IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 13, 2017, accepted January 19, 2018, date of publication January 26, 2018, date of current version March 19, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2798573

Attended Visual Content Degradation Based Reduced Reference Image Quality Assessment

JINJIAN WU [1], YONGXU LIU[1], LEIDA LI[2], AND GUANGMING SHI[1], (Senior Member, IEEE)

[1]Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, School of Artificial Intelligence, Xidian University , Xi'an 710000, China

[2]School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou 221116, China

Corresponding author: Jinjian Wu (jinjian.wu@mail.xidian.edu.cn)

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant JB181706; and in part by the National Natural Science Foundation of China under Grant 61772388, Grant 61472301, Grant 61632019, and Grant 61621005.

ABSTRACT Reduced-reference (RR) image quality assessment (IQA), which aims to use a small amount of the reference information but achieve high accuracy, is greatly demanded in quality-orientated systems. In order to design a better RR IQA model which performs consistently with the subjective perception, the inner mechanism of the human visual system (HVS) is usually investigated and imitated. In this paper, the attention mechanism is thoroughly analyzed and used for RR IQA modeling. Generally, the HVS is more sensitive to the distortion on the attended region than that on the unattended region. Thus, the saliency of each region is calculated to highlight its importance, and a saliency weighted local structure (SWLS)-based histogram is created for visual structure degradation measurement. Meanwhile, the distortion may cause attention shift (changing the attended region). In other words, the difference of attention between the reference and distorted images can efficiently represent the quality degradation. Therefore, the attention distribution is analyzed with the salient map, and an orientation located global saliency (OLGS)-based histogram is built for attention shift measurement. Finally, combining the quality degradations from both SWLS and OLGS, a novel attended visual content degradation-based RR IQA method is introduced.[1] Experimental results demonstrate that the proposed method uses only several values (18 values) and performs consistently with the subjective perception. Moreover, the proposed attention procedure can be easily extended to the existing RR IQA models and improve their performances.

INDEX TERMS Reduced-reference (RR), image quality assessment (IQA), visual attention, saliency, content degradation.

I. INTRODUCTION

With a tremendous growth of visual signal, images/videos take up 70% Internet traffic. Due to the near-ubiquitous presence of images/videos, a reliable and efficient objective image quality assessment (IQA) method is greatly demanded for many quality-oriented image acquisition/processing systems [1], [2].

During the past decades, a large amount of IQA methods have been proposed. According to the volume of the reference image required during the quality prediction, existing IQA methods are usually classified into three categories: 1) Full-Reference (FR) IQA [3], where the whole reference image is required; 2) Reduced-Reference (RR) IQA [4], where only a small part (several values) of the reference information is available; and 3) No-Reference (NR) IQA [5], for which nothing about the reference image is available. With the guidance of the reference image, local features are always extracted and compared with a point-to-point way between the reference and distorted images. As a result, the FR IQA models perform consistently with subjective perception [6]. However, the reference image is not always available in many applications, and a NR IQA model is required for such condition [7]. Without any guideline from the reference image, it is extremely difficult to design a reliable and efficient NR IQA model. As a compromise between the two categories, the RR IQA technique requires only a small amount of the reference information and achieves a high performance [8]. In this work, we focus on RR IQA modeling.

Reliable RR IQA models aim to use as less reference data as possible, and achieve a higher prediction

[1]The source code of the proposed method will be available at http://web.xidian.edu.cn/wjj/en/index.html

(a) (b) (c)

FIGURE 1. An example of distortion on attended *V.S.* unattended regions. (a) MOS= 6.0, PSNR= 27.3. (b) MOS= 3.80, PSNR= 27.2. (c) MOS= 2.84, PSNR = 27.2.

accuracy [9]. To this end, global features, which can efficiently represent the visual content of an image with limited values, are usually designed for quality prediction [8]. Statistical researches on images indicate that the contents of nature scenes follow a certain type of statistical distribution (e.g., generalized Gaussian distribution) [10], which is called as natural scene statistic (NSS) feature. Distortions on the natural scene will change its NSS characteristic to some extent, which makes the natural image unnatural [11]. According to this assumption, the distribution of the wavelet coefficients was analyzed for NSS feature extraction during quality prediction in [12] (i.e., the Wavelet-domain Natural Image Statistic Metric, WNISM). Moreover, the NSS features were analyzed on the divisive normalization based wavelet coefficients [13], the discrete cosine transform (DCT) coefficients [14], the curvelets and the contourlets coefficients [15] for RR IQA, respectively. Besides, according to the visual cognition theory, the degradation on visual information was measured for RR IQA, e.g., the visual information fidelity (VIF) [16], the orientation selectivity based visual pattern degradation (OSVP) [17], the scaled entropy degradation (RRED) [18], the entropy of weighted wavelet coefficients (RED-LOG) [19], and so on. Though these existing methods have greatly improved the performance of the RR IQA, there is still a remarkable gap between RR IQA models and the subjective perception.

Most of the existing RR IQA algorithms equally measure the degradations from the whole image. However, the human visual system (HVS) presents obvious attention mechanism, with which the HVS focuses only on these important regions for detailed perception and withdraws the other regions [20], [21]. In other words, attended regions play more important roles than the other unattended regions for visual perception, and degradations on attended regions have a larger influence on the quality assessment. In recent years, the attention mechanism has already been taken into account for FR IQA, in which the salient map (saliency is a strong predictor for attention/gaze allocation) is created to highlight these attended regions during quality assessment [22]. However, the size of the salient map (always with a same size as the reference image) is too large for RR IQA. How to highlight the attended regions with several values for the RR IQA is still an open problem. Moreover, distortions always change the attended regions (between the reference and distorted images) [23], how to measure the effect of distortion on attention shift is another challenge.

In this work, we introduce an attended visual content degradation (AVCD) based RR IQA method. Firstly, the effects of distortions on attended regions are thoroughly investigated and analyzed. Next, the quality degradation on the local structure is measured. By extracting the local structure with the visual orientation pattern [17] and highlighting with its salient value, a saliency weighted local structure (SWLS) based histogram is created for visual content degradation measurement. Meanwhile, the degradation on the attention shift is estimated. According to the distribution of attended regions, the histogram of gradient (HOG) [24] is calculated on the salient map, and an orientation located global saliency (OLGS) based histogram is built for the attention shift measurement. Finally, considering degradations from both SWLS and OLGS between the reference and distorted images, the novel AVCD based RR IQA model is proposed.

II. ATTENDED VISUAL CONTENT DEGRADATION

It is well known that our human eyes cannot attend to all things at once, and we focus our attention on selections of the input scene. So that the brain can successfully and effectively perceive these regions of interest (ROI) [25]. In general, the ROI in an image refers to the object regions with high informative contents, which can not only reduce the further processing complexity but also increase the processing accuracy [26]. Thus, ROI detection (visual attention estimation) has attracted tons of research attention and has been successfully used in many visual recognition tasks [27].

Attended regions (ROIs) play more important roles than the other regions during image perception, and thus, the distortion on attended regions will cause more serious quality degradation than that on unattended regions. In order to give an intuitive interpretation, an example of distortions on different image regions are shown in Fig. 1 (three images are distorted by the JPEG transmission error (JTE)). Though the noise levels in these images are almost the same (the Peak

(a) (b) (c)

FIGURE 2. An example of attention shift caused by distortion. (a) ORG. (b) J2K, MOS= 2.0. (c) JTE, MOS= 2.57.

Signal to Noise Ratio (PSNR) for Fig. 1 (a)-(c) are 27.3dB, 27.2dB, and 27.2dB, respectively), the quality degradation degrees for them are obviously different. As can be seen, Fig. 1 (a) (with mean opinion score MOS $=$ 6.00) has a better quality than the other two images (Fig. 1 (b) and (c) with MOS $=$ 3.80 and MOS $=$ 2.84, respectively). With further analysis, we have found that the JPEG transmission error in Fig. 1 (a) mainly occurred at the background reef area (located at the bottom of the image), and little at the foreground objection regions (e.g., the lighthouse). In other words, the distortion is mainly located at unattended regions, while little at attended regions. As a result, the quality degradation in Fig. 1 (a) is limited. For Fig. 1 (b), the distortion is mainly located at the contour of the left parrot, especially on its body (rather than the head). Since the attended objects are partly distorted, the quality of Fig. 1 (b) is obviously degraded. When on Fig. 1 (c), most of the distortion is located at these highly attended regions (e.g., the contours of the hats), and its quality is seriously degraded. Since distortions on attended regions cause more serious quality degradation than that on unattended regions, the attended objective regions should be highlighted during quality assessment.

Moreover, distortions may change the attended regions between the reference and distorted images. As shown in Fig. 2, the original girl face image (Fig. 2 (a)) is distorted by JPEG 2000 noise (J2K) (Fig. 2 (b)) and JPEG transmission error noise (JTE) (Fig. 2 (c)). Obviously, the two eyes and the mouth regions will attract our attention when looking at the original image (Fig. 2 (a)). However, distortion may degrade the attended objects and make them unattended. As shown in Fig. 2 (b), the J2K has smoothed out the structures of the right eye and the mouth, and the two regions will no more attract our attention in the distorted image. Meanwhile, the background region with severe distortion may attract our attention. As shown in Fig. 2 (c), the distortion on the jaw generates a fake edge, which will attract our attention. Therefore, the shift of attended regions can reflect the characteristic of the distortion.

Considering the effect of distortion on visual attention during image perception, both the local weighing and the global shifting of attention should be considered for visual content degradation measurement during RR IQA.

III. DEGRADATION MEASUREMENT

In this section, the SWLS is firstly extracted for local visual content degradation measurement. Next, the OLGS is estimated for global attention shifting calculation. Finally, considering the degradations on both SWLS and OLGS, a novel AVCD based RR IQA model is built.

A. SWLS BASED LOCAL CONTENT EXTRACTION

The attended places normally refer to the informative object regions, and such regions play more important roles during image perception and understanding. Thus, distortions in the attended regions are more sensitive to the HVS, and the attended contents should be highlighted during quality assessment. To this end, we try to weight the local content with its saliency for visual content extraction in this subsection.

It is well known that the HVS is highly adaptive to extract the local structure for image/video perception. And thus, the local structure is firstly analyzed for visual content extraction. Researches on visual cognition further state that the HVS is extremely adapted to summarize the organizing rules, especially by extracting these repeated local structures [28]. As an efficient representation of the repeated local structures of images, the pattern is usually adopted for visual content representation [29]. Moreover, the HVS presents substantially orientation selectivity (OS) mechanism for pattern extraction [30]. According to the similarity of the preferred orientations that neighbor neurons reacted, the local receptive field presents obvious OS [31].

With the inspiration from the OS mechanism (the interactions among neurons in a local receptive field), the relationships between neighbor pixels are analyzed with their preferred orientations, and an OS based visual pattern is introduced with the arrangement of the relationships [17]. Firstly, the preferred orientation of each pixel is calculated as its gradient direction $\mathcal{G}_d$,

$$\mathcal{G}_d(x) = \arctan \frac{\mathcal{M}_v(x)}{\mathcal{M}_h(x)}, \tag{1}$$

where $\mathcal{M}_v$ is the gradient magnitude along the vertical direction, and $\mathcal{M}_h$ is that along the horizontal direction.

Next, the relationship between a central pixel x and its neighbor pixel x_i is analyzed as the similarity of their preferred orientations. There are two opponent responses between neurons in a local receptive, i.e., excitatory and inhibitory interactions. By mimicking this, the relationship $\mathcal{R}(x|x_i)$ is represented with a binary form,

$$\mathcal{R}(x|x_i) = \begin{cases} 1 & \text{if } |\mathcal{G}_d(x) - \mathcal{G}_d(x_i)| < \mathcal{T} \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where '1' means the two pixels present excitatory interaction, and '0' means inhibitory interaction. $\mathcal{T}$ is the judging threshold. According to the subjective visual masking experiment [8], the threshold $\mathcal{T}$ is set as 6° in this work (there exists strong masking effect if the angle difference of two orientation gratings is smaller than 12°).

Finally, the OS based visual pattern (OSVP, $\mathcal{P}$) that a local region ($\mathcal{L}$) possesses is represented by the arrangement of the relationship between the central pixel $x \in \mathcal{L}(x)$ and its neighbors $x_i \in \mathcal{L}(x)$,

$$\mathcal{P}(x) = \mathcal{A}(\mathcal{R}(x|x_1, \cdots, x_n)) = \{\mathcal{R}(x|x_1), \cdots, \mathcal{R}(x|x_n)\}, \tag{3}$$

where n is the size of the neighborhood (i.e., the number of the neighbor pixels). Considering the computational complexity, n is set as 8 in this work (same as that in [17]).

With Eq. (3), the pattern form $\mathcal{P}(x)$ for each pixel is built. In order to reduce the number of OSVPs, these similar ones should be combined. Since patterns with the same number of highly related elements (i.e., the number of '1') are more likely to represent similar information, these OSVPs are combined and represented by a same pattern form (only 9 pattern forms are reserved) [17].

Besides the pattern form, the HVS is also highly sensitive to the luminance changes (LC). In this work, the LC of each pixel is calculated as the corresponding magnitude of gradient,

$$\mathcal{M}(x) = \sqrt{(\mathcal{M}_v(x))^2 + (\mathcal{M}_h(x))^2}. \tag{4}$$

The visual content of a pixel x can be represented as the joint of OSVP and LC, i.e., $\{\mathcal{P}(x), \mathcal{M}(x)\}$. Thus, by summarizing the LC of each pixel according to its OSVP, the visual content of an image can be mapped into an OSVP based histogram ($\mathcal{H}_{\text{OSVP}}$).

$$\mathcal{H}_{\text{OSVP}}(k) = \sum_{m=1}^{M} \mathcal{M}(x_m) \cdot \delta(\mathcal{P}(x_m), \mathcal{P}_k) \tag{5}$$

$$\delta(\mathcal{P}(x_m), \mathcal{P}_k) = \begin{cases} 1 & \text{if } \mathcal{P}(x_m) = \mathcal{P}_k \\ 0 & \text{else,} \end{cases} \tag{6}$$

where M is the total pixel number in the image, and $\mathcal{P}_k$ represents the k-th OSVP.

Since the attended regions play a more important role than other regions (as illustrated in the former section), we try to highlight these attended regions and suppress the other regions with a salient map for visual content extraction. To this end, the salient value of each pixel is required. In the past decades, a large amount of saliency estimation models have been proposed. As a simple but powerful saliency estimator, the classical graph-based saliency model [32] is employed,

$$\mathcal{S}(x) = f(A_1(x), \cdots, A_n(x)), \tag{7}$$

where $\mathcal{S}(x)$ is the salient value for pixel x, $A_i(x)$ is the unusual/surprise of x at the $i-th$ feature channel, and n is the channel number (more information about saliency detection can be found in [32]).

By weighting the local structure with it salient value, the visual content of an image is mapped into a SWLS based histogram ($\mathcal{H}_{\text{SWLS}}$),

$$\mathcal{H}_{\text{SWLS}}(k) = \sum_{m=1}^{M} \mathcal{S}(x_m) \cdot \mathcal{M}(x_m) \cdot \delta(\mathcal{P}(x_m), \mathcal{P}_k). \tag{8}$$

An intuitive example on the visual content extraction is shown in Fig. 3, where Fig. 3 (a)-(d) are the original image, the salient map, $\mathcal{H}_{\text{OSVP}}$ from (5), and $\mathcal{H}_{\text{SWLS}}$ from (8), respectively.

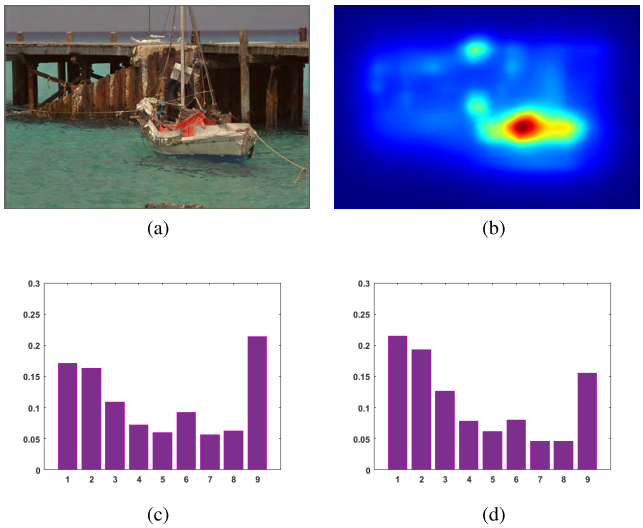


FIGURE 3. An intuitive example on the visual content extraction. (a) Original Image. (b) Salient Map. (c) $\mathcal{H}_{OSVP}$. (d) $\mathcal{H}_{SWLS}$.

B. OLGS BASED GLOBAL ATTENTION ESTIMATION

Distortions may directly change the salient regions between the reference and distorted images (as illustrated in the former section). And thus, the saliency itself is an efficient feature to represent the quality degradation. An example of attention shift is shown in Fig 4. Due to the effect from distortions (i.e., J2K in Fig 4 (b) and JTE in Fig 4 (c)), the attention regions are obviously shifted (comparing the salient maps Fig 4 (d)-(f)). However, the size of the salient map (with a same size of the reference image) is too large for RR IQA measurement (RR IQA only uses a small amount of values).

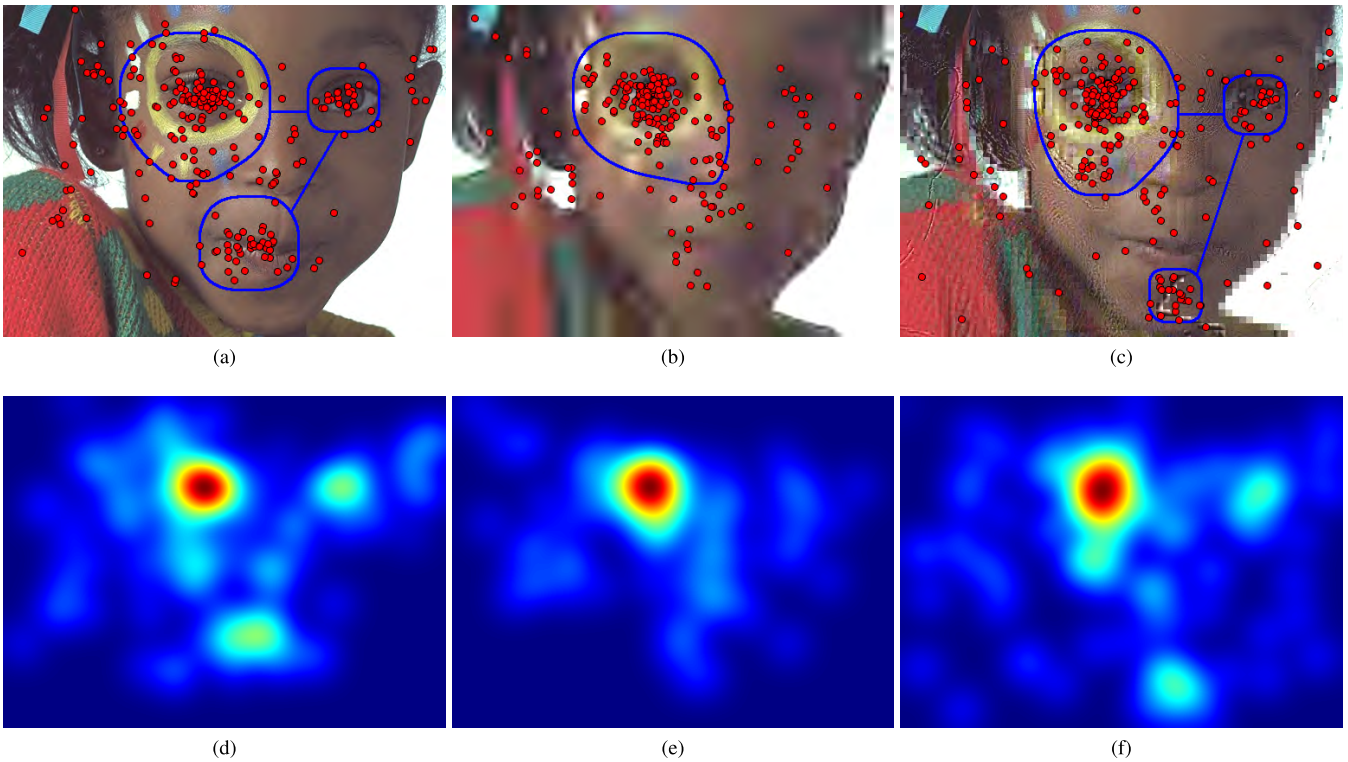


FIGURE 4. Attention shift caused by different distortion types from the Eye-tracker. (a) Gazing map for ORG. (b) Gazing map for J2K. (c) Gazing map for JTE. (d) Salient map for ORG. (e) Salient map for J2K. (f) Salient map for JTE.

How to efficiently represent the saliency information with several values for RR IQA is still an open problem (to the best of our knowledge, there is not existing work about this so far).

In order to efficiently represent the saliency characteristic with several values, the content description for salient map is firstly analyzed. The primary function of the salient map is salient object localization, for which these regions with larger values correspond to salient objects. As an efficient descriptor for the low-level feature, the Histogram of Oriented Gradients (HOG) has been successfully applied for object localization [33], e.g., human detection [24], face localization [34], and so on. Thus, the HOG is adopted to represent the characteristic of the salient map for attention shift measurement.

The HOG is directly analyzed on the salient map of an image I. According to Eq. (7), the saliency of I is calculated, and a salient map $\mathcal{S}$ is acquired. Next, the gradient direction ($\mathcal{G}_{\mathcal{S}}(x)$) and magnitude ($\mathcal{M}_{\mathcal{S}}(x)$)of each pixel in $\mathcal{S}$ are calculated with Eq. (1) and Eq. (4). Since the gradient direction is with a wide range (can be any value in $[-180°, 180°)$), it is usually equally normalized into N (e.g., 9) discrete directions. And a salient map can be mapped into an OLGS based histogram with N-bin HOG,

$$\mathcal{H}_{\mathrm{OLGS}}(k) = \sum_{m=1}^{M} \mathcal{M}_{\mathcal{S}}(x_m) \cdot \delta\left(\mathrm{HOG}(x_m), \mathrm{HOG}_k\right) \tag{9}$$

$$\delta\left(\mathrm{HOG}(x_m), \mathrm{HOG}_k\right) = \begin{cases} 1 & \text{if } \mathrm{HOG}(x_m) = \mathrm{HOG}_k \\ 0 & \text{else,} \end{cases} \tag{10}$$

where $\mathrm{HOG}(x_m)$ is the corresponding bin of the gradient direction for pixel x_m, and HOG_k is the k-th bin for the resulting histogram. With Eq. (9), a salient map can be represented with N-bin histogram. An intuitive example of OLGS is shown in Fig. 5, whose corresponding salient maps are shown in Fig. 4.

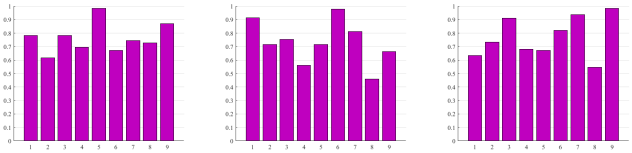


FIGURE 5. HOG based histogram for the OLGS representation. (a) ORG. (b) J2K. (c) JTE.

C. QUALITY PREDICTION

The quality is measured as the attended visual content degradation in two aspects: 1) the degradation on the local structure/content and 2) the attention shift. Since the RR-IQA aims to use only several values during quality prediction, the local content of an image is represented by a SWLS based histogram (i.e., $\mathcal{H}_{\mathrm{SWLS}}$ with Eq. (8)), and the attention localization is represented by an OLGS based histogram

(i.e., $\mathcal{H}_{\mathrm{OLGS}}$ with Eq. (9)). Thus, the quality degradation is measured as the changes on the two histograms.

For a given reference image I^r and its corresponding distorted image I^d, the degradation on the local structure ($\mathcal{Q}_{\mathrm{SWLS}}$) is measured as the distance $\mathcal{D}$ between their corresponding SWLS based histograms,

$$\begin{aligned}\mathcal{Q}_{\mathrm{SWLS}}(I^d|I^r) &= \mathcal{D}(\mathcal{H}^r_{\mathrm{SWLS}}, \mathcal{H}^d_{\mathrm{SWLS}}) \\ &= \frac{1}{N}\sum_{k=1}^{N}\frac{2\cdot\mathcal{H}^r_{\mathrm{SWLS}}(k)\cdot\mathcal{H}^d_{\mathrm{SWLS}}(k)}{\mathcal{H}^r_{\mathrm{SWLS}}(k)^2+\mathcal{H}^d_{\mathrm{SWLS}}(k)^2}\end{aligned} \tag{11}$$

where $\mathcal{H}^r_{\mathrm{SWLS}}$ ($\mathcal{H}^d_{\mathrm{SWLS}}$) is the SWLS based histogram of the reference (distorted) image, and N is the bin number of the histogram.

Meanwhile, the attention shift ($\mathcal{Q}_{\mathrm{OLGS}}$) is measured as the changes on the OLGS based histograms between the reference ($\mathcal{H}^r_{\mathrm{OLGS}}$) and the distorted ($\mathcal{H}^d_{\mathrm{OLGS}}$) images.

$$\begin{aligned}\mathcal{Q}_{\mathrm{OLGS}}(I^d|I^r) &= \mathcal{C}(\mathcal{H}^r_{\mathrm{OLGS}}, \mathcal{H}^d_{\mathrm{OLGS}}) \\ &= \frac{1}{N}\sum_{k=1}^{N}\frac{2\cdot\mathcal{H}^r_{\mathrm{OLGS}}(k)\cdot\mathcal{H}^d_{\mathrm{OLGS}}(k)}{\mathcal{H}^r_{\mathrm{OLGS}}(k)^2+\mathcal{H}^d_{\mathrm{OLGS}}(k)^2}\end{aligned} \tag{12}$$

where $\mathcal{C}$ is the change between two histograms.

Considering the degradations from the two aspects, the final quality is predicted as,

$$\mathcal{Q}(I^d|I^r) = \mathcal{Q}^{\alpha}_{\mathrm{SWLS}}\cdot\mathcal{Q}^{\beta}_{\mathrm{OLGS}}, \tag{13}$$

where α and β are two weighting parameters, and we set $\alpha = \beta = 1$ for simplicity in this work.

IV. EXPERIMENTAL RESULT ANALYSIS

In this section, the efficiency of the proposed model is firstly verified by the eye tracker. Then, a thorough comparison between the proposed model and these existing RR IQA methods is demonstrated. Finally, the attention mechanism is transferred into the existing RR IQA models to further verify its efficiency.

A. EFFICIENCY ANALYSIS

The HVS is extremely sensitive to changes on attended regions. And thus, the distortion on such kind of regions will cause more severely quality degradation than that on unattended regions. In order to thoroughly demonstrate the relation between the attention and the distortion, a subjective viewing experiment with the eye-tracker was designed. The SensoMotoric Instruments (SMI) RED eye tracker was adopted to record the eye movements with a sampling rate of 120 Hz. Meanwhile, a standard office environment was set up for this experiment. All the stimuli were displayed on a LED screen, which with a resolution of 1920×1080. The viewing distance between participants and the screen was about 70cm, which was slightly adapted for a stable and accurate result according to the eye-tracking software.

Twelve participants (7 males and 5 females, ranging from 20 to 30 years of age) were invited and all of them were inexperienced with image quality assessment. These participants were instructed to look at the stimuli without any task. Each stimulus was represented for 7 seconds, and followed by a gray image with 3 seconds.

After the eye-tracking experiment, the eye-tracking data was analyzed with the SMI BeGaze analysis software, and fixations were extracted and exported. Each individual fixation is drew by a circle in the image to represent the fixation point, thus creating a gaze map. Meanwhile, the fixation duration was ignored during the gaze map construction.

Distortions on attended objects will cause more severe quality degradation than that on background regions. An example is shown in Fig. 6, for which the first row shows three images (the lighthouse, the lady and the flower) distorted by the JTE noise, and the second row shows their corresponding fixation maps (where red points stand for human fixations from eye-tracking experiment and blue lines are drawn to indicate the main attention regions). As shown in Fig. 6 (a), the distortion mainly occurs on the reef (the background region). When we look at this image, most of our attentions focus on the lighthouse and the red houses (with little distortion), while limited on the background reef region, as shown in the corresponding fixation map. In other words, the distortion is mainly on the unattended background region. For Fig. 6 (b), we mainly focus our attention on the face, the jewelry, and her hands. Meanwhile, only the left hand and a small part of the jewelry are distorted. In other words, distortion presents partly on the attended regions for Fig. 6 (b). While for Fig. 6 (c), we mainly focus our attention on the flower regions. At the meantime, such regions are contaminated by obvious distortion. In summary, the distortions on the three contaminated images are different (i.e., mainly background, partly foreground, and mainly foreground), which result in different perceptual quality degradations.

The quality values for the three images are listed in Table 1. Since most of attended regions are free from distortion, Fig.6 (a) has the best subjective quality (with MOS $= 5.6191$) even under the highest noise level (with PSNR $= 23.96$). Since parts of the attended regions are distorted in Fig.6 (b), it presents a worse quality (with MOS $= 5.5610$) than Fig.6 (a), even though the noise level in Fig.6 (b) (with PSNR $= 27.598$) is much lower than that in Fig.6 (a). Fig.6 (c) has the worst quality (with MOS $= 2.8636$), though its noise level (with PSNR $= 24.855$) is slightly lower than that in Fig.6 (a). Therefore, the HVS is extremely sensitive to the distortion on the attended region, which result in much severe quality degradation.

The proposed AVCD method can accurately predict the quality degradation for such kind of image set as shown in Fig.6. A comparison with several latest RR IQA methods (i.e., RED-LOG [19], FSI [35], OSVP [17], RRED [18], and WNISM [12]) are shown in Table 1. By considering the importance of attention regions for quality prediction, the proposed AVCD based RR IQA returns an accurately quality prediction result which is consistent with the subjective perception (i.e., MOS value). However, the other RR IQA

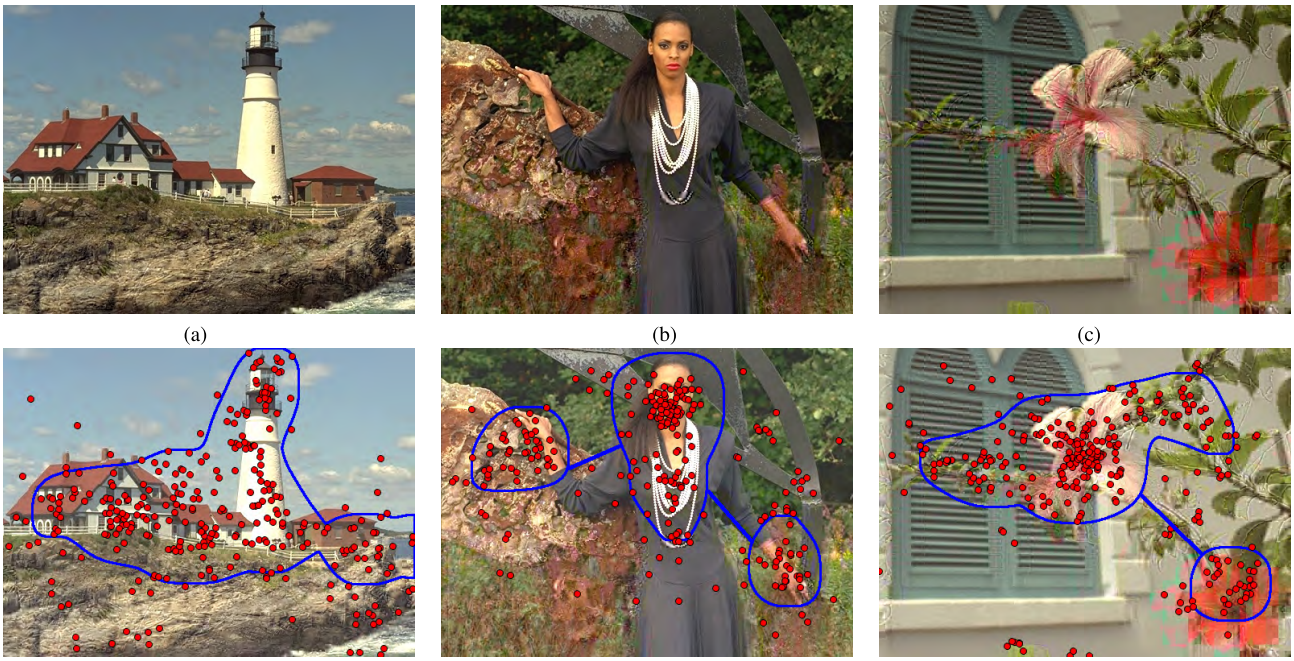


FIGURE 6. An illustration of distortions on attended and unattended regions with fixatioin maps. (a) *lighthouse*. (b) *lady*. (c) *flower*.

TABLE 1. Qualities of RR-IQAs on Fig. 6.

RR-IQAs \ IMG	*lighthouse*	*lady*	*flower*
PSNR	23.960	27.598	24.855
MOS	5.6191	5.5610	2.8636
AVCD	0.9995	0.9990	0.9676
RED-LOG	5.1002	**4.4576**	7.3732
FSI	0.0166	**0.0085**	0.0136
OSVP	0.9988	**0.9997**	0.9747
RRED	1.9675	**2.0359**	1.5640
WNISM	3.9116	**3.0026**	7.9957

TABLE 2. Qualities of RR-IQAs on Fig. 7.

RR-IQAs \ Dist	Lossy	JTE	Block-wise
MOS	5.1667	3.5294	3.1892
AVCD	0.9931	0.9821	0.9771
SWLS	0.9931	**0.9992**	0.9772
RED-LOG	4.9516	**4.8274**	5.9395
FSI	0.125	0.1639	**0.0308**
OSVP	0.9966	**0.9990**	0.9845
RRED	0.5483	**2.9969**	2.3556
WNISM	5.4465	**2.8583**	5.0715

methods show confusing results. As shown in Table 1, all of these compared RR IQA methods suggest that the *lady* image (Fig. 6 (b) which has the lowers noise level) should possess the best visual quality among the three.

Moreover, the attended region may be shifted by the severe distortion. An example is given in Fig. 7, in which the first row shows a same scene is contaminated by three types of distortion (i.e., Compression Lossy, JTE, and Block Wise), and the second row shows their corresponding fixation maps. For Fig. 7 (a), the compression lossy dose not change much on the visual content, and the fixations uniformly spread on all of these objectives. JTE in Fig. 7 (b) severely destroys the people (e.g., the man in the middle and the two kids in the right side). As a result, the attention is obviously shifted. As shown in Fig. 7 (b), much attention is attracted on the body region of the man, and less on the two kids. For Fig. 7 (c), the image is distorted by block-wise distortion. As a result, these blocked regions attract some attention. In summary, by comparing the fixation region among the three contaminated images (with a same scene), attentions are obviously shifted by these distortions.

The performance of these RR IQA methods are tested and compared within Fig. 7, and the results are listed in Table 2. Moreover, the quality prediction with only the SWLS part is also listed. As can be seen, without the part about attention shift, the SWLS cannot properly predict the quality of Fig. reffig:exp-ga (b). Since the proposed AVCD method has considered the attention shift during quality prediction, an accurate quality prediction result is given. Meanwhile, other RR IQA methods always show a different tendency from subjective perceptual results (MOS values). In summary, by taking both local attention weight and global attention shift into account, the proposed AVCD method has greatly improved the prediction accuracy and performs consistently with subjective perception.

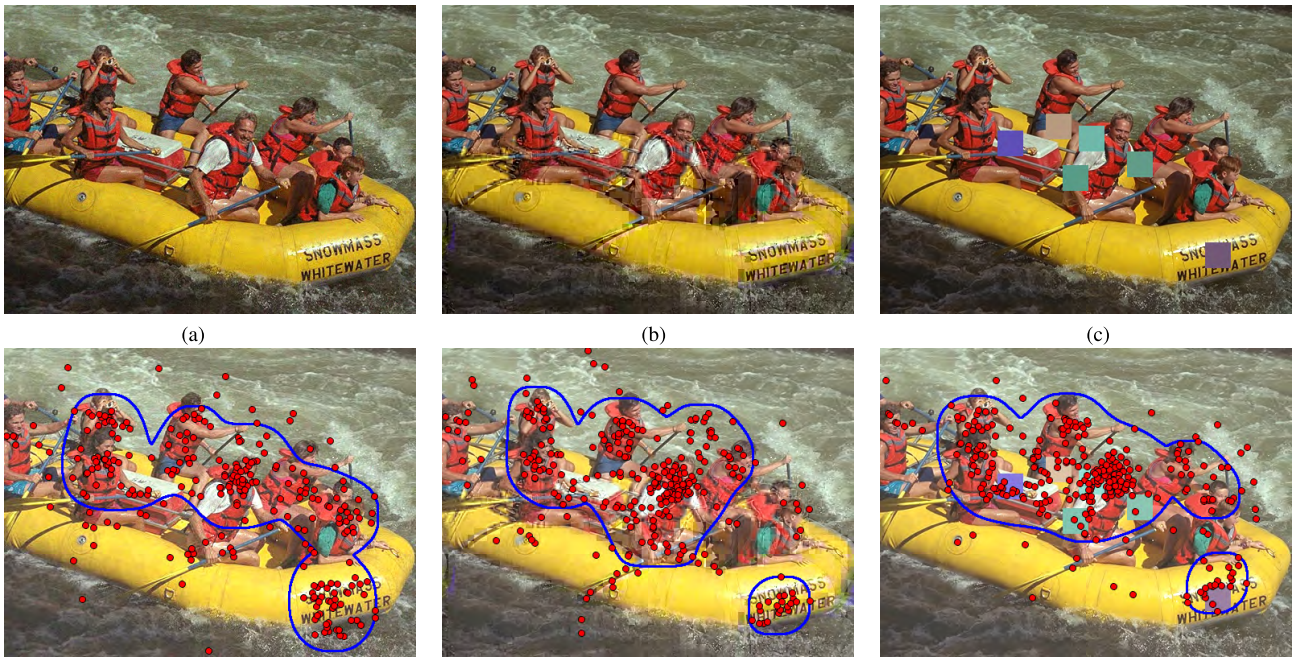


FIGURE 7. An illustration of attention shift on the same scene with fixatioin maps. (a) Compression Lossy. (b) JTE. (c) Block Wise.

B. PERFORMANCE COMPARISON ON DATABASES

In order to make a comprehensive analysis, the proposed AVCD based RR IQA is verified on three publicly available IQA databases, *CSIQ* [36], *LIVE* [37] and *TID2013* [38]. The *CSIQ* database is made up of 866 distorted images and 30 reference images with 6 distortion types; The *LIVE* database contains 779 distorted images corresponding to 29 reference images with five distortion types; and the *TID2013* database comprises 3000 distorted images of 24 types of distortion with 25 reference images. For a concrete demonstration of performance, six latest and accessible RR IQA methods (i.e., RED-LOG, FSI, OVSP, RRVIF [16], RRED and WNISM) and three classical FR IQA indices (i.e., PSNR, SSIM, and MS-SSIM [6]) are introduced for comparison.

For a quantitative performance comparison of RR IQA algorithms, three widely used performance metrics are employed, which are Pearson linear correlation coefficient (PLCC), Spearman rank-order correlation coefficient (SRCC), and root mean squared error (RMSE), to measure the correlation between predicted quality values and ground truth (i.e., the subjective mean opinion score (MOS) or the difference of MOS (DMOS)). PLCC measures the accuracy of prediction, while SRCC evaluates the monotonicity. A better IQA algorithm will result in a higher PLCC and SRCC, and a lower RMSE. Before the evaluation, the predicted quality value Q is always regressed with a logistic nonlinearity mapping function,

$$\mathrm{Q}' = \beta_1\left(\frac{1}{2} - \frac{1}{1 + exp(\beta_2(\mathrm{Q} - \beta_3))}\right) + \beta_4\mathrm{Q} + \beta_5, \quad (14)$$

where Q' is the regressed value, and $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$, are the parameters used in the regression model.

The performance of these IQA methods on each individual distortion type is firstly compared, and the comparison result is listed in Table 3 (only the SRCC value is given for simplicity, and the best results are highlighted in bold). As can be seen, with very limited values for all of the RR IQA algorithms (18, 9, 1, 9, 18 and 18 scalars for AVCD, RED-LOG, FSI, OSVP, RRED and WNISM, respectively), the proposed AVCD achieves the best performance for JPEG compression distortion on three databases, and shows a good performance on white noise (best on *LIVE* and *TID2013*, but a little worse on *CSIQ*). Moreover, for the *TID2013* which contains 24 distortion type (complex and challenge for IQA algorithms), the AVCD still shows a remarkable advantages over other RR IQA methods.

Besides, by comparing with the classical FR IQA metrics, the proposed AVCD also shows some advantages. The performances of the proposed AVCD on many distortion types (e.g., JPEG, JP2K, Blur, JTE, JPEG2000 transmission errors, etc.) have been beyond classical PSNR and SSIM, and approximating to MS-SSIM. Therefore, even with a quite small amount of the reference information (18 values), the AVCD can outperform PSNR and SSIM in many cases, and also show a comparable performance against the MS-SSIM in some cases.

In order to give an intuitive view, the hit-counts of RR IQA methods on these distortion types of each database are counted in Table 4. As can be seen, the proposed AVCD gets 2 (performs the best on 2 distortion types) out of 6 distortion types on *CSIQ*, 2 out of 5 distortion types on *LIVE*, and 14 out of 24 distortion types on *TID2013*. Meanwhile, the proposed AVCD possesses the highest hit-count number on each database, and also for the whole three databases (18 times,

TABLE 3. Performance (SRCC) evaluation of RR-IQA models on each individual distortion type.

Database	Distortion Type	RR							FR		
		AVCD	RED-LOG	FSI	OSVP	RRVIF	RRED	WNISM	PSNR	SSIM	MS-SSIM
	No. of Scalars	18	6	1	9	2	18	18	N	N	N
CSIQ	AWGN	0.880	0.874	0.849	0.891	**0.947**	0.913	0.822	0.936	0.925	0.947
	JPEG	**0.956**	0.926	0.951	0.951	0.838	0.945	0.903	0.888	0.922	0.963
	JP2K	0.937	0.949	0.934	0.932	0.944	**0.955**	0.940	0.936	0.921	0.968
	FNoise	0.889	0.863	0.814	0.891	**0.907**	0.905	0.802	0.934	0.892	0.933
	BLUR	**0.965**	0.939	0.963	0.959	0.851	0.964	0.918	0.929	0.925	0.971
	Contrast	0.905	0.930	**0.955**	0.929	0.893	0.939	0.910	0.862	0.740	0.953
LIVE	JP2K	0.937	0.952	0.902	0.896	**0.955**	0.948	0.939	0.895	0.936	0.963
	JPEG	**0.967**	0.950	0.962	0.964	0.877	0.949	0.919	0.881	0.944	0.981
	White-Noise	**0.976**	0.931	0.923	0.975	0.972	0.946	0.872	0.985	0.963	0.973
	Gaussian-Blur	0.939	0.935	0.964	0.928	0.753	**0.965**	0.920	0.782	0.894	0.954
	Fastfading	0.935	**0.964**	0.886	0.940	0.901	0.918	0.939	0.891	0.941	0.947
TID2013	gauss noise	**0.904**	0.875	0.709	0.856	-	0.825	0.686	0.929	0.853	0.866
	Additive noise	0.784	**0.818**	0.599	0.721	-	0.789	0.616	0.898	0.774	0.773
	Spatially correlated noise	**0.901**	0.802	0.703	0.893	-	0.803	0.673	0.920	0.862	0.854
	Masked noise	0.627	0.504	**0.722**	0.448	-	0.650	0.625	0.832	0.810	0.807
	High frequency noise	0.884	**0.907**	0.771	0.819	-	0.890	0.747	0.914	0.847	0.865
	Impulse noise	**0.898**	0.564	0.704	0.788	-	0.687	0.677	0.897	0.799	0.763
	Quantization noise	0.814	0.793	0.262	0.723	-	**0.862**	0.656	0.881	0.806	0.871
	Gaussian blur	0.963	0.895	0.950	0.903	-	**0.967**	0.924	0.914	0.963	0.967
	Image denoising	**0.928**	0.880	0.831	0.901	-	0.915	0.854	0.948	0.910	0.927
	JPEG compression	**0.916**	0.898	0.858	0.889	-	0.914	0.843	0.919	0.910	0.927
	JPEG2000 compression	0.933	0.932	0.906	0.895	-	**0.939**	0.916	0.884	0.905	0.950
	JPEG transmission errors	**0.874**	0.807	0.365	0.816	-	0.773	0.819	0.813	0.818	0.848
	JPEG2000 transmission errors	**0.904**	0.826	0.642	0.869	-	0.742	0.760	0.888	0.877	0.889
	Non eccentricity pattern noise	**0.757**	0.479	0.446	0.740	-	0.750	0.502	0.686	0.759	0.797
	Block-wise distortions	0.155	0.469	**0.559**	0.334	-	0.431	0.286	0.099	0.617	0.502
	Mean shift	**0.767**	0.669	0.617	0.537	-	0.576	0.485	0.767	0.777	0.791
	Contrast change	0.256	0.539	**0.568**	0.526	-	0.488	0.585	0.440	0.364	0.463
	Change of color saturation	**0.634**	0.019	0.243	0.173	-	0.242	0.277	0.101	0.406	0.410
	Multiplicative Gaussian noise	**0.888**	0.807	0.638	0.832	-	0.756	0.595	0.891	0.775	0.779
	Comfort noise	**0.892**	0.858	0.581	0.818	-	0.872	0.709	0.841	0.819	0.853
	Lossy compression	0.820	0.906	0.402	0.825	-	**0.927**	0.722	0.914	0.911	0.907
	Image color quantization with dither	**0.900**	0.656	0.281	0.830	-	0.873	0.482	0.927	0.789	0.855
	Chromatic aberrations	**0.885**	0.803	0.866	0.831	-	0.883	0.858	0.887	0.889	0.878
	Sparse sampling and reconstruction	0.932	0.888	0.883	0.923	-	**0.940**	0.920	0.904	0.903	0.948

TABLE 4. Hit-count on each database.

Database	AVCD	RED-LOG	FSI	OSVP	RRVIF	RRED	WNISM
CSIQ	2	0	1	0	2	1	0
LIVE	2	1	0	0	1	1	0
TID2013	14	2	3	0	-	5	0
Overall	18	3	4	0	3	7	0

while the second one possesses 7 times), which further confirms that the proposed AVCD shows a remarkable advantage against other RR IQA methods on each individual distortion type.

Moreover, the overall performance on the whole database of these RR IQA methods are compared, and their corresponding PLCC, SRCC, and RMSE values are listed in Table 5. On the *CSIQ* database, AVCD achieve the highest PLCC and SRCC, and the lowest RMSE, which demonstrates that AVCD not only outperforms other RR IQA algorithms, but also shows better than the classical MS-SSIM in FR IQA method. On the *LIVE* database, AVCD performs a slightly worse than RED-LOG, but still has an advantage over the other RR IQA methods, as well as PSNR and SSIM. As for the *TID2013* database, the proposed AVCD still holds a dominant performance, and only worse than MS-SSIM. Additionally, the weighted average performances (weighted mean value of PLCC and SRCC) are computed and listed in the bottle of Table 5. As can be seen, the proposed AVCD method outperforms (has a remarkable improvement on both PLCC and SRCC values) the other RR IQA methods. Meanwhile, by comparing the proposed AVCD method with the FR IQA methods, it can be seen that though only a small amount of reference information is used, the proposed AVCD method is comparable with the FR IQA metrics: has larger weighted PLCC and SRCC values than that of PSNR and SSIM, and even achieves a larger SRCC value than the MS-SSIM. In summary, from the comparison results on the three databases it can be seen that the proposed AVCD outperforms the existing RR IQA methods, and is comparable with the classical FR IQA methods.

In addition, the statistical significance of the proposed AVCD is evaluated to further demonstrate the effectiveness. The F-test [39] is introduced to compute the statistical significance of AVCD against these IQA methods. For the computation of F-test, the residual between the quality prediction from a certain IQA model (after nonlinear mapping) and the ground truth is calculated firstly. With the variance of the residual, a ratio F is obtained from the residual variances of two IQA models on the same database. Finally, the judging threshold $F_{critical}$ (which is depended on the number of residuals and the level of confidence we expect) is used for

TABLE 5. IQA performance comparison on three databases.

Database	Algo. / Crit.	RR							FR		
		AVCD	RED-LOG	FSI	OSVP	RRVIF	RRED	WNISM	PSNR	SSIM	MS-SSIM
CSIQ (866)	PLCC	**0.913**	0.856	0.807	0.843	0.598	0.758	0.737	0.800	0.815	0.899
	SRCC	**0.914**	0.858	0.781	0.849	0.633	0.777	0.757	0.800	0.838	0.913
	RMSE	**0.107**	0.136	0.155	0.141	0.210	0.172	0.178	0.158	0.152	0.115
LIVE (779)	PLCC	0.907	**0.937**	0.881	0.862	0.722	0.830	0.743	0.872	0.904	0.941
	SRCC	0.916	**0.946**	0.883	0.867	0.738	0.834	0.755	0.876	0.910	0.951
	RMSE	11.488	**9.522**	12.931	13.838	18.900	15.244	18.276	13.360	11.669	9.259
TID2013 (3000)	PLCC	**0.800**	0.746	0.584	0.724	0.577	0.748	0.629	0.702	0.686	0.831
	SRCC	**0.800**	0.698	0.398	0.654	0.451	0.689	0.523	0.703	0.627	0.786
	RMSE	**0.744**	0.826	1.007	0.856	1.013	0.822	0.964	0.883	0.902	0.69
Weighted Average	PLCC	**0.839**	0.799	0.675	0.769	0.605	0.764	0.668	0.749	0.747	0.862
	SRCC	**0.841**	0.769	0.551	0.726	0.533	0.730	0.606	0.750	0.714	0.837

TABLE 6. Performance comparison with F-test (statistical significance with 95% confidence).

DB \ Algo.	RR						FR		
	RED-LOG	FSI	OSVP	RRVIF	RRED	WNISM	PSNR	SSIM	MS-SSIM
CSIQ	1	1	1	1	1	1	1	1	1
LIVE	-1	1	1	1	1	1	1	0	-1
TID2013	1	1	1	1	1	1	1	1	-1

judgement: if $F > F_{critical}$, the IQA model on the denominator of the ratio F is better than the one on the numerator; if $F < \frac{1}{F_{critical}}$, the IQA model on the denominator of the ratio F is worse than the one on the numerator; otherwise, the two models are statistically indistinguishable.

In this work, the confidence is set as 95%, and the statistical significance between the proposed AVCD and these IQA models are shown in Table 6. Symbols like '1', '0', and '−1' in Table 6 means that the proposed AVCD is better ('1'), indistinguishable ('0'), and worse ('−1') than the corresponding method on the current database, respectively. By comparing with these RR IQA methods (i.e., RED-LOG, FSI, OSVP, RRVIF, RRED, and WNISM), we can see that almost all of the values are ''1'' except for the one against RED-LOG under the LIVE database. Therefore, the proposed AVCD method performs significantly better than the other RR IQA methods on all of the three databases (except for the RED-LOG on the LIVE database). Moreover, by comparing with the three FR IQA metrics (i.e., PSNR, SSIM, and MS-SSIM), the proposed AVCD performs significantly better than PSNR on all of the three databases, performs significantly better than SSIM on CSIQ and TID2013, and performs comparably with MS-SSIM. The comparison results from F-test is consistent with that from these three correlation criteria (within Table. 5). With the above analysis we can conclude that the proposed AVCD method performs better than these existing RR IQA metrics, and has a comparable performance with the classical FR IQA metrics.

C. THE FUNCTION OF ATTENTION ON RR-IQA MODELS

The attention mechanism is extended to these existing IQA models to further verify its effectiveness. Since the visual attention reveals the inner processing of the HVS, it will always benefit the IQA models (make these IQA models perform more consistently with subjective perception). The OLGS part of the proposed model is independent of visual structure extraction, which can be directly added into these existing RR IQA methods. Therefore, the OLGS features are extracted, and then combined with the original features extracted by these RR IQA methods (i.e., RED-LOG, FSI, OSVP, RRED, and WNSIM) for quality prediction.

TABLE 7. RR-IQA performance comparison after introducing attention feature on TID2013.

Crit.	PLCC		SRCC	
Algo.	Original	After	Original	After
RED-LOG	0.746	**0.750**	0.698	**0.709**
FSI	0.584	**0.610**	0.398	**0.587**
OSVP	0.724	**0.784**	0.654	**0.771**
RRED	0.748	**0.750**	0.689	**0.702**
WNISM	0.629	**0.672**	0.523	**0.664**

The performances comparison of these methods (original) against their extension (after) on the TID2013 database is listed in Table 7. As can be seen, both the PLCC and SRCC values are increased on all of these methods, especially for FSI (the SRCC value is increased from 0.398 to 0.587) and WNISM (the SRCC value is increased from 0.523 to 0.664) which achieve a remarkable improvement. Therefore, the performances of these existing models are improved by combining with OLGS. So far, we can conclude that the proposed AVCD method can not only perform consistently with the subjective perception and outperform these existing

RR IQA methods, but also can benefit these existing methods by adding the attention based features (i.e., OLGS).

V. CONCLUSION

In this work, a novel RR IQA method with visual attention has been proposed. As an important visual mechanism, attention reveals the content/region selection during perception. And thus, the effect of distortion on attention is thoroughly analyzed. The HVS is extremely sensitive to changes on the attended regions, and distortions on such regions will cause more severely quality degradation than that on the unattended regions. Meanwhile, the distortion degrades the structure of the object, which may cause attention shift (change the attended regions between the reference image and distorted image).

The attention has been estimated with saliency for quality prediction. The visual content of each region has been firstly weighted by its saliency, and a SWLS based histogram has been created for the local structure degradation measurement. Then, the attention distribution has been counted, and an OLGS based histogram has been built for global attention shift measurement. Finally, considering the degradation from SWLS and OLGS, a novel ACVD RR IQA method has been proposed. Experimental results have demonstrated that the proposed ACVD method performs highly consistent to the subjective perception, and it can be extended to improve these existing RR IQA models.

REFERENCES

[1] D. Liu, F. Li, and H. Song, "Image quality assessment using regularity of color distribution," *IEEE Access*, vol. 4, pp. 4478–4483, 2016.

[2] K. Gu, G. Zhai, W. Lin, and M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 284–297, Jan. 2016.

[3] J. Wu, W. Lin, G. Shi, and A. Liu, "Perceptual quality metric with internal generative mechanism," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 43–54, Jan. 2013.

[4] X. Min, K. Gu, G. Zhai, M. Hu, and X. Yang, "Saliency-induced reduced-reference quality index for natural scene and screen content images," *Signal Process.*, vol. 145, pp. 127–136, Apr. 2018.

[5] L. Li, Y. Yan, Z. Lu, J. Wu, K. Gu, and S. Wang, "No-reference quality assessment of deblurred images based on natural scene statistics," *IEEE Access*, vol. 5, pp. 2163–2171, 2017.

[6] 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.

[7] X. Min, K. Ma, K. Gu, G. Zhai, Z. Wang, and W. Lin, "Unified blind quality assessment of compressed natural, graphic, and screen content images," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5462–5474, Nov. 2017.

[8] J. Wu, W. Lin, G. Shi, Y. Zhang, W. Dong, and Z. Chen, "Visual orientation selectivity based structure description," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4602–4613, Nov. 2015.

[9] R. A. Manap and L. Shao, "Non-distortion-specific no-reference image quality assessment: A survey," *Inf. Sci.*, vol. 301, pp. 141–160, Apr. 2015.

[10] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annu. Rev. Neurosci.*, vol. 24, no. 1, pp. 1193–1216, 2001.

[11] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[12] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," *Proc. SPIE*, vol. 5666, pp. 149–159, Mar. 2005.

[13] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 202–211, Apr. 2009.

[14] L. Ma, S. Li, F. Zhang, and K. N. Ngan, "Reduced-reference image quality assessment using reorganized DCT-based image representation," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 824–829, Aug. 2011.

[15] X. Gao, W. Lu, D. Tao, and X. Li, "Image quality assessment based on multiscale geometric analysis," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1409–1423, Jul. 2009.

[16] J. Wu, W. Lin, G. Shi, and A. Liu, "Reduced-reference image quality assessment with visual information fidelity," *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1700–1705, Nov. 2013.

[17] J. Wu, W. Lin, G. Shi, L. Li, and Y. Fang, "Orientation selectivity based visual pattern for reduced-reference image quality assessment," *Inf. Sci.*, vol. 351, pp. 18–29, Jul. 2016.

[18] R. Soundararajan and A. C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 517–526, Feb. 2012.

[19] S. Golestaneh and L. J. Karam, "Reduced-reference quality assessment based on the entropy of DWT coefficients of locally weighted gradient magnitudes," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5293–5303, Nov. 2016.

[20] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[21] Y. Fang, W. Lin, B.-S. Lee, C.-T. Lau, Z. Chen, and C.-W. Lin, "Bottom-up saliency detection model based on human visual sensitivity and amplitude spectrum," *IEEE Trans. Multimedia*, vol. 14, no. 1, pp. 187–198, Feb. 2012.

[22] W. Zhang, A. Borji, Z. Wang, P. Le Callet, and H. Liu, "The application of visual saliency models in objective image quality assessment: A statistical evaluation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1266–1278, Jun. 2016.

[23] W. Zhang and H. Liu, "Toward a reliable collection of eye-tracking data for image quality research: Challenges, solutions, and applications," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2424–2437, May 2017.

[24] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1. Jun. 2005, pp. 886–893.

[25] M. I. Posner and S. E. Petersen, "The attention system of the human brain," *Annu. Rev. Neurosci.*, vol. 13, no. 1, pp. 25–42, 1990.

[26] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[27] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.

[28] P. A. van der Helm and E. L. J. Leeuwenberg, "Accessibility: A criterion for regularity and hierarchy in visual pattern codes," *J. Math. Psychol.*, vol. 35, no. 2, pp. 151–213, May 1991.

[29] U. Grenander and M. Miller, *Pattern Theory: From Representation to Inference*. New York, NY, USA: Oxford Univ. Press, 2007.

[30] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol.*, vol. 160, no. 1, pp. 106–154, 1962.

[31] D. Hansel and C. van Vreeswijk, "The mechanism of orientation selectivity in primary visual cortex without a functional map," *J. Neurosci.*, vol. 32, no. 12, pp. 4049–4064, 2012.

[32] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in Neural Information Processing Systems*, B. Schölkopf, J. C. Platt, and T. Hoffman, Eds. Cambridge, MA, USA: MIT Press, 2007, pp. 545–552.

[33] J. Zhang, K. Huang, Y. Yu, and T. Tan, "Boosted local structured HOG-LBP for object localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 1393–1400.

[34] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Proc. IEEE Conf. IEEE Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2879–2886.

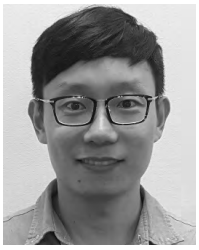
[35] Y. Liu, G. Zhai, K. Gu, X. Liu, D. Zhao, and W. Gao, "Reduced-reference image quality assessment in free-energy principle and sparse representation," *IEEE Trans. Multimedia*, vol. 20, no. 2, pp. 379–391, Feb. 2018.

[36] E. C. Larson and D. M. Chandler. (2004). *Categorical Image Quality (CSIQ) Database*. [Online]. Available: http://vision.okstate.edu/csiq

[37] H. R. Sheikh, K. Seshadrinathan, A. K. Moorthy, Z. Wang, A. C. Bovik, and L. K. Cormack. (2006). *Image and Video Quality Assessment Research at Live*. [Online]. Available: http://live.ece.utexas.edu/research/quality/
[38] N. Ponomarenko *et al.*, "Color image database TID2013: Peculiarities and preliminary results," in *Proc. 4th IEEE Eur. Workshop Vis. Inf. Process. (EUVIP)*, Jun. 2013, pp. 106–111.
[39] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*. Boca Raton, FL, USA: CRC Press, 2003.

JINJIAN WU received the B.Sc. and Ph.D. degrees from Xidian University, Xi'an, China, in 2008 and 2013, respectively. From 2011 to 2013, he was a Research Assistant at Nanyang Technological University, Singapore, where he was a Postdoctoral Research Fellow from 2013 to 2014. From 2013 to 2015, he was a Lecturer at Xidian University, where he has been an Associated Professor at the School of Electronic Engineering since 2015. His research interests include visual perceptual modeling, saliency estimation, quality evaluation, and just noticeable difference estimation. He received the Best Student Paper Award from ISCAS 2013. He served as the Special Section Chair for the IEEE Visual Communications and Image Processing in 2017 and the Section Chair/the Organizer/a TPC Member for ICME2014–2015, PCM2015–2016, ICIP2015, and QoMEX2016.

YONGXU LIU received the B.S. degree from Xidian University, Xi'an, China, in 2016, where he is currently working toward the M.S. degree. His research interests include visual perceptual modeling and image/video quality assessment.



LEIDA LI received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2004 and 2009, respectively. In 2008, he joined the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Taiwan, as a Visiting Ph.D. Student. From 2014 to 2015, he was a Visiting Research Fellow at the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he is currently a Senior Research Fellow. He is also a Full Professor at the School of Information and Control Engineering, China University of Mining and Technology, China. His research interests include multimedia quality assessment, information hiding, and image forensics.

GUANGMING SHI (SM'10) received the B.S. degree in automatic control, the M.S. degree in computer control, and the Ph.D. degree in electronic information technology from Xidian University, Xi'an, China, in 1985, 1988, and 2002, respectively. He had studied at the University of Illinois and The University of Hong Kong. Since 2003, he has been a Professor at the School of Electronic Engineering, Xidian University, where he is currently the Academic Leader on circuits and systems. He has authored or coauthored over 200 papers in journals and conferences. His research interests include compressed sensing, brain cognition theory, multirate filter banks, image denoising, low-bitrate image and video coding, and the implementation of algorithms for intelligent signal processing. He received the Cheung Kong Scholar Chair Professorship by the Ministry of Education in 2012. He served as the Chair for the 90th MPEG and 50th JPEG of the international standards organization and the Technical Program Chair for FSKD'06, VSPC 2009, the IEEE PCM 2009, SPIE VCIP 2010, and the IEEE ISCAS 2013.