evidenced by the results in Table 5, the execution time of the SCEA criterion is nearly twice that of the average criterion. There are three general explanations as follows. Firstly, each enhanced image usually enriches its color and contrast information, so the complexity to evaluate

TABLE 5. Complexity comparison between the average criterion and SCEA criterion on subset B (Seconds).

Enhancement algorithms	Average criterion	SCEA criterion
IF	10.1019	19.5540
DTWB	9.9755	18.2683
HE	10.2324	18.7670
CLAHE	10.6320	18.2794
CS	10.4197	19.1720
UCM	10.3292	19.4102

an enhanced image is slightly greater than that to evaluate its initial image. Secondly, the SCEA criterion performs the image quality metric twice to evaluate an image before and after enhancement, while the average criterion only performs that once. Thirdly, the complexity of the SCEA or average criterion is mainly in the operation of performing a metric. Given an imageset with image quantity N, the computational complexity of the average criterion is $O(N)$, while that of the SCEA criterion is nearly $O(2N)$. Overall, the complexity of the SCEA criterion and average criterion is still in the same order of magnitude.

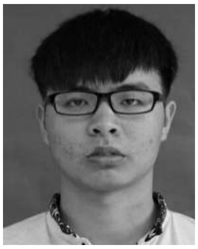
VI. CONCLUSION

To assess the performance of imageset enhancement, this paper proposes a subset-guided consistency enhancement assessment criterion for different image enhancement algorithms. By leveraging both prior model-based metric and data-driven approach, the proposed criterion is more robust than the existing average criterion when choosing a consistency enhancement algorithm for the rest of an imageset without reference. The proposed criterion can be a useful supplement to the average criterion during imageset enhancement, especially in high-risk challenging situations. In the future work, we tend to construct a publicly available underwater imageset with the degradation or worst cases and the corresponding benchmark. Moreover, more intelligent assessment criterion will be investigated for various enhancement algorithms by combining prior model with dynamic learning.

REFERENCES

[1] M. Han, Z. Lyu, T. Qiu, and M. Xu, "A review on intelligence dehazing and color restoration for underwater images," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published.
[2] R. Cui, L. Chen, C. Yang, and M. Chen, "Extended state observer-based integral sliding mode control for an underwater robot with unknown disturbances and uncertain nonlinearities," *IEEE Trans. Ind. Electron.*, vol. 64, no. 8, pp. 6785–6795, Aug. 2017.
[3] J. S. Jaffe, "Underwater optical imaging: The past, the present, and the prospects," *IEEE J. Ocean. Eng.*, vol. 40, no. 3, pp. 683–700, Jul. 2015.
[4] Y. Xu, J. Wen, L. Fei, and Z. Zhang, "Review of video and image defogging algorithms and related studies on image restoration and enhancement," *IEEE Access*, vol. 4, pp. 165–188, 2016.
[5] H. Lu, Y. Li, L. Zhang, and S. Serikawa, "Contrast enhancement for images in turbid water," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 32, no. 5, pp. 886–893, 2015.
[6] M. S. Hitam, E. A. Awalludin, W. N. J. H. W. Yussof, and Z. Bachok, "Mixture contrast limited adaptive histogram equalization for underwater image enhancement," in *Proc. Int. Conf. Comput. Appl. Technol.*, Sousse, Tunisia, Jan. 2013, pp. 1–5.
[7] E. Trucco and A. T. Olmos-Antillon, "Self-tuning underwater image restoration," *IEEE J. Ocean. Eng.*, vol. 31, no. 2, pp. 511–519, Apr. 2006.
[8] C. Li and X. Zhang, "Underwater image restoration based on improved background light estimation and automatic white balance," in *Proc. 11th Int. Congr. Image Signal Process., Biomed. Eng. Informat.*, Beijing, China, Oct. 2018, pp. 1–5.
[9] V. Voronin, E. Semenishchev, S. Tokareva, A. Zelenskiy, and S. Agaian, "Underwater image enhancement algorithm based on logarithmic transform histogram matching with spatial equalization," in *Proc. 14th IEEE Int. Conf. Signal Process.*, Beijing, China, Aug. 2018, pp. 434–438.
[10] G. E. Güraksin, U. Köse, and Ö. Deperlioğlu, "Underwater image enhancement based on contrast adjustment via differential evolution algorithm," in *Proc. Int. Symp. Innov. Intell. Syst. Appl.*, Sinaia, Romania, Aug. 2016, pp. 1–5.
[11] K. Iqbal, M. Odetayo, A. James, R. A. Salam, and A. Z. H. Talib, "Enhancing the low quality images using unsupervised colour correction method," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Istanbul, Turkey, Oct. 2010, pp. 1703–1709.
[12] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," 2019, *arXiv:1901.05495*. [Online]. Available: https://arxiv.org/abs/1901.05495
[13] 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.
[14] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, Jan. 2019.
[15] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE J. Ocean. Eng.*, vol. 41, no. 3, pp. 541–551, Jul. 2016.
[16] D. Akkaynak and T. Treibitz, "A revised underwater image formation model," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Rec. (CVPR)*, Jun. 2017, pp. 6723–6732.
[17] H. Blasinski, T. Lian, and E. Farrell, "Underwater image systems simulation," in *Proc. Imag. Appl. Opt.*, Jun. 2017, pp. 1–2.
[18] S. C. Tan, S. Wang, X. Zhang, S. Wang, S. Wang, S. Ma, and W. Gao, "Visual information evaluation with entropy of primitive," *IEEE Access*, vol. 6, pp. 31750–31758, 2018.
[19] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 6062–6071, Dec. 2015.
[20] M. Yang, A. Sowmya, Z. Wei, and B. Zheng, "Offshore underwater image restoration using reflection-decomposition-based transmission map estimation," *IEEE J. Ocean. Eng.*, to be published.
[21] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, "Real-world underwater enhancement: Challenges, benchmarks, and solutions," 2019, *arXiv:1901.05320*. [Online]. Available: https://arxiv.org/abs/1901.05320
[22] A. Wunderlich, F. Noo, B. D. Gallas, and M. E. Heilbrun, "Exact confidence intervals for channelized Hotelling observer performance in image quality studies," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 453–464, Feb. 2015.
[23] L. Sha, W. Wu, and B. Li, "Novel image set compression algorithm using rate-distortion optimized multiple reference image selection," *IEEE Access*, vol. 6, pp. 66903–66913, 2018.
[24] D. Akkaynak, T. Treibitz, T. Shlesinger, Y. Loya, R. Tamir, and D. Iluz, "What is the space of attenuation coefficients in underwater computer vision," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Rec. (CVPR)*, Jul. 2017, pp. 568–577.
[25] J. Li, K. A. Skinner, R. M. Eustice, and M. Johnson-Roberson, "WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images," *IEEE Robot. Autom. Lett.*, vol. 3, no. 1, pp. 387–394, Jan. 2018.
[26] J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 2049–2062, Apr. 2018.
[27] M. W. Jian, Q. Qi, J. Y. Dong, Y. Yin, W. Zhang, and K.-M. Lam, "The OUC-vision large-scale underwater image database," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Hong Kong, Jul. 2017, pp. 1297–1302.

HAO LIU (M'11) received the B.S. degree in image communication from the Nanjing University of Posts & Telecommunications, Jiangsu, China, in 2000, and the Ph.D. degree in signal and information processing from Shanghai Jiao Tong University, Shanghai, China, in 2006. Since 2009, he has been with the Faculty of the College of Information Science and Technology, Donghua University, where he is currently an Associate Professor. He has published more than 30 papers with Scientific Citation Index. His research interests include image processing and communication.



DONG WEI received the B.S. degree in information engineering from the Hubei University of Science and Technology, in 2017. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering, College of Information Science and Technology, Donghua University, Shanghai, China. His research interests include image quality evaluation and video coding.

DEMIN LI received the Ph.D. degree in electronic and computer engineering from the Nanjing University of Science and Technology, China, in 1998. From 2002 to 2005, he was a Research Scientist with Nanjing University, China. He is currently a Professor with the Department of Telecommunication and Electronic Engineering, College of Information Science and Technology, Donghua University, Shanghai, China. He has authored more than 60 research papers in journals and conferences. His recent research interests include telecommunication system engineering, wireless mobile networking, mobile decision theory, and mobile decision support systems. He is currently as an Associate Chairman of the Circuits and Systems Committee, Shanghai.