Blind Quality Assessment of Tone-Mapped Images Based on Visual-Processing Features

DONGHUI WAN\textsuperscript{1,2}, XIUHUA JIANG\textsuperscript{1,3}, CHENG GUO\textsuperscript{1,3}, AND QING SHEN\textsuperscript{2}

\textsuperscript{1}School of Information and Communication Engineering, Communication University of China, Beijing 100024, China
\textsuperscript{2}School of Electronic Information, Huzhou College, Huzhou 313000, China
\textsuperscript{3}Peng Cheng Laboratory, Shenzhen 518000, China

Corresponding author: Xiuhua Jiang (jiangxiuhua@cuc.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61802123; and in part by the Primary Research and Development Plan of Zhejiang Province, China, under Grant 2020C01097.

\section*{ABSTRACT}
Albeit high dynamic range (HDR) images contain way richer information than low dynamic range (LDR) images, they need to be tone mapped for the visualization on traditional display devices. The community has developed diverse tone-mapping operators (TMOs), and they produce images with varying qualities. Therefore, developing effective evaluation methods consistent with human visual perception is an urgent necessity for determining the best TMO in specific scenarios, or accordingly optimizing the parameters of a certain TMO. Towards this end, we propose a new blind quality assessment model by simulating the mechanism of human visual processing (HVS), which can adaptively adjust the sensitivity according to the chromatic properties of scenes at the beginning and well represent the perception process. Specifically, we first obtain all the retinal response maps of four visual sensitivities from three opponent color channels. After that, gradient similarities in each color channel are calculated as global features to represent the procedure of visual experience, and local mean values and standard deviations are computed in the brightest and darkest regions of the maps to measure the distortions caused by over- and under-exposure. Meanwhile, the maps’ natural scene statistics (NSS) are utilized as the third set of features. Finally, all these features are blended for quality assessment by support vector regression (SVR). Extensive experiments on two public benchmark databases show our method correlates highly with subjective perception and outperforms other state-of-the-art quality assessment methods.

\section*{INDEX TERMS}
Quality assessment, high dynamic range, tone-mapped images, no reference, visual processing.

\section*{I. INTRODUCTION}
High dynamic range (HDR) technologies provide advances in capturing and displaying the real-world scene radiance ranging from direct sunlight to faint starlight [1]. Compared with low dynamic range (LDR) images, HDR images are able to preserve more visual information and details. Though HDR images and videos have gained increasing popularity and hardware support over the past decades, the majority of current display devices are still in LDR, and even some existing consumer HDR displays don’t support the full range of HDR content. Hence, HDR content needs to be tone-mapped to match the range of display devices for visualization. During tone-mapping, the visual quality of HDR images is degraded meanwhile information is lost, due to the compression of the dynamic range.

A well-designed tone-mapping operator (TMO) is supposed to preserve as much visual information from HDR and produce LDR images with excellent perceptual experience. A large number of TMOs with different emphasis have been developed in the last decades [2]. Since the content and structure contained in HDR images are largely-diversified, a TMO behaving well on one HDR image may not be good on another. Hence, it’s critical to choose a proper
TMO or optimize a current TMO based on its tone-mapping performance, which is usually assessed through tone-mapped images (TMIs).

In the field of image quality assessment (IQA), subjective evaluation is the most accurate, but time-consuming and expensive for setting a standard viewing environment and recruiting graders. Such a dilemma also exists in TMIs quality assessment, therefore, developing an objective metric for TMIs is of significant importance.

Many objective IQA methods have been proposed [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], and they are generally classified into full reference (FR), reduced-reference (RR) and no-reference (NR) methods. Typical FR methods assumed that reference and test images were in the same dynamic range, and they extracted features from both reference and test images to calculate their similarity for quality evaluation. Yet, for our task i.e., TMIs quality assessment where to-be-evaluated tone-mapped LDR images have largely lower dynamic ranges than HDR images, it’s difficult to directly make a fair comparison between LDR and HDR. To this end, the tone-mapped quality index (TMQI) [18] used a modified SSIM [3] where the original luminance term (using mean value $\mu$) is deprecated to enable the structure similarity calculation between different dynamic ranges, and a natural scene statistics (NSS) term was added to polish the method. Then, combined with TMQI, the feature similarity index for tone-mapped images (FSITM) [19] compared the local phase similarity of original HDR and tone-mapped LDR images. Instead of simply uniform averaging, different perceptual methods [20], [21] have been used to pool the local qualities and boosted the performance of TMQI. Lastly, TMQI-II [22] achieved a better correlation with subjective scores by new configurations in structure similarity and NSS calculation.

Above methods are all FR. However, since large-data HDR reference images are still scarce, meanwhile FR methods require extra bandwidth to transfer bulky HDR reference counterparts (usually 32-bit quantized), it’s necessary to design a NR method which requires only tone-mapped images.

Traditional NR methods [8], mostly designed for evaluating synthesized distortions, adopted common features related to natural images and usually underperformed on TMIs quality assessment since they didn’t consider the specific distortion type. It’s necessary to assess TMIs quality by paying attention to distinct distortions, including structure corruption, abnormal exposure and colorfulness fade, etc.

Some NR TMIs metrics have started doing so, for example, Gu et al. [23] recognized that a good TMI should well preserve the original information, structure, and naturalness, hence they utilized the entropy values of nine intermediate images generated by darkening/brightening the TMI as their features. They also measured how much structure was preserved in the TMI by calculating the mean value of the gradients above a small threshold, and adopted natural statistics similar to [18]. Apart from this, Kundu et al. [24] proposed a NR approach based on standard space-domain NSS features and HDR-specific gradient-based features. Yue et al. [28], [29] extracted multiple quality-sensitive features in terms of colorfulness, exposure, and naturalness to blindly predict TMIs’ scores. Wang et al. [30] exploited local degradation characteristics and global statistical properties to extract features. The attributes of an image i.e., texture, structure, colorfulness, and naturalness, were considered locally or globally. Yue et al. [31] simulated the responses of single-opponent (SO) and double-opponent cells (DO), and obtained textual information and structural information from the SO and DO responses.

Albeit the success of the above blind (NR) TMIs assessment metrics, their ideas and performance still need to be ameliorated. For example, human beings are able to view and make proper assessment on images through human visual system (HVS). Consequently, exactly simulating HVS can be an effective way to improve TMIs assessment method. Currently, almost all metrics mentioned above extracted quality-related features from images directly, neglecting the visual processing of images. Yue et al. [31] obtained features from SO and DO responses by simulating HVS chromatic behaviors and achieved an improved performance. However, they merely employed ultimate responses and didn’t regard that HVS will perceive an image with the adaptation procedure, which is specific for information acquisition [39]. To this end, our work aims to obtain multiple response images of different visual sensitivities and extract comprehensive features for TMIs quality evaluation by simulating the chromatic procedure of HVS.

Our main ideas are listed as follows:

1) Since the retina of human eyes can adaptively adjust the sensitivity according to image content in the first few seconds to best obtain useful information, our method obtains the retinal responses of three opponent color channels in four sensitivity scales from low to high. These retinal responses serve as the cognitive process of HVS when viewing an image.

2) The responses of each opponent color channel change as the sensitivity increases. These changes should not be same for the images of different qualities, and are useful for quality evaluation, hence we calculate gradient similarities to represent the changes in viewing responses.

3) We also find that the local areas of TMIs may appear over- or under-exposed since the dynamic range might be inappropriately compressed by a specific TMO. Therefore, we segment TMIs into three regions, and extract features from over- and under-exposed local regions respectively.

4) HVS is highly adapted to natural environment after millions of years of evolution. Hence, we also use the natural scene statistics (NSS) of response maps to enhance the performance of our method.

The remainder of this paper is organized as follows. Section II detailly describes how above ideas are implemented in the
The proposed method. Then, in Section III, the edge of our method over competitors is provided by experiments on two publicly available databases (TMID [18] and ESPLIVE [24]), and multiple ablation studies are also carried out. Finally, conclusions are drawn in section IV.

II. PROPOSED METHOD

The flowchart of the proposed method is shown in Fig. 1. Similar as successfully adopted learning-based IQA methods [24], [25], [26], [27], our method firstly extracts features from only tone-mapped images i.e., in NR manner, then learns a support vector regression (SVR) to predict quality scores. Different from common practices, our method does not directly extract features from TMIs themselves, instead, we simulate the HVS retinal process of viewing an image, and derive four response maps of four different viewing conditions. The cone responses can be formulated as:

\[ F_R(x, y) = I_R(x, y) \ast g(x, y, \sigma) \] (1)
\[ F_G(x, y) = I_G(x, y) \ast g(x, y, \sigma) \] (2)
\[ F_B(x, y) = I_B(x, y) \ast g(x, y, \sigma) \] (3)

where \( R, G, \) and \( B \) are the red, green, and blue channels, respectively, \( \ast \) is a convolution operator, and \( g(x, y, \sigma) \) is a 2-D gaussian filter to simulate the receptive field (RF) of cones, in which the parameter \( \sigma \) is used for controlling the size of RF.

\[ g(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp \left( -\frac{x^2 + y^2}{2\sigma^2} \right) \] (4)

Then, horizontal cells laterally modulate those cone signals for the reception of ganglion layer [35]. The modulation can be looked as a gain control mechanism for cone responses [36], which is given by:

\[ P_c(x, y) = \frac{F_c(x, y)}{\hat{F}_c} \quad c \in \{R, G, B\} \] (5)

with

\[ \hat{F}_c = \left( \text{mean}(F_c(x, y))^n \right)^{\frac{1}{n}} \] (6)

where \( \hat{F}_c \) simulates the modulation of horizontal cells, and \( n \) is used to emphasize the bright pixels [37].

Retinal ganglion cells (RGCs) receive the modulated cone signals and process them with the color-opponent mechanism [38]. Fig. 2 shows the receptive field of RGCs, in which the subunits inhibit each other at first and then inhibit the center. It can be expressed as follows:

\[ G_{\text{sub}}(x, y) = \max[0, \ P_G(x, y) - A_{UG} \times P_G(x, y) \ast g(x, y, \sigma_B)|_{P_G(x, y) = 0}] \] (7)
\[ G_{\text{sur}}(x, y) = A_S G \times G_{\text{sub}}(x, y) \ast g(x, y, \sigma_S) \] (8)
\[ R_{G(x,y)} = \max[0, \ P_R(x, y) \ast g(x, y, \sigma_C) - G_{\text{sur}}(x, y)] \] (9)
where \( x \) and \( y \) are the spatial indices, \( G_{\text{sub}}(x, y) \) denotes the subunit response, \( G_{\text{sur}}(x, y) \) is the surround response, and \( R \cdot G(x, y) \) is the final response in the R-G opponent channel. The operator max makes the result non-negative. \( A_{UG} \) and \( A_{SG} \) represent the sensitivities of the subunit and surround respectively. The responses of R-G and B-Y channels can also be similarly calculated using (7), (8), and (9), where \( Y \) is the yellow signal with \( P_y = \frac{P_R + P_G}{2} \). The parameters of gaussian functions are listed in Table 1.

### Table 1. Parameters of gaussian function.

| Function name | Standard deviation | Size |
|---------------|--------------------|------|
| \( g(x, y, \sigma) \) | \( \sigma = 3 \) | \( 3 \times 3 \) |
| \( g(x, y, \sigma_u) \) | \( \sigma_u = 0.5 \) | \( 3 \times 3 \) |
| \( g(x, y, \sigma_s) \) | \( \sigma_s = 1.5 \) | \( 7 \times 7 \) |
| \( g(x, y, \sigma_r) \) | \( \sigma_r = 0.5 \) | \( 3 \times 3 \) |

According to [39], RGCs are capable of adaptively adjusting the sensitivity to the viewing environment in the first few seconds. During the adaptation procedure, the color variation in each channel may reflect viewing experience and is critical to the quality assessment. Strictly referring to [34], \( A_{UR} = A_{UG} = A_{UB} = \frac{k}{1} \), and \( A_{SR} = A_{SG} = A_{SB} = \frac{4}{3} \), where \( k \) means the \( k \)th sensitivity adjustment.

### B. FEATURE EXTRACTION

1) GLOBAL FEATURES \((F_G)\)

As mentioned above, the retina adjusts the sensitivity from low to high when viewing an image. When the sensitivity increases, the color maps of each channel vary and can reflect viewing experience. In our method, \( K = 4 \), the maximum value for \( k \), is used to set the total number of sensitivities. Our first consideration is to catch the variation from calculating the gradient similarity. The gradient map is seen as one proper measurement of image structure [40], [41], which can be calculated as:

\[
G(x, y) = \sqrt{(I(x, y) \ast s_h(x, y))^2 + (I(x, y) \ast s_v(x, y))^2}
\]

(10)

with

\[
\begin{align*}
    s_h(x, y) &= \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}, \\
    s_v(x, y) &= \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}
\end{align*}
\]

(11)

where \( s_h(x, y) \) and \( s_v(x, y) \) are the horizontal and vertical filter kernels of Sobel filter, respectively. Mathematically, Sobel operator is implementation-friendly and effective.

Then, the gradient similarity is calculated between the first and the \( k \)th color maps in terms of the sensitivity in the same opponent color channel, which is expressed by:

\[
SIM_{k,Ch}(x, y) = \frac{2G_{1,Ch}(x, y)G_{k,Ch}(x, y) + C}{G_{1,Ch}^2(x, y) + G_{k,Ch}^2(x, y) + C},
\]

where \( Ch \) is the opponent channel of R-G, G-R or B-Y, and a small constant \( C = 0.0001 \) is appended to remove the instability of the equation. The mean value of each similarity map is used as one feature, so here we get 9 features, categorized as \( F_G \).

Fig. 3 shows three images tone mapped from a same HDR image by three different TMOs on TMQI database. Here, subjective quality scores of (a), (b) and (c) are 5.65, 1.55 and 7.9 in the form of MOS, respectively, where the higher score represents the lower quality experience. Compared to (b) of best visual quality, (a) is under-exposed and (c) is over-exposed and both of them lose a lot of details. To be specific, it’s too dark to clearly see the indoor scene from (a) and the pixels of outdoor part in (c) nearly all appear white. Fig. 4 shows the variation of the gradient similarity of Fig. 3 with the sensitivity \( k \). All curves fall with the increase of \( k \), and it can be found that the curves drop faster for Fig. 3(b) in all channels compared to those for Fig. 3(a) or Fig. 3(c). It may be contributed to that: 1) the higher \( k \) means the bigger difference created between the \( k \)th and the first color maps and then the smaller gradient similarity. 2) Fig. 3(a) and Fig. 3(c) are much flatter than Fig. 3(b) and smaller changes will happen when \( k \) increases. For the two severely ill-exposed images, the dropping rates of Fig. 3(c) are all lower than those of Fig. 3(a), which are also consistent with the subjective perception of Fig. 3(a) having better quality experience. Hence, the features of \( F_G \) have the potential for representing TMIs quality.

2) LOCAL FEATURES \((F_L)\)

After clarifying the indispensability of above features, we start to introduce more based on HVS. Consideration here is that, during tone mapping, the highest and the lowest luminance values are significantly compressed, leading to detail loss and bad viewing experience. That is, the quality deterioration is mostly susceptible in these two regions of TMIs [42]. It can also be observed in Fig. 3 that few details
are visible in the darkest area (Fig. 3(a)) and the brightest area (Fig. 3(c)).

Therefore, for these features, we segment the darkest and brightest regions by the lowest 10% and highest 10% values of input TMIs, respectively. This threshold is calculated in the LAB color space. Afterwards, the mean value $\mu$ and standard deviation $\sigma$ of these two segmented regions of the color maps of each opponent channel are calculated to evaluate their quality loss. The $\sigma$ is used to measure detail and the $\mu$ for brightness measurement. By combining such segmentation and color maps of each opponent channel, we get total 48 features grouped as $F_L$.

3) NSS FEATURES ($F_S$)

Also, natural images have certain statistical properties [43], [44], which may be violated by distortions. Similarly, instead of directly modeling the statistics from the input images, we calculate them for every color map produced by HVS. As in [8], the statistical model first processes the color map $\varphi$ by the mean subtraction and divisive normalization operator (MSCN), which is generated by:

$$\hat{\varphi}(x, y) = \frac{\varphi(x, y) - \mu_{\varphi}(x, y)}{\sigma_{\varphi}(x, y) + 1}$$  \hspace{1cm} (13)

with

$$\mu_{\varphi}(x, y) = \sum_{m=-M}^{M} \sum_{n=-N}^{N} w_{m,n} \varphi(x + m, y + n)$$  \hspace{1cm} (14)

$$\sigma_{\varphi}(x, y) = \sqrt{\sum_{m=-M}^{M} \sum_{n=-N}^{N} w_{m,n} \left[ \varphi(x + m, y + n) - \mu_{\varphi}(x, y) \right]^2}$$  \hspace{1cm} (15)

where $\mu_{\varphi}(x, y)$ is the local mean and $\sigma_{\varphi}(x, y)$ is the local standard deviation, and $w_{m,n}$ is one element in a circularly symmetric 2D $(2M + 1) \times (2N + 1)$ Gaussian kernel to weight the corresponding neighboring pixel. In our method, we set $M = 7, N = 7$ as in [24].

Fig. 5 shows the MSCN distributions of the images in Fig. 3. The (a1), (a2) and (a3) in the first row respectively show the distributions of the three opponent color channels (i.e., R-G, G-R and B-Y) of Fig. 3(a), where $k$ is the sensitivity. Similarly, the second row represents the distributions of Fig. 3(b) and the last row corresponds to Fig. 3(c). For the normal-exposure TMI of Fig. 3(b), the distributions in three color channels are more Gaussian-like than the other two tone-mapped images, which means it’s more “natural”. As you can see, with the increase of sensitivity $k$, the graphical shapes of distributions of Fig. 3(b) vary obviously, but such shapes of Fig. 3(a) and Fig. 3(c) almost keep unchanged. Therefore, we utilize a generalized Gaussian distribution (GGD) to get the statistical features for each scale of response maps to make “naturalness” evaluation. The GGD modeling MSCN is given by:

$$f(x, \alpha, \gamma^2) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp \left[ -\left( \frac{|x|}{\beta} \right)^{\alpha} \right]$$  \hspace{1cm} (16)

where $\alpha$ and $\beta$ are the shape and scale parameters of the GGD, respectively.
FIGURE 5. MSCN distributions of the images of Fig. 3. (a1), (a2) and (a3) show the MSCN distributions of R-G, G-R, and B-Y channels of Fig. 3(a), respectively. Where $k$ means the visual sensitivity. Similarly, the next two rows represent the distributions of Fig. 3(b) and Fig. 3(c).

with

$$\beta = \gamma \sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}}$$  \hspace{1cm} (17)

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}, \quad x > 0$$  \hspace{1cm} (18)

where $\Gamma(x)$ is the gamma function, $\alpha$ and $\beta$ are the shape and scale parameters, respectively. Since $\alpha$ and $\beta$ are taken as quality-aware features, we obtain 24 features here as $F_S$.

C. LEARNING TO A METRIC

Here, we describe how the final metric is derived from above features. Specifically, we adopt the response maps of four scales, which are generated by simulating the HVS processing mechanism. Three kinds of above features, $F_G$, $F_L$, and $F_S$, are extracted from these four response maps in each of the three opponent color channels, and we have got features of total number 81. Following the successful attempt of most NR-IQA methods, we employ SVR to pool all the extracted features. The SVR function [45] is given by:

$$\min_{w, b, \gamma, \gamma'} \frac{1}{2} |w|^2 + \lambda \sum_{i=1}^S (y_i + \gamma'_i)$$

$$\text{s.t.} \quad w^T \phi(x_i) + b - y_i \leq \eta + \gamma_i$$

$$\gamma_i, \gamma'_i \geq 0, \quad i = 1, 2, \ldots, S$$  \hspace{1cm} (19)

where $y_i$ and $\gamma'_i$ are slack variables, $b$ is a bias parameter. $x_i$ and $y_i$ are the input feature and quality score, respectively. $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ is the kernel function, and a radial basis function kernel (RBF) is adopted, which is defined as:

$$K(x_i, x_j) = \exp \left( -\mu ||x_i - x_j||^2 \right)$$  \hspace{1cm} (20)

Finally, the prediction model is built after setting all training parameters. We use 80% images from 2 datasets (see later experiment) for training.
III. EXPERIMENT
A. EXPERIMENT CONFIGURATION
1) EVALUATION CRITERIA
All objective IQA methods will be recognized precise only when their metrics are of high relevance with subjective scores. We use four criteria recommended by the Video Quality Experts Group (VQEG) [46] i.e., root mean squared error (RMSE), Pearson linear correlation coefficient (PLCC), Spearman rank-order correlation coefficient (SRCC), and Kendall’s rank-order correlation (KRCC), to measure such correlation. RMSE and PLCC are utilized to measure consistency and accuracy respectively, and SRCC, as well as KRCC, evaluates monotonicity. Lower RMSE and higher PLCC, SRCC, and KRCC stand for better performance. Also, to remove the nonlinearity of objective scores, a logistic regression model of five parameters is adopted before calculating RMSE and PLCC, which is given as follows:

\[
S = a_1 \left[ \frac{1}{2} - \frac{1}{1 + e^{a_2(S_p-a_3)}} \right] + a_4S_p + a_5
\]

where \(S_p\) is the input objective prediction value, and \(a_1, a_2, a_3, a_4\) and \(a_5\) are the parameters fitted by a nonlinear regression method.

2) DATASETS AND COMPETITORS
We evaluate the performance of the proposed method on two popular databases, TMID [18] and ESPL-LIVE [24].

a: TMID DATABASE
Developed by University of Waterloo, Canada, TMID consists of 15 HDR images and each of them is tone mapped by 8 different TMOs (total 120 TMIs). It also provides subjective quality scores by twenty observers, in the form of mean opinion score (MOS) ranging from 1 to 8 where a lower score means a higher perceptual quality.

b: ESPL-LIVE DATABASE
Created by the Laboratory of Image and Video Engineering at the University of Texas, USA, ESPL-LIVE Contains 1811 images generated via three approaches, including tone mapping, multi-exposure fusion and post processing. Its subjective evaluation was carried out on a crowdsourcing platform providing average 110 ratings for each image. After that, the collecting subjective ratings were used to calculate the MOS values.

c: COMPETITORS
We compare our method with some state-of-the-art (SOTA) quality evaluation methods, including SSIM [3], FSIM [19], NIQE [8], BRISQUE [9], GM-LOG [15], TMQI [18], BTMQI [23], HIGRADE-2 [24], BLIQUE-TMI [10], Yue’s method [29] and Wang’s method [30].

Among them, SSIM, FSIM and TMQI are FR and thus can’t be implemented on the ESPL-LIVE database which has no original HDR images, and the rest methods are NR. Also, SSIM, FSIM, NIQE, BRISQUE and GM-LOG are designed for evaluating synthesized distortions of standard dynamic range (SDR) images, while the rest are tailored for TMIs. To be specific, SSIM made evaluation by calculating the similarities of brightness, contrast and structure between reference and distorted images [3]; FSIM compared the similarities of the phase congruency and gradient magnitude between image pairs [19]; BRISQUE used the generalized Gaussian distribution (GGD) and asymmetric generalized Gaussian distribution (AGGD) to evaluate the “naturalness” of normalized luminance [9]; NIQE considered the sharpness of image regions, and introduced the multivariate Gaussian model (MVG) to represent the statistics of natural images [8]; GM-LOG computed the joint statistics of the gradient magnitude (GM) map and the Laplacian of Gaussian (LOG) response [15]; TMQI removed the luminance comparison component of SSIM and combined a naturalness measure to assess the TMIs quality [18]. BTMQI evaluated TMIs by considering information, naturalness and structure [23]. HIGRADE-2 laid emphasis on extracting NSS features from space domain and gradient domain in three LAB color channels [24]; BLIQUE-TMI computed moment and statistics in both brightness and color domains [10]. Yue’s method focused on the degradations of colorfulness, structure and naturalness [29]; Wang’s method extracted comprehensive features after exploiting the local and global characteristics of TMIs [30].

Test set consists of 20% (the rest of 80% training set) images from above two datasets. The testing and training process is repeated 1000 times and the median indices are applied as the final results.

B. RESULTS AND ANALYSIS
TABLE 2 shows the comparison results on the TMID and ESPL-LIVE databases. The indices (i.e., RMSE, PLCC, SRCC and KRCC) of best performance are indicated in bold. From the results, we can see that: First of all, those methods designed for synthesized SDR distortions, such as SSIM, FSIM, NIQE, BRISQUE and GM-LOG, can’t achieve satisfactory performance. Specifically, the SRCC values of NIQE, BRISQUE and GM-LOG are all below 0.5 on the ESPL-LIVE database. The main reason is they were designed for predicting common distortions, such as blurriness, blocking, noise, etc., which are very unlike distortions caused by tone mapping. Hence, it makes sense these metrics are not competent for TMIs evaluation. Next, we can see that FR method of SSIM obtains the worst results, and the GM-LOG outperforms it by approximately 93% in terms of SRCC. This phenomenon indicates traditional FR metrics supposing the same dynamic range of compared image pairs are not proper for TMI pairs with big gaps of dynamic ranges. Last, Yue’s and Wang’s methods are the most comparable blind methods, with respective SRCC values of 0.73 and 0.75 on the ESPL-LIVE database, slightly underperforming over ours. As demonstrated in TABLE 2, the proposed method performs better than the state-of-the-arts.
TABLE 2. Comparison results on the TMID and ESPL-LIVE databases. The best results are bolded.

| Metrics | Type | TMID | ESPL-LIVE |
|---------|------|------|-----------|
|         |      | RMSE | PLCC      | SRCC      | KRCR |
| SSIM    | FR   | 1.6529 | 0.3568 | 0.3143 | 0.2234 | -   | -    | -  |
| FSIM    | FR   | 1.4652 | 0.5612 | 0.5184 | 0.3453 | -   | -    | -  |
| NQIE    | NR   | 1.7555 | 0.4900 | 0.4522 | 0.3191 | 10.001 | 0.0768 | 0.0904 | 0.0603 |
| BRISQUE | NR   | 1.2369 | 0.5542 | 0.5065 | 0.4218 | 9.2159 | 0.3901 | 0.3493 | 0.2367 |
| GM-LOG  | NR   | 1.1430 | 0.6585 | 0.6073 | 0.4336 | 9.6135 | 0.5075 | 0.4993 | 0.3458 |
| TMQI    | FR   | 1.2236 | 0.7716 | 0.7407 | 0.5588 | -   | -    | -  |
| BTM1QI  | NR   | -   | 0.7647 | 0.7966 | 0.5994 | 7.8674 | 0.6175 | 0.6152 | 0.4834 |
| ITGRAD-2| NR   | -   | 0.8582 | 0.8007 | 0.6361 | 6.8354 | 0.7304 | 0.7295 | 0.5340 |
| BLIQUE-TMI | NR | -   | 0.8700 | 0.8340 | -   | -    | 0.7120 | 0.7040 | -  |
| Yue     | NR   | -   | 0.838 | 0.7908 | 0.6794 | 6.7796 | 0.7384 | 0.7328 | -  |
| Wang    | NR   | -   | 0.8840 | 0.8430 | -   | -    | 0.7430 | 0.7500 | -  |
| Proposed | NR   | 1.1378 | 0.8857 | 0.8620 | 0.6985 | 6.7664 | 0.7645 | 0.7517 | 0.5401 |

TABLE 3. Performance comparison on the three individual distortion sets of ESPL-LIVE. The best performed metrics are in bold.

| Metrics | Tone Mapping | Multi-Exposure Fusion | Post-Processing | Overall |
|---------|--------------|-----------------------|-----------------|---------|
|         | PLCC | SRCC | PLCC | SRCC | PLCC | SRCC | PLCC | SRCC |
| BRISQUE | 0.3364 | 0.2711 | 0.5525 | 0.5320 | 0.5499 | 0.4948 | 0.3901 | 0.3493 |
| GM-LOG  | 0.4935 | 0.4754 | 0.6264 | 0.6254 | 0.6167 | 0.5900 | 0.5075 | 0.4993 |
| BTM1QI  | 0.5853 | 0.5812 | 0.6163 | 0.6098 | 0.5776 | 0.5469 | 0.6175 | 0.6152 |
| ITGRAD-2| 0.7662 | 0.7563 | 0.7172 | 0.7056 | 0.6057 | 0.5823 | 0.7304 | 0.7295 |
| Yue     | 0.6852 | 0.6737 | 0.7274 | 0.7202 | 0.6655 | 0.6337 | 0.7384 | 0.7328 |
| Wang    | 0.7820 | 0.7330 | 0.7150 | 0.7330 | 0.6660 | 0.6340 | 0.7430 | 0.7500 |
| Proposed | 0.8013 | 0.7662 | 0.7275 | 0.7087 | 0.7054 | 0.6975 | 0.7645 | 0.7517 |

TABLE 4. Performance results with different training metrics.

| Train-Test Sizes | TMID | ESPL-LIVE |
|------------------|------|-----------|
|                  | PLCC | SRCC      | PLCC | SRCC |
| 20%-80%          | 0.7713 | 0.7625 | 0.6562 | 0.6425 |
| 30%-70%          | 0.8102 | 0.7986 | 0.6837 | 0.6783 |
| 40%-60%          | 0.8363 | 0.8233 | 0.7097 | 0.7042 |
| 50%-50%          | 0.8597 | 0.8379 | 0.7212 | 0.7136 |
| 60%-40%          | 0.8703 | 0.8503 | 0.7324 | 0.7245 |
| 70%-30%          | 0.8794 | 0.8591 | 0.7442 | 0.7316 |
| 80%-20%          | 0.8857 | 0.8620 | 0.7645 | 0.7517 |
| 90%-10%          | 0.8913 | 0.8683 | 0.7721 | 0.7608 |

To further verify the superiority of the proposed method, we conduct experiments on individual distortion sets from ESPL-LIVE i.e., tone mapping, multi-exposure fusion and post-processing. It can be observed from TABLE 3 that our method obtains the best performance, and only Wang’s method delivers slightly better SRCC than ours on multi-exposure fusion. Also, our method has a better performance on tone mapping than the other two distortions. To be specific, the SRCC values of tone mapping, multi-exposure fusion and post-processing are 0.7662, 0.7087 and 0.6975, respectively, indicating the effectiveness and robustness of our method.

C. ABLATION STUDIES

1) IMPACT OF TRAINING SET SIZE
As mentioned in [23] and [24], the performance of SVR based method is related to the size of training set. Here, we investigate this relationship by varying the training set percentage from 10% to 90% and the remaining as the testing set. For each value of percentage, the non-overlap random split is also carried out 1000 times, and the median indices are shown in TABLE 4.

As seen, the performance improves gradually when the percentage of training set increases. Accordingly, we set the portions of the training and testing to 80% and 20% respectively in the final model. Besides, the results show the high robust of the proposed method as the performance is still competent when the training percentage is 30%.

2) ON NUMBER OF SENSITIVITY VALUES
As introduced in Section II-B, the number of sensitivity values \( K \) determines the amount of the extracted features for evaluation. To explore its impact, we try \( K \epsilon \{ 2, 4, 6, 8 \} \) and test on the ESPL-LIVE database.

The results in Fig. 6 show that the performance raises synchronously with \( K \), however, the performance increase brought by bigger \( K \) is relatively slight and “uneconomical” when \( K > 4 \). Therefore, we make a trade-off between the complexity and performance of the algorithm, and neutrally set \( K = 4 \). This setting is also consistent with the HVS’s property of assessing an image mainly on the first few seconds, which is detailed above.

FIGURE 6. Performance comparison under different sensitivity index \( K \) settings on the ESPL-LIVE database. (a) Result comparison of PLCC, (b) Comparison of SRCC.
TABLE 5. Experimental results of different feature combinations on the two databases.

| Features | TMID | ESPL-LIVE |
|----------|------|-----------|
| $F_G$, $F_L$, $F_S$ | PLCC | SRCC | PLCC | SRCC |
| √ × × | 0.6132 | 0.5967 | 0.5463 | 0.5289 |
| × √ × | 0.7964 | 0.7813 | 0.7205 | 0.7192 |
| √ √ × | 0.5758 | 0.5534 | 0.5109 | 0.4975 |
| √ × √ | 0.8441 | 0.8326 | 0.7572 | 0.7511 |
| × √ √ | 0.7482 | 0.7361 | 0.6218 | 0.6197 |
| × √ × | 0.8126 | 0.7213 | 0.6084 | 0.6138 |
| √ √ √ | 0.8857 | 0.8620 | 0.7645 | 0.7517 |

FIGURE 7. Experimental comparison with differently setting the proportions of the brightest and the darkest regions. (a) PLCC comparison, (b) SRCC comparison.

3) ON OVER-/UNDER-EXPOSURE THRESHOLD
As mentioned above, over- and under-exposure in TMIs might be brought on some TMOs. In section II-B, the mean values and standard deviations of two segmented regions in every color map are calculated as features.

To reveal the impact of segmentation thresholds of two regions on the derived features. We set (over-exposure threshold) as (5%, 5%), (10%, 10%), (15%, 15%), (20%, 20%), (25%, 25%), respectively. Fig. 7 shows the performance on the ESPL-LIVE database. We empirically choose (10%, 10%) since it boosts the performance by more than 2% compared to (5%, 5%) and (15%, 15%).

4) ON COMBINATION OF FEATURES
We have adopted three groups of features: global gradient similarity features ($F_G$), local mean and standard deviation features ($F_L$), and statistics features ($F_S$). Although our method has highly superior performance, the contribution of each feature set is still unclear. For this purpose, we also make comparison of the distinct combinations of the three feature sets.

The results of ablation studies on the TMID and ESPL-LIVE datasets are shown in TABLE 5. As seen, the contribution of the feature groups to the performance lies in: 1) For individual feature set evaluation, $F_L$ obtains the best performance on both datasets, validating that over- and under-exposure are critical for TMIs quality assessment. 2) $F_S$ produces a relatively worse performance than other two (i.e., $F_G$ and $F_L$) on the both datasets, due to the incapability of traditional NSS expressing the specific distortions of TMIs. 3) The performance boosts when features are combined, and the more features involved, the better performance will be. Thus, the three feature sets are complementary and should be fused together for quality evaluation.

D. RUN TIME
To be efficient for practical deployment, a good quality assessment model should be with low complexity and consume little time. Therefore, we also investigate the run time of the proposed method, as well as the competing methods. All algorithms are operated in MATLAB 2016b on a windows 10 laptop, with a 2.5 GHz CPU processor and 8 GB RAM. TABLE 6 shows the average run time of all methods implemented on the TMID database. Obviously, our method performs moderate, about 1.73 seconds/image, consuming much less time than HIGRADE-2. As for real-time applications, our notebook computer is a little incompetent and it is feasible for reducing the proposed method’s run time by providing higher computational power.

E. FURTHER DISCUSSION
By mimicking the HVS procedure, we have built a blind quality assessment model for TMIs. Experiments show the edge of our model over the previous researches. The superiority of our method can mainly be attributed to that: First of all, we extract features concerning TMIs’ specific distortions, such as local features to measure the distortions in over- and under-exposed regions. The ablation results in TABLE 5 show that a desirable performance can be achieved with just applying $F_L$. In contrast, the methods of SSIM, FSIM, NIQE, BRISQUE and GM-LOG were designed for evaluating synthetical distortions, which are basically different from distortions in TMIs. Hence, they are not competent in evaluating TMIs (in TABLE 2). It is important to note SSIM has the worst performance, suggesting that the direct luminance comparison should not be used in TMIs assessment. Secondly, all features are obtained from visual response maps. Since HVS plays the first important and decisive role in assessing images, simulating HVS is regarded as one of the most effective solutions to IQA problems [29], [31]. All the competing methods in this section straightforwardly extract quality-related features from to-be-evaluated TMIs, thus it is supposed that our extracted features possibly better express visual experience. From Fig. 4 and Fig. 5, it is clear to see the collecting data from the visual procedure have the ability of well representing the quality. Lastly, our method gathers features in three opponent color channels. Color destruction will occur following the dynamic range compression of tone mapping [28], [29]. For example, obvious colorfulness losses can be seen from ill-exposed Fig. 3(a) and Fig. 3(c). To address such an issue, the most recently proposed NR methods, including HIGRADE-2, BLIQUE-TMI, Yue and Wang, assess the chromatic information and achieve promising results, which are all better than those of benchmark FR TMQI. Therefore, our substantial attention to color properties may be a positive factor for the proposed method’s outperformance.
Albeit the advantages of the proposed method, there are still several aspects to be improved: 1) We can adopt more precise HVS models to simulate the procedure of visual process more accurately. 2) Currently, the over- and under-exposed regions are segmented by simple threshold. In the future work, we plan to use a more appropriate content-related partition methods as in [30]. 3) At present, we use “hand-crafted” features and adopt a SVR learning-scheme. Such features and learning mechanism might be insufficient compared with learning-based deep neural network (DNN). Hence, we can utilize DNN for automatic feature extraction of TMIs assessment, on the premise of developing a new massive database for such data-driven approaches.

IV. CONCLUSION

In this article, we propose a new effective blind (NR) quality model for TMIs using extracted features by simulating the procedure of HVS. Specifically, HVS is adaptive to the content of images, and the retinal responses of different sensitivities are used to indicate the visual processing. To envelop the variation of visual responses, gradient similarities are calculated between the first and the kth sensitivity color maps in each opponent color channel. Also, due to the over-or under-exposure caused by the dynamic range compression, we segment to-be-assessed TMIs into three parts, and calculate local means and standard deviations as features in the darkest and brightest regions in all response maps. We also employ natural statistics for all the response maps to ameliorate the performance. Then, SVR is utilized to pool all these obtained features and build a regression model for quality prediction. Experimental results on the TMID and ESPL-LIVE databases demonstrate the proposed method is superior over the other state-of-the-art assessment metrics.

REFERENCES

[1] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski, *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. San Mateo, CA, USA: Morgan Kaufmann, 2010.
[2] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced High Dynamic Range Imaging*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2017.
[3] 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.
[4] Z. Wang, E. P. Simoncelli, and A. C. Bovik, “Multiscale structural similarity for image quality assessment,” in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 2, Jul. 2003, pp. 1598–1402.
[5] A. Rehman and Z. Wang, “Reduced-reference image quality assessment by structural similarity estimation,” *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3378–3389, Aug. 2012.
[6] Y. Liu, G. Zhai, K. Gu, X. Liao, D. Zhao, and W. Gao, “Reduced-reference image quality assessment in free-energy principle and sparse representation,” *IEEE Trans. Multimedia*, vol. 20, no. 2, pp. 379–391, Feb. 2018.
[7] Y. Fang, J. Liu, Y. Zhang, W. Lin, and Z. Guo, “Reduced-reference quality assessment of image super-resolution by energy change and texture variation,” *J. Vis. Comm. Image Represent.*, vol. 60, no. 1, pp. 140–148, Apr. 2019.
[8] A. Mittal, R. Soundararajan, and A. C. Bovik, “Making a ‘completely blind’ image quality analyzer,” *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Nov. 2013.
[9] A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
[10] Q. Jiang, F. Shao, W. Lin, K. Gu, G. Jiang, and H. Sun, “Optimizing multistage discriminative dictionaries for blind image quality assessment,” *IEEE Trans. Multimedia*, vol. 20, no. 8, pp. 2035–2048, Aug. 2018.
[11] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, “A psychovisual quality metric in free-energy principle,” *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 41–52, Jan. 2011.
[12] D. Lee and N. K. Platanioti, “Toward a no-reference image quality assessment using statistics of perceptual color descriptors,” *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3875–3889, Aug. 2016.
[13] Y. Fang, J. Yan, J. Liu, S. Wang, Q. Li, and Z. Guo, “Objective quality assessment of screen content images by uncertainty weighting,” *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 2016–2027, Apr. 2017.
[14] Y. Zhou, L. Li, J. Wu, K. Gu, W. Dong, and G. Shi, “Blind quality index for multiply distorted images using biorder structure degradation and non-local statistics,” *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 3019–3032, Nov. 2018.
[15] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, “Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features,” *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4850–4862, Nov. 2014.
[16] K. Ma, W. Liu, K. Zhang, Z. Duannu, Z. Wang, and W. Zuo, “End-to-end blind image quality assessment using deep neural networks,” *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1202–1213, Mar. 2018.
[17] Q. Wu, H. Li, F. Meng, and K. N. Ngan, “A perceptually weighted rank correlation indicator for objective image quality assessment,” *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2499–2513, May 2018.
[18] H. Yeganeh and Z. Wang, “Objective quality assessment of tone-mapped images,” *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 657–667, Feb. 2013.
[19] H. Z. Nafchi, A. Shahkohi, R. F. Moghaddam, and M. Cheriet, “FSITM: A feature similarity index for tone-mapped images,” *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1026–1029, Aug. 2015.
[20] D. Kundu and B. L. Evans, “Visual attention guided quality assessment of tone-mapped images via scene statistics,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 96–100.
[21] H. R. Nasrinpour and N. D. Bruce, “Saliency weighted quality assessment of tone-mapped images,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 4947–4951.
[22] K. Ma, H. Yeganeh, K. Zeng, and Z. Wang, “High dynamic range image compression by optimizing tone mapped image quality index,” *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3086–3097, Oct. 2015.
[23] K. Gu, S. Wang, G. Zhai, and S. Ma, “Blind quality assessment of tone-mapped images via analysis of information, naturalness, and structure,” *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 432–443, Mar. 2016.
[24] D. Kundu, D. Ghaediaram, A. C. Bovik, and B. L. Evans, “No-reference quality assessment of tone-mapped HDR pictures,” *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2957–2971, Jun. 2017.
[25] G. Yue, C. Hou, K. Gu, T. Zhou, and H. Liu, “No-reference quality evaluator of transparently encrypted images,” *IEEE Trans. Multimedia*, vol. 21, no. 9, pp. 2184–2194, Sep. 2019.
[26] G. Yue, C. Hou, K. Gu, T. Zhou, and G. Zhai, “Combining local and global measures for DIBR-synthesized image quality evaluation,” *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 2075–2088, Apr. 2019.
[27] Q. Jiang, Z. Liu, K. Gu, F. Shao, X. Zang, H. Liu, and W. Lin, “Single image super-resolution quality assessment: A real-world dataset, subjective studies, and an objective metric,” *IEEE Trans. Image Process.*, vol. 31, pp. 2279–2294, 2022.
[28] G. Yue, C. Hou, and T. Zhou, “Blind quality assessment of tone-mapped images considering colorfulness, naturalness, and structure,” *IEEE Trans. Ind. Electron.*, vol. 66, no. 5, pp. 3784–3793, May 2019.
[29] G. Yue, W. Yan, and T. Zhou, “Referenceless quality evaluation of tone-mapped HDR and multiplexposure fused images,” IEEE Trans. Ind. Inform., vol. 16, no. 3, pp. 1764–1775, Mar. 2020.

[30] X. Wang, Q. Jiang, F. Shao, K. Gu, G. Zhai, and X. Yang, “Exploiting local degradation characteristics and global statistical properties for blind quality assessment of tone-mapped HDR images,” IEEE Trans. Multimedia, vol. 23, no. 1, pp. 692–705, 2021.

[31] G. Yue, C. Hou, K. Gu, S. Mao, and W. Zhang, “Biologically inspired blind quality assessment of tone-mapped images,” IEEE Trans. Ind. Electron., vol. 65, no. 3, pp. 2525–2536, Mar. 2018.

[32] H. J. A. Dartnall, J. K. Bowmaker, and J. D. Mollon, “Human visual pigments: Microspectrophotometric results from the eyes of seven persons,” Proc. Roy. Soc. London, B, Biol. Sci., vol. 220, no. 1218, pp. 115–130, 1983.

[33] T. Gollisch and M. Meister, “Eye smarter than scientists believed: Neural computations in circuits of the retina,” Neuron, vol. 65, no. 2, pp. 150–164, Jan. 2010.

[34] X.-S. Zhang, S.-B. Gao, R.-X. Li, C.-Y. Li, and Y.-J. Li, “A retinal mechanism inspired color constancy model,” IEEE Trans. Image Process., vol. 25, no. 3, pp. 1219–1232, Mar. 2016.

[35] R. H. Masland, “The neuronal organization of the retina,” Neuron, vol. 76, no. 2, pp. 266–280, Oct. 2012.

[36] C. J. Jospelevitch, “Human retinal circuitry and physiology,” Psychol. Neurosci., vol. 1, no. 2, pp. 141–165, Jul./Dec. 2008.

[37] G. D. Finlayson and E. Trezzi, “Shades of gray and colour constancy,” in Proc. Color Imag. Conf., 2004, pp. 37–41.

[38] F.-T. Qiu and C.-Y. Li, “Mathematical simulation of disinhibitory properties of concentric receptive field,” Acta Biophys. Sinica, vol. 11, no. 2, pp. 214–220, 1995.

[39] G. Vasserman, E. Schneiderman, and R. Segev, “Adaptive colour contrast coding in the salamander retina efficiently matches natural scene statistics,” PLoS ONE, vol. 8, no. 10, Oct. 2013, Art. no. e79163.

[40] K. Gu, G. Zhai, X. Yang, and W. Zhang, “An efficient color image quality metric with local-tuned-global model,” in Proc. IEEE Int. Conf. Image Process. (ICIP), Oct. 2014, pp. 506–510.

[41] A. Liu, W. Lin, and M. Narwaria, “Image quality assessment based on gradient similarity,” IEEE Trans. Image Process., vol. 21, no. 4, pp. 1500–1512, Apr. 2012.

[42] S. Gao, W. Han, Y. Ren, and Y. Li, “High dynamic range image rendering with a luminance-chromaticity independent model,” in Proc. Int. Conf. Intell. Sci. Big Data Eng., 2015, pp. 220–230.

[43] W. S. Geisler, “Visual perception and the statistical properties of natural scenes,” Annu. Rev. Psychol., vol. 59, pp. 167–192, Jan. 2008.

[44] D. L. Ruderman, “The statistics of natural images,” Netw. Comput. Neural Syst., vol. 5, no. 4, pp. 517–548, Jul. 1994.

[45] B. Scholkopf and A. J. Smola, Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. Cambridge, MA, USA: MIT Press, 2002.

[46] VQEG. (Mar. 2000). Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment. [Online]. Available: http://www.vqeg.org/