No-Reference Quality Evaluator of Transparently Encrypted Images

Guanghui Yue, Chunping Hou, Ke Gu, Tianwei Zhou, and Hantao Liu, *Member*, *IEEE*

***Abstract*—In the past years, various encrypted algorithms have been proposed to fully or partially protect the multimedia content in view of practical applications. In the context of digital TV broadcasting, transparent encryption only protects partial content and fulfills both security and quality requirements. To date, only a few reference-based works have been reported to evaluate the quality of transparently encrypted images. However, these works are incapable of reference-unavailable conditions. In this paper, we conduct the first attempt that proposes a novel quality evaluator in the absence of reference images. The key strategy of proposed metric lies in extracting features by considering the motivation of transparently encrypted images. Specifically, given that encrypted images prevent the content to be easily recognized, several features, including correlation coefficient, information entropy and intensity statistic, are preliminarily extracted to estimate visual recognizability. Meanwhile, considering that encrypted images avoid to be in extremely low-quality, we also capture many features to measure the distortions on multiple quality-sensitive image attributes, such as naturalness, structure, and texture. Finally, the quality evaluator is built by bridging all extracted features and corresponding quality scores via a regression module. Experimental results demonstrate that the proposed method is superior to the mainstream no-reference quality evaluation methods designed for synthetically distorted images and possesses a close approximation to state-of-the-art reference-based methods designed for encrypted images.**

***Index Terms*—Quality evaluation, visual security, encrypted image, transparent encryption, no-reference.**

I. Introduction

NOWADAYS, image data is under explosive growth, storage, and propagation due to widely pervasive usage of cameras, cloud storage devices and convenient availability of network access [1], [2]. However, the copyright protection becomes increasingly difficult due to the wanton propagation and even the illegal attacks in transmission network. Multimedia copyright crisis gradually emerges as a headache problem that challenges the ethics, law, and technology.

In recent years, numerous security schemes have been developed to protect the copyright for both individual privacy and commercial confidentiality [1], [3]–[6]. Among those schemes, encryption is a generally acknowledged approach. Specifically, it protects image content by inserting ciphertext, which can be unencrypted by an authorized user with plaintext. For multimedia data, the encryption scheme is specifically designed to meet the particular application environment [7], [8]. One well-known application is the digital TV broadcasting, where transparent encryption are needful. In this application, on the one hand, the encrypted images should avoid to be in high-quality for ensuring the security to unauthorized viewers. On the other hand, the encrypted images should also be in relative high-quality to attract more unauthorized users to buy the copyright [9]. It is apparent that the visual quality is the key-defining property in the definition of security. Therefore, evaluating the quality of transparently encrypted images is meaningful to the encryption and broadcasting systems. In general, image quality can be evaluated subjectively and objectively. Subjective evaluation is the most reliable method as it, to the greatest extent, reflects the intuitive subjective feeling. However, it has intrinsical limitations, e.g., labor-intensive, time-consuming, and expensive, thereby hindering its widely practical applications. Hence, designing objective solutions has attracted increasing attention from scholars towards automatically evaluating visual quality of encrypted multimedia [10]–[12].

In the literature, there have been several discussions on visual quality evaluation [13]–[21]. Generally, existing metrics can be mainly classified into two categories, i.e., reference-based metrics and reference-free metrics (also denoted as no-reference (NR), blind, or referenceless metric). To date, a great variety of reference-based metrics have been reported to evaluate the synthetically distorted images with good performance by computing the relationship between the reference image and its distorted version [22]–[26]. However, as discussed in Section III-B, the encrypted image possesses different characteristics compared to the one generated by synthetical distortions, such as contrast change, blockiness, noise, and so on. Traditional metrics are unsuitable for evaluating the quality of encrypted images. Therefore, it is highly desired to design specific metrics for encrypted images. Until now, only a few specific metrics have been reported by measuring quality-sensitive image attributes, like structure, luminance, and edge. Since the content of a low-quality encrypted image is usually

This work was supported partly by the National Natural Science Foundation of China under Grants (Nos. 61731003, 61520106002, 61871274, 61801305, and 81571758), National Natural Science Foundation of Guangdong Province (Nos. 2017A030313377 and 2016A030313047), Shenzhen Peacock Plan (No. KQTD2016053112051497), and Shenzhen Key Basic Research Project (Nos. JCYJ20170818142347251 and JCYJ20170818094109846). (Corresponding author: Chunping Hou.)

G. Yue is with the National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518060, China. (e-mail: guanghuiyue.doctor@gmail.com.)

C. Hou is with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China. (e-mail: hcp@tju.edu.cn.)

K. Gu is with Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China. (e-mail: guke.doctor@gmail.com).

T. Zhou is with the College of Management, Shenzhen University, Shenzhen 518060, China. (e-mail: tianweizhoudr@gmail.com.)

H. Liu is with the School of Computer Science and Informatics, Cardiff University, Cardiff CF24 3AA, U.K. (e-mail: liuh35@cardiff.ac.uk.)

hard to be recognized, these metrics believe that the visual quality reflects visual security to some extent and name the quality evaluation as security evaluation. For instance, Mao and Wu [27] designed two quantitative visual security metrics, namely ESS and LSS, based on edge similarity and luminance similarity between the reference and encrypted images. In [12], three metrics were proposed based on structure similarity, image entropy, and spatial intensity correlation, respectively. Likewise, Yao *et al*. [28] captured local pixel similarity correlation between the corrupted image and its reference one to quantify visual security. Jolfaei and Mirghadri [29] designed three approaches according to the pixel value change, position change, and value-position change. Tong *et al*. [11] proposed a local feature based visual security (LFBVS) metric. To be specific, both gradient similarity and luminance similarity between the reference image and the processed one were primarily extracted, followed by a error pooling procedure to combine two similarities. In [10], Xiang *et al*. proposed a novel visual security index by adaptively combining texture and edge similarities. Specifically, the texture similarity was estimated as the distance of features, extracted in gray level co-occurrence matrix (GLCM), between the reference and corrupted images, while the gradient similarity was estimated by means of multi-threshold edge maps.

As can be seen, all these metrics mentioned above are reference-based, which indicates that they merely suit for reference available situations, yet are impossible for the broader application range. In practice, we cannot always obtain the reference information. One well-known example is the application scenario of digital TV broadcasting. The broadcaster usually provides a low-quality version of the broadcast program to the users. If the quality of the program is too low, it will not attract users, thereby losing its business value. In general, monitoring the viewership traffic and perceptual quality at the receiver can provide a timely reply about the attractiveness to audiences, based on which the broadcaster can dynamically change the quality level of the broadcast program source to optimize the business value. Therefore, an effective quality evaluator is needful to monitor the visual quality of the program at the receiver. However, the reference TV programme is not transmitted, and thus, existing reference-based methods are incompetent to such a task. In this sense, a NR metric is highly desired.

In this paper, we conduct the first attempt that proposes a NR metric to evaluate visual quality of encrypted images. Compared with existing metrics, the proposed metric provides several contributions: 1) This is a pioneering attempt that evaluates visual quality of encrypted images without access to reference images. Given that transparent encryption considers both security protection and quality attractiveness, the proposed metric is designed by extracting features from measurements of visual recognizability and quality-related image attributes. All extracted features are integrated via a regression model to indicate the ultimate visual quality. 2) To estimate visual recognizability, we propose a simple yet effective way, i.e., extracting correlation coefficients, information entropy, and intensity statistics. Inspired by traditional quality evaluation metrics, some quality-related features are captured by analyzing image naturalness, structure and texture information. Specifically, in the absence of reference images, structural information is captured by computing correlation among multi-scale decomposed high order phase consistency maps. This measurement is different from existing works. 3) Experiments on the publicly available database demonstrate that the proposed method possesses a close approximation to state-of-the-art reference-based quality evaluation methods designed for encrypted images and obtains outstanding performance compared to mainstream NR quality evaluation methods designed for synthetically distorted images.

The rest of this article is organized as follows. Section II presents the brief overview of the NR quality evaluation metrics and provides the motivation of this article. In Section III, we detail the proposed NR metric for encrypted images. In Section IV, we conduct extensive experiments and insightful analyses are provided based on the results. Finally, the conclusion is drawn in Section V.

II. Related Works and Motivation

A. No-Reference Visual Quality Evaluation

At present, many NR visual quality evaluation metrics have been reported, and they can be broadly divided into specific-purpose and general-purpose metrics [30]–[35]. In case of a known distortion, such as blockiness, blurriness/sharpness, and contrast change, some specific-purpose quality evaluation metrics are designed. For instance, Li *et al*. [36] proposed a NR evaluation method for blocking artifacts via analysis of Tchebichef moment. Considering the properties of human visual system (HVS), Li *et al*. [20] also designed a novel sharpness evaluation metric by learning multi-scale features in spectral and spatial domains. Wang *et al*. [37] extracted gradient statistical properties to indicate image blurriness, and proposed a NR image blurriness evaluator by utilizing extreme learning machine. In [38], a NR contrast evaluation method was designed based on analysis of information. Fang *et al*. [39] took moment and entropy features into account and built a model based on natural scene statistics to estimate image contrast.

Different from specific-purpose metrics, general-purpose metrics aim to simultaneously quantify multiple synthetical distortions by one quality evaluation metric. Until now, various general-purpose NR metrics have been reported. According to properties of natural images, representative works have been proposed by analyzing natural scene statistics in spatial and transformed domains [40], [41]. Besides, other categories of features are also utilized. For example, Liu *et al*. [42] captured local image entropy in both spectral and spatial domains to estimate the perceptual quality. Li *et al*. [31] extracted local binary pattern (LBP) and luminance histogram to measure image distortions. Inspired by HVS characteristics, Fang *et al*. [43] designed a NR metric for screen content by extracting luminance-relevant and texture-relevant features. In [44], Gu *et al*. utilized three groups of features inspired by classical HVS and free-energy-based brain theory as well as image naturalness. Then, the image distortion intensity was estimated via a regression function. In addition, some metrics

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2913315, IEEE Transactions on Multimedia

were also designed for multiply distorted images [34], [35], [45]. For instance, Li *et al*. [45] took the gradient-weight histogram of LBP map as the quality-sensitive feature and built a prediction model (namely GWH-GLBP) for multiply-distorted images via support vector regression (SVR). In [35], a new model was proposed by extracting structural features from the improved LBP descriptor, which was modified in consideration of HVS characteristics. Likewise, inspired by the HVS, Zhou *et al*. [34] extracted two groups of quality-sensitive features to evaluate multiply-distorted images. More specifically, the first group aimed to represent spatial contrast by extracting structure-related features, and the second group was used to reflect spatial resolution by extracting nonlocal self-similarity statistics.

B. Motivation

In general, transparent encryption should meet both security and quality requirements. Due to the specific requirements, the encrypted images show different appearances compared with the synthetically distorted images. Specifically, when an image is encrypted at mild or moderate degree of strength, its content can still be discernible with local or global structure distortions. One example is shown in Fig. 1(b). It is clear that, some local regions are corrupted slightly in comparison with Fig. 1(a) regarding structure, texture, luminance, etc. The whole image is clear and possesses similar appearance as that processed by synthetical distortions, such as blockiness, blurriness. In such a situation, the image is of relatively high visual quality. Traditional evaluation metrics designed for synthetically distorted images may be competent for evaluating encrypted images to some extent. However, in the case of high-strength encryption, the image content (especially the main structure) suffers from dramatic changes and therefore is hard to be intelligible, causing unpleasant visual experience and almost losing its business value. In contrast, an image is corrupted but still discernible even with severe synthetical distortions. Fig. 1 presents an intuitive comparison between the encrypted image and compressed image. Fig. 1(a) is the reference image, and Fig. 1(c) and Fig. 1(d) are its encrypted and compressed versions, respectively. Clearly, the high-strength encryption severely affects the content of the reference image, and the encrypted image is hard to be recognized without hint. By contrast, compression distorts local structural and textural information with unnatural global perception. Despite this, it does not affect the overall recognition of image content.

From the above analysis, it is clear that the encrypted image presents different characteristics as the encrypted strength changed. An effective quality evaluation method for encrypted images should carefully consider these characteristics. On the one hand, some measures are needful to estimate the visual recognizability because the contents are hard to be recognized in the extreme low-quality case. On the other hand, it requires to draw lessons from traditional quality evaluation metrics that measure the image quality by analyzing quality-sensitive attributes. However, all of existing quality evaluation metrics designed for encrypted images mainly extract features, such as information entropy, pixel spatial correlation, to estimate image recognizability, but ignore the estimation of distortions on image attributes. Actually, the quality evaluation task can also profit from features that reflect the image attributes, e.g., structure and texture. In the literature, there are a great number of NR metrics designed for evaluating synthetically distorted images from measuring structure and texture. Unfortunately, they only achieve limited performance in evaluating the encrypted images [9]. This may be attributed to that they ignore the measurement of visual recognizability. Overall, existing methods are insufficient for effectively evaluating the encrypted images. As stated in Introduction (Section I), a NR metric is more favored due to its broader application range. Motivated by the above analyses, we propose a NR quality evaluation metric for encrypted images by measuring both visual recognizability and image attributes.

III. The Proposed Metric

Fig. 2 illustrates the framework of the proposed metric, which consists of the training stage and testing stage. In the training stage, many features are extracted first, followed by the construction of a regression model, which reflects the relationship between features and quality scores. Then, the perceptual quality score of a test image can be estimated by feeding its extracted features into the pre-trained regression model. Undoubtedly, feature extraction is the key point of the proposed metric. In this paper, we extract sufficient features on the statistical property, image naturalness, structure, and texture by comprehensively analyzing the characteristics of encrypted images. Specifically, the features on statistical property measure the visual recognizability, and the remainder ones quantify the distortions on image attributes.

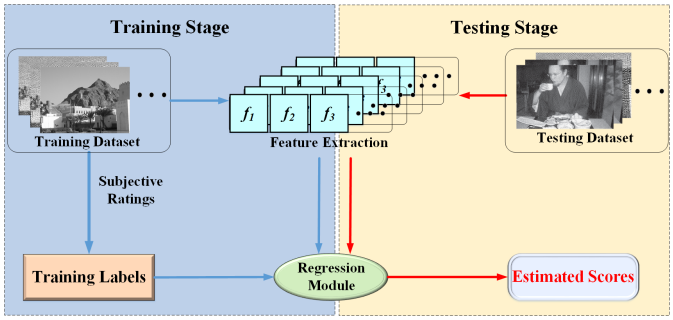
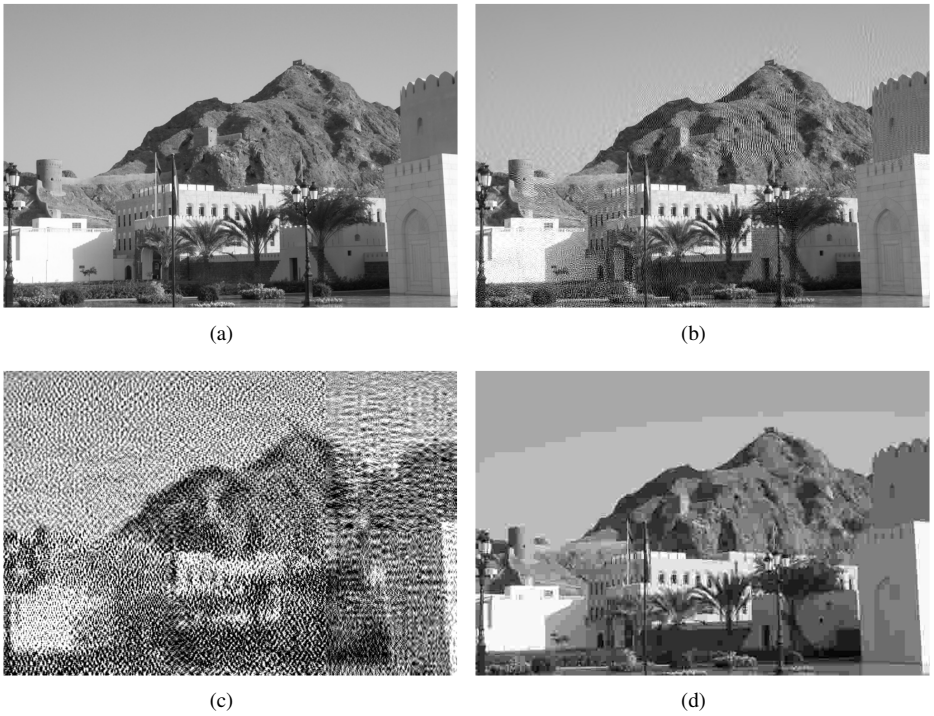


Fig. 2. Framework of the proposed metric.

A. Statistical Property

Digital images are full of complex structures with strong correlations among pixels. The HVS can successfully recognize image contents by extracting structural information. For security purpose, an image is post-processed by randomizing the image pixel values. Under extreme conditions, the statistical distribution of pixel values tends to be uniform, indicating the convincing security [46]. According to these observations, it is a good way to estimate the visual security by measuring the pixel distribution. In this study, we extract various features through analyzing statistical properties of an image. Given an

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2913315, IEEE Transactions on Multimedia



(a) (b) (c) (d)

Fig. 1. (a) is the reference image. (b) and (c) are the generated images processed by slight- and high- strength encryptions. (d) is the compressed version (QP = 5).

input image, its standard deviation S_1 is first computed by:

$$S_1 = \left(\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \boldsymbol{E}(x))^2 \right)^{\frac{1}{2}}, \tag{1}$$

where N denotes the number of image pixels, x_i is the gray value of the i-th image pixel. $\boldsymbol{E}(\cdot)$ is the expectation operator. Apart from the standard deviation, the skewness S_2 and kurtosis S_3 characteristics are subsequently measured:

$$S_2 = \frac{\boldsymbol{E}(x - \boldsymbol{E}(x))^3}{S_1^3}, \tag{2}$$

$$S_3 = \frac{\boldsymbol{E}(x - \boldsymbol{E}(x))^4}{S_1^4} - 3. \tag{3}$$

In addition, we also calculate high central moments S_4:

$$S_4 = \boldsymbol{E}(x - \boldsymbol{E}(x))^k, \tag{4}$$

where k determines the order of central moment. In this paper, we set k as 5 and 6, thereby obtaining two kinds of high central moments.

Revisiting Fig. 1, scrupulous readers may find that, natural images preserve a regular distribution in the spatial domain, whereas such regularity fades with respect to the encrypted images. Inspired by this, we analyze the pixel values' correlation in the spatial domain to indicate the degree of visual security. To get diverse information, we capture such characteristics along four directions, i.e., the horizontal direction, vertical direction, main-diagonal direction, and secondary-diagonal direction. In each direction, the correlation coefficient of the image I_1 and its neighboring-shift version I_2 can be formulated by:

$$C = \frac{1}{H \cdot W \cdot \sigma_1 \cdot \sigma_2} \sum_{h=1}^{H} \sum_{w=1}^{W} (I_1^{h,w} - \overline{I_1})(I_2^{h,w} - \overline{I_2}) \tag{5}$$

where W and H denote the width and height of the image, respectively. σ_1 and σ_2 calculated by Eq. (1) are standard deviations of I_1 and I_2. $\overline{I_1}$ and $\overline{I_2}$ are the mean values of I_1 and I_2. $I_1^{h,w}$ and $I_2^{h,w}$ are the pixel values of I_1 and I_2 in the (h,w) position. In this study, the neighboring-shift version is simply obtained by shifting pixel location in a direction with one pixel distance. Since we consider four directions, four correlation coefficients are obtained at one time in total. Furthermore, it is widely known that entropy provides an excellent idea about self information estimation. Therefore, we also utilize entropy to quantify the uncertain degree of an image. The entropy E_s can be calculated by

$$E_s = -\sum_{u=0}^{U} p_u \cdot \log p_u. \tag{6}$$

In Eq. (6), p_u denotes the probability density of the u-th pixel value. For 8-bits image, $U = 255$.

B. Image Naturalness

Generally, digital images represent the commonness and preserve the original naturalness of the majority of natural scenes. Such naturalness has been successfully investigated, discussed, and subsequently unveiled in the form of natural scene statistic (NSS) regularity [47]. In the literature, considering that distortions can corrupt NSS regularity, many research attempts have explored the usage of NSS in evaluating

image quality [40], [48]–[51]. For example, NSS was used in evaluating the JPEG2000 compression in [48], and in evaluating more synthetical distortions in [40], such as JPEG compression, Gaussian noise, etc. In [50], it was used in evaluating the quality of 3D virtual views. However, there is no attempt in investigating the feasibility of utilizing NSS in evaluating encrypted images. Actually, directly changing pixel intensities inevitably corrupts image contents, which may result in infringing NSS regularity and causing unnatural perception. In view of this, we utilize the NSS regularity to measure the naturalness damage caused by encryption algorithms. Specifically, for an image V, it is preprocessed to obtain the locally normalized luminance $\widehat{V}$ via the local mean removal and divisive normalization operation:

$$\widehat{V}(i,j) = \frac{V(i,j) - \mu_V(i,j)}{1 + \sigma_V(i,j)} \tag{7}$$

where $\sigma_V(i,j)$ and $\mu_V(i,j)$ denote the standard deviation and local mean (which are calculated by convoluting V with a $P \times Q$ filter kernel) of the (i,j)-th pixel, respectively. As stated in [40], the type and size of filter kernel do not sensitively affect the performance. In this work, we emulate [40] and adopt a 2D Gaussian kernel with size 7×7. Without loss of generality, the mean subtracted contrast normalized (MSCN) coefficients (obtained by Eq. (7)) of natural images exhibit Gaussian-like appearance, which is easily infringed by distortions. To investigate whether encrypted images also obtain such regularity, we give a simple example, as shown in Fig. 3. As can be seen from Fig. 3(a), for the natural image (i.e., Fig. 1(a)), its MSCN coefficients exhibit Gaussian-like distribution. In contrast, MSCN coefficients of the encrypted image (i.e., Fig. 1(c)) show more Laplacian-like appearance, as shown in Fig. 3(b). This indicates that measuring the distribution of MSCN coefficients is a good way to characterize the naturalness distortion of encryption. Here, we use the generalized Gaussian distribution (GGD) with zero mean to estimate the distribution properties of MSCN coefficients. Formally, the GGD model can be expressed by

$$g(\kappa, \alpha, \delta^2) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} exp(-(|\kappa|/\beta)^\alpha), \tag{8}$$

where $\beta = \delta\sqrt{\Gamma(1/\alpha)/\Gamma(3/\alpha)}$, $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt$. In Eq. (8), κ is the MSCN coefficient, parameters α and δ^2 determine the shape and variance of the distribution, respectively. In this study, we collect α and δ^2 as two features to estimate image naturalness.

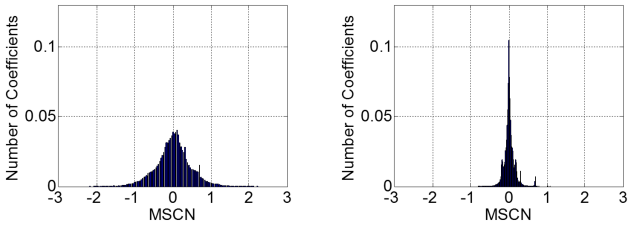


Fig. 3. MSCN coefficients' distributions of: (a) Fig. 1(a) and (b) Fig. 1(c).

C. Image Structure and Texture

Perceptual quality profits from three components, i.e., structure, texture and background [52]. Among them, structural information plays a primary role in visual perception, especially in object recognition. Textural information, representing abundant details of image contents, occupies the second significant place. On the contrary, background regularly indicates the smooth areas, thereby contributing less on the recognition of image contents. Hence, we mainly take structural and textural information into account for the quality evaluation task.

Classical gradient descriptors, such as Prewitt, Sobel, and Lapacian, have been extensively used in structural information extraction. However, they merely mathematically consider pixel values' relationship but ignore human visual characteristics, which should be paid attention to in image processing, especially in perceptual related tasks. Considering that the HVS focuses on the point where each harmonic component contains the same phase, phase congruence (PC) is a biologically plausible structure extraction method [53]. Given a point a, its PC is simply computed as:

$$PC(a) = \frac{\sum_n W(a)\lfloor A_n(a) \cdot \Delta\vartheta_n(a) - \gamma_n \rfloor}{\sum_n A_n(a) + \varepsilon}, \tag{9}$$

with

$$W(a) = (exp[(u - t(a)) \cdot v] + 1)^{-1}, \tag{10}$$

$$\Delta\vartheta_n(a) = \cos(\vartheta_n(a) - \overline{\vartheta_n}(a)) - |\sin(\vartheta_n(a) - \overline{\vartheta_n}(a))|, \tag{11}$$

where symbol $\lfloor \cdot \rfloor$ aims to set negative value to zero; $A_n(a)$ and $\vartheta_n(a)$ represent the amplitude and phase of n-th Fourier component at position a, respectively; γ_n predicts the noise extent; ε is a small positive constant to preserve stability; $t(a) = (1/N)\sum_n A_n(a)(A_{max}(a) + \varepsilon)^{-1}$ denotes the manipulating function by weighting; N is the scale number; u offers a cutoff value for penalizing low PC values under it; v, as the gain variable, controls the cutoff sharpness; $\overline{\vartheta_n}(a)$ represents the mean value of phase. Since it is out of scope to investigate these parameters' influence on the PC map, in this study, we directly set them according to [53]. Fig. 4(b) portrays the PC map of Fig. 4(a). It can be intuitively observed that PC is component to reflect image structures.

However, scrupulous readers may observe that PC map merely represents general outline but ignores the structures in details, which is also important for content understanding. To cope with this problem, we extract multiple high-order PC maps by taking the first order PC map as the input and then calculating its PC map. Through M times iteration, we finally obtains M PC maps. Fig. 4(c) shows the second order PC map. As can be seen, it preserves more detailed structures than Fig. 4(b). Revisiting Fig. 1, it is clear that an encrypted image suffers from structural distortion. There is no doubt that such distortion inevitably induces PC map variation as well as its high-order versions. Since these PC maps exist image information transmission, they reasonably possess the instinct correlation. Therefore, we hold the hypothesis that encryption affects correlation between adjacent PC maps and utilize the mutual information to measure the relevance of PC maps.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2913315, IEEE Transactions on Multimedia

(a) (b) (c)

Fig. 4. Illustrations of PC maps. (a) is the gray-scale image, (b) and (c) are its first and second order PC maps, respectively. For convenient viewing, the edges are highlighted in black lines.

Given two PC maps I_j and $I_{j+1}, j \in \{1, 2, \cdots, M-1\}$, their mutual information can be calculated by:

$$\mathbf{M}(I_j, I_{j+1}) = E(I_j) + E(I_{j+1}) - E(I_j, I_{j+1}) \tag{12}$$

where $E(I_j)$ and $E(I_{j+1})$ calculated by Eq. (6) are the entropy values of I_j and I_{j+1}, respectively. $E(I_j, I_{j+1})$ denotes the joint entropy and can be obtained by:

$$E(I_j, I_{j+1}) = -\sum_{h,k} P_{I_j, I_{j+1}}(h,k) \cdot \log P_{I_j, I_{j+1}}(h,k), \quad h, k \in \{0, 1, \cdots, 255\}. \tag{13}$$

Moreover, we also capture the local entropy of the PC map to further represent image structural variation. More specifically, the local entropy is calculated using Eq. (6) with window size 8×8. Then, the mutual information of two local entropy maps is also estimated according to Eq. (12). It is worth nothing that the structural variation estimated via Eqs. (9) - (13) is reference-free and entirely different from previous works. Given that blurriness usually corrupts image structures, we also adopt one efficient and effective sharpness evaluation method (i.e., the LPC_SI algorithm) to measure the structural distortion caused by blurriness [54].

Our second consideration is the textural information. Two widely used and proved textural descriptors are chosen in this study, i.e., GLCM and LBP. According to [55], GLCM can be obtained through the following steps: First, an image is converted to a matrix C_I by mapping the image pixel value range to $\mathcal{L}$ gray values. Second, for gray values l_1 and l_2 in C_I ($l_1, l_2, \in \{1, 2, \cdots, \mathcal{L}\}$), $m_{d,\theta}(l_1, l_2)$ counts their occurrences to be a neighbor with a distance d in the direction of θ. Finally, the co-occurrence matrix is expressed as $M_{d,\theta}$. After obtaining GLCM, we extract two features from it (i.e., energy and homogeneity) to indicate textural information. Energy is a global feature, reflecting GLCM's distribution and roughness. By definition, it can be expressed as:

$$F_E = \sum_{l_1=1}^{\mathcal{L}} \sum_{l_2=1}^{\mathcal{L}} P_{d,\theta}(l_1, l_2)^2, \tag{14}$$

where $P_{d,\theta}(l_1, l_2)$ stands for the probability of $m_{d,\theta}(l_1, l_2)$ in $M_{d,\theta}$, and it can be computed by:

$$P_{d,\theta}(l_1, l_2) = m_{d,\theta}(l_1, l_2) / \sum_{l_1=1}^{\mathcal{L}} \sum_{l_2=1}^{\mathcal{L}} m_{d,\theta}(l_1, l_2). \tag{15}$$

Homogeneity, as a local feature, quantifies the local change of pixel intensities in C_I. Mathematically, it is formulated as:

$$F_H = P_{d,\theta}(l_1, l_2) / (1 + |l_1 - l_2 - 2|). \tag{16}$$

Overall, the selected two features of GLCM consider both local and global characteristics to represent textural information. Generally speaking, the computation of GLCM with one pixel distance and four directions is sufficient to reflect the textural information, and we therefore set d and θ as 1 and $\{0°, 45°, 90°$ and $135°\}$, respectively.

LBP describes the local textural feature by calculating the relationship between gray values of the center pixel and its surrounding neighbors [56]. By definition, the rotation invariant uniform LBP can be expressed as:

$$\Upsilon_{P,R}^{riu2} = \begin{cases} \sum_{i=0}^{P-1} s(I(n_i), I(n_c)) & , if \;\; \mathcal{U}(\Upsilon_{P,R}) \leq 2 \\ P+1 & , otherwise \end{cases}, \tag{17}$$

where P is the number of sampling points, and R is the radius value of neighbors. $I(n_c)$ is the gray value of center pixel (n_c), and $I(n_i)$, stands for the gray value of i-th neighbor ($i = \{0, 1, \cdots, P-1\}$). The superscript ^{riu2} indicates rotation invariant uniform patterns when $\mathcal{U}(\cdot)$ is less than 2. Formally, $\mathcal{U}(\cdot)$ is computed as the number of bitwise transitions:

$$\begin{aligned} \mathcal{U}(\Upsilon_{P,R}) = &\|s(I(n_{P-1}), I(n_c)) - s(I(n_0), I(n_c))\| \\ &+ \sum_{i=1}^{P-1} \|s(I(n_i), I(n_c)) - s(I(n_{i-1}), I(n_c))\| \end{aligned}, \tag{18}$$

where

$$s(I(n_i), I(n_c)) = \begin{cases} 1, & if \;\; I(n_i) - I(n_c) \geq 0 \\ 0, & if \;\; I(n_i) - I(n_c) < 0 \end{cases}. \tag{19}$$

According to Eqs. (17)-(18), we can observe that the textural information of an image can be represented by *P+2* different patterns. By definition, P and R directly determine the properties of the LBP descriptor. In this study, we set $P = 8$ and $R = 1$ as it is sufficient to quantify the textural information in such settings [31], [43].

D. Visual Quality Estimation

So far, we have extracted many features for one image in consideration of statistical properties, image naturalness, structure and texture, and directly combined them into a

feature vector. Our next task is to build a regression model $\mathbf{R}_m$, which maps feature vectors to the associated quality scores. In the literature, many machine learning tools can be chosen for the regression task. Given that SVR has been widely proved to be effective in NR visual quality evaluation owing to its simple but effective characteristics [40], [42], [44], in this study, we choose the libSVM package [57] to implement SVR as the competing metrics do for the fair comparison. Given the training data $\mathcal{D} = \{(o_1, \tau_1), (o_2, \tau_2), \cdots (o_r, \tau_r)\}$, where o_i and τ_i, $i = 1, 2, \cdots, r$, are the feature vector and quality score of the i-th image, the standard form of SVR can be formulated as:

$$\begin{aligned} \underset{\mathbf{w},\gamma,\mathbf{z},\mathbf{z'}}{\text{minimize}} \quad & \frac{1}{2}||\mathbf{w}||^2 + l\sum_{i=1}^{r}(z_i + z_i^{'}) \\ s.t. \quad & \mathbf{w}^T\varphi(o_i) + \sigma - \tau_i \leq p + z_i \\ & \tau_i - \mathbf{w}^{\mathrm{T}}\varphi(o_i) - \sigma \leq p + z_i^{'} \\ & z_i, z_i^{'} \geq 0, i = 1, 2, \cdots, r. \end{aligned} \tag{20}$$

where l and p are parameters that meet $l > 0$ and $p > 0$. $K(o_i, o_j) \equiv \varphi(o_i)^{\mathrm{T}}\varphi(o_j)$ is the kernel function. Here, the radial basis function (RBF) is used as the kernel function due to the fact that it has fast learning convergence and can effectively approximate nonlinear function. The RBF can be expressed as:

$$K(o_i, o_j) = e^{-\gamma\|o_i - o_j\|^2}, \tag{21}$$

where γ is the variance of the kernel function. Based on the training data $\mathcal{D}$, parameters l, p, and γ can be finally determined, and the prediction model $\mathbf{R}_m$ is accordingly built. In the testing stage, the visual quality score ρ of a test image can be estimated by feeding its feature vector $\mathbf{F}$ into the pre-trained model $\mathbf{R}_m$:

$$\rho = \mathbf{R}_m(\mathbf{F}). \tag{22}$$

IV. Experimental Results

A. Experimental Protocol

1) Database: The IVC-SelectEncrypt database is the only encrypted image database reported in the literature [58]. This database was created by the researchers from the IRCCyN lab [1]. It consists of 200 corrupted images created from 8 reference images (with resolution of 960 × 540). Each reference image was processed by 5 different encryption techniques with 5 levels. The involved encryption techniques were traditional encryption, truncation of the codestream, wavelet packet encryption, window encryption with no error concealment, and window encryption with error concealment. The quality scores of images were generated subjectively using a "Pair Comparison" protocol with 21 observers. During the experiment, viewers were sat at a distance of 6 times screen height in a controlled environment. Each image's quality is given in the form of mean opinion score (MOS), ranging from 1 to 5. A higher MOS value indicates to higher visual quality.

[1]http://www.irccyn.ec-nantes.fr/~autrusse/Databases/SelectiveEncryption/

2) Evaluation Criteria: For the present study, we utilize four evaluation criteria to evaluate and compare the proposed method with existing methods, including Root mean-squared error (RMSE), Spearman rank correlation coefficient (SRCC), Kendall's rank correlation coefficient (KRCC), and Pearson linear correlation coefficient (PLCC) [2]. By definition, these criteria can be respectively expressed as:

$$\Lambda_R = \sqrt{\frac{1}{K}\sum_{i=1}^{K}(p_i - m_i)^2}, \tag{23}$$

$$\Lambda_S = 1 - \frac{6\sum_{i=1}^{K} D_i^2}{K(K^2 - 1)}, \tag{24}$$

$$\Lambda_K = \frac{2(K_1 - K_2)}{K(K-1)}, \tag{25}$$

$$\Lambda_P = \frac{\sum_{i=1}^{K}(p_i - \overline{p})(m_i - \overline{m})}{\sqrt{\sum_{i=1}^{K}(p_i - \overline{p})^2(m_i - \overline{m})^2}}. \tag{26}$$

In the above equations, K is the total number of images in the dataset. p_i and m_i represent the predicted score and its associated subjective score of the i-th image. D_i is the difference of the i-th image's rank in the objective and subjective scores. K_1 and K_2 indicate the number of concordant and discordant pairs found in the dataset. $\overline{p}$ and $\overline{m}$ are the mean values of predicted scores and subjective scores, respectively. Theoretically, an excellent metric should obtain higher values (with the maximum of 1) of PLCC, SRCC and KRCC but smaller value (with the minimum of 0) of RMSE. Before the calculation of PLCC and RMSE, a five-parameter logistic regression function is needed to remove the nonlinearity of the estimated scores:

$$f(x_o) = \varsigma_1 \cdot \left[\frac{1}{2} - \frac{1}{1 + e^{\varsigma_2 \cdot (x_o - \varsigma_3)}}\right] + \varsigma_4 \cdot x_o + \varsigma_5 \tag{27}$$

where x_o and $f(x_o)$ denote the estimated score set and mapped score set of the testing image set, respectively. Taking x_o as the input and MOS as the output of Eq. (27), parameters $(\varsigma_1, \varsigma_2, \cdots, \varsigma_5)$ can be determined by the *nlinfit* function in MATLAB. The values of these parameters are initialized according to the video quality expert group [59] and optimized with nonlinear least squares optimization between $f(x_o)$ and MOS values. Since different objective methods have different x_o, the fitting parameters are different. In spite of this, the mapped score $f(x_o)$ is with the same range, so it is fair to make comparisons among different objective methods.

3) Implementation Details: The proposed metric is learning-based, so a dataset should be split into two non-overlapping subsets, i.e., training subset and testing subset. The performance is evaluated on the testing subset after learning the prediction model on the training subset. In this work, we strictly follow previous works [20], [43], [44] and set the train-test split as 80%-20%. To ensure the results are not restricted to specific train-test splits, we randomly separate the train subset and test subset 500 times, and report the results in

[2]For the sake of simplicity, we mark them as $\mathbf{\Lambda_R}$, $\mathbf{\Lambda_S}$, $\mathbf{\Lambda_K}$, and $\mathbf{\Lambda_P}$ according to priority.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2913315, IEEE Transactions on Multimedia

the form of median value. Besides, another parameter directly affects the overall performance, i.e., the order number used in structure extraction. In this study, we set the order number as 4 according to the experiment discussed in Section IV-C.

B. Overall Performance Comparison

We compare the proposed method against state-of-the-art visual security evaluation metrics and mainstream synthetical distortion evaluation metrics. The former case includes ESS [27], LSS [27], LFBVS [11], and Xiang's method [10]. The later case consists of SSIM [22], BPRI [14], NIQE [60], BRISQUE [40], NFERM [44], SSEQ [42], NRSL [31], and GWH-GLBP [45]. For learning-free methods, the performance is evaluated on the entire database. Whereas, similar to the proposed method, we divide one dataset into training and testing subsets for learning-based metrics (i.e., BRISQUE, NFERM, SSEQ, NRSL, and GWH-GLBP), and measure the performance on the testing subset after learning the prediction model on the training subset. Specifically, the result is also reported as the median value of 500 random train-test split results. To ensure the fair comparison and avoid unnecessary mistakes, all the results (except those of ESS, LSS, and Xiang's method [3]) are computed by implementing the source codes released by authors.

TABLE I
COMPARISON RESULTS WITH REFERENCE-BASED METRICS.

Criteria	ESS [27]	LSS [27]	LFBVS [11]	Xiang [10]	SSIM [22]	Proposed
$\mathbf{\Lambda_P}$	0.901	0.920	0.891	**0.950**	0.880	0.894
$\mathbf{\Lambda_S}$	0.909	0.932	0.891	**0.949**	0.867	0.892
$\mathbf{\Lambda_K}$	0.747	0.786	0.712	**0.819**	0.689	0.726
$\mathbf{\Lambda_R}$	0.576	0.520	0.601	**0.413**	0.629	0.598

Experimental results, in the form of PLCC, SRCC, KRCC, and RMSE, are tabulated in Table I and Table II. For convenient viewing, the best result in each type is highlighted in boldface. From the data in tables, we can observe some significant and meaningful clues. First, all reference-based metrics (ESS, LSS, LFBVS, Xiang's method, and SSIM), no matter designed for evaluating visual security or visual quality, have encouraging performance. Since reference-based metrics possess the reference information, they advantage in obtaining a high performance by measuring the relationship between the corrupted image and its reference version. On the contrary, reference-free metrics do not require reference information, thereby obtaining relatively inferior performance, as shown in Table II. In spite of this, the proposed metric still spares no effort to solve the assessment problem and ultimately possesses a close approximation to these reference-based metrics. More concretely, it is superior to SSIM, LFBVS and is slightly inferior to ESS. Second, the proposed method outperforms those reference-free metrics, e.g., BPRI, NIQE, BRISQUE, NFERM, SSEQ, NRSL, and GWH-GLBP. Much to our excitation, the PLCC value of proposed metric is in advance of the runner-up reference-free method (NRSL) approximately 5%. Third, learning-based methods basically outperform the learning-free method (NIQE). This may be attributed to that machine learning in learning-based methods is more competent to build the relationship between features and quality scores than the empirical combination in learning-free methods. Based on the above analysis, we have sufficient confidence that the proposed metric can estimate the visual quality of encrypted images effectively and is consistent with subjective feelings.

As introduced above, the proposed metric comprehensively considers two components, i.e., estimating visual recognizability and measuring image attributes, to estimate the quality of encrypted images. Therefore, we further investigate the contribution of each component on the final performance. For this purpose, we take features depicted in Section III-A as an individual component to represent visual recognizability, while utilize those in Section III-B and Section III-C to estimate distortions in image attributes, respectively. Then, the features in each component and their associated quality scores are trained via SVR as the same procedure mentioned in Section IV-A. Finally, the experimental results are reported in median value of 500 random train-test splits, as tabulated in the right side of Table I. For the sake of simplicity, we utilize "M_1" and "M_2" to denote the generated models in two components, respectively. From Table I, we can observe that both components obtain fair performance but with positive effects. In other words, only considering visual recognizability or measuring image attributes cannot achieve the expected results. Such conclusion dovetails with previous works well [10], [29]. In spite of this, promising performance is obtained when considering both components, as the proposed metric does.

Moreover, to draw a more reliable conclusion, we also investigate the statistical significance between the proposed metric and competing ones individually (i.e., BRISQUE, NFERM, SSEQ, NRSL and GWH-GLBP) via Rank-Wilcoxon test (with the 95% confidence interval). During the statistical test, SRCC values of the proposed metric and a competing one across 500 trails are regarded as the test samples. A null hypothesis is that the mean correlation of the SRCC values of proposed metric is equal to that of the competing one with the 95% confidence interval. The alternate hypothesis is that the mean correlation of the SRCC values of proposed metric is better or worse than that of the competing one. Experimental results demonstrate that the proposed metric is statistically better to all of competing metrics.

Furthermore, we are also interested in whether the proposed metric is able to compare and rank the quality of multiple images generated by different encryption algorithms with different levels. Since the proposed metric is learning-based, it is impossible to estimate the quality scores of all images in the database at the same time. In this study, we adopt the leave-one-out strategy utilized in [39] to evaluate the proposed metric (as well as the competing learning-based ones). Specifically, the database is first split into 8 image sets according to the image content. For each set, there are 25 images generated by 5 encryption algorithms with 5 levels. Then, 7 image sets are used to learn a prediction model, and the remaining one

[3]Actually, their results are directly excerpted from reference [10], which has the similar test environment like us.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2913315, IEEE Transactions on Multimedia

TABLE II
COMPARISON RESULTS WITH REFERENCE-FREE METRICS.

Criteria	BPRI [14]	NIQE [60]	BRISQUE [40]	NFERM [44]	SSEQ [42]	NRSL [31]	GWH-GLBP [45]	M_1	M_2	Proposed
$\mathbf{\Lambda_P}$	0.530	0.228	0.815	0.798	0.786	0.840	0.824	0.825	0.859	**0.894**
$\mathbf{\Lambda_S}$	0.437	0.217	0.796	0.784	0.770	0.830	0.761	0.816	0.849	**0.892**
$\mathbf{\Lambda_K}$	0.316	0.149	0.630	0.644	0.580	0.655	0.585	0.616	0.670	**0.726**
$\mathbf{\Lambda_R}$	1.124	1.291	0.773	0.794	0.811	0.711	0.757	0.720	0.686	**0.598**

TABLE III
SRCC COMPARISON RESULTS ON THE INDIVIDUAL IMAGE SET.

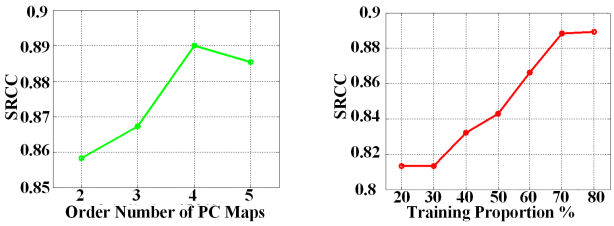
Image Set	Reference-Based		Reference-Free							
	LFBVS [11]	SSIM [22]	BPRI [14]	NIQE [60]	BRISQUE [40]	NFERM [44]	SSEQ [42]	NRSL [31]	GWH-GLBP [45]	Proposed
Cimg6013	0.825	**0.854**	0.301	0.486	0.777	0.736	0.851	0.787	0.699	**0.893**
Cimg6178	0.879	**0.881**	0.382	0.245	0.884	0.884	0.897	0.816	0.910	**0.941**
Cimg7593	**0.950**	0.867	0.521	0.114	0.855	0.852	0.915	0.900	0.880	**0.964**
Cimg7869	0.844	**0.850**	0.543	0.282	0.837	0.863	0.837	0.879	**0.903**	0.876
Img_0596	**0.912**	0.885	0.511	0.375	0.855	0.857	0.687	**0.945**	**0.945**	0.934
Img_0764	0.897	**0.906**	0.437	0.389	0.878	0.883	**0.901**	0.867	0.829	0.795
Img_0768	**0.927**	0.896	0.152	0.749	0.694	0.642	0.424	0.491	**0.822**	0.801
P1010004	**0.931**	0.922	0.709	0.347	0.544	0.617	0.541	0.727	0.784	**0.911**
Average	**0.896**	0.883	0.444	0.373	0.790	0.792	0.757	0.801	0.847	**0.890**

is used for testing. This procedure repeats 8 times, and each time is with a different testing set. Table III presents the SRCC values of each metric over the 8 image sets. For convenient viewing, the best result is highlighted in boldface. From the table, we have the following observations. First, none of these metrics always achieves the best result among all image sets. Second, the proposed metric performs stably across different image sets and totally occupies the top place 4 times among the 8 comparisons with competing NR metrics. In contrast, some metrics (e.g., BPRI, NIQE, BRISQUE, and SSEQ) show a significant change across different image sets. Third, the proposed metric outperforms all competing NR metrics with respect to the average SRCC value of obtained across all image sets. According to the comparison results, we believe that the proposed metric possesses more potential in ranking the quality of encrypted images than others.

C. Parameters Sensitivity

During feature extraction, the order number M used in the PC map construction mainly determines the overall performance of the proposed metric. In this section, we further investigate its impact. For this purpose, we set M with various values, ranging from 2 to 5, and construct model as described in Section III-D to obtain the associated results. Experimental result is intuitively depicted in Fig. 5(a). As can be seen, SRCC value slightly increases with the increasement of M and reaches its peak when M = 4. Hence, we set it as 4 in this study reasonably.

As the proportion of training subset to the entire database highly determines the overall performance of learning-based metrics, we further conduct an experiment to explore the impacts of proportion of the training subset. To be specific, we range the training subset from 20% to 80% at an interval of 10%, and the remaining subset for testing. Experimental results are depicted in Fig. 5(b). Intuitively, SRCC gradually increases with the training proportion increasing. Besides, we find that the performance is still considerable when training subset is 40% (with SRCC about 0.84). Hence, we can believe that the proposed method is stable.



(a) (b)

Fig. 5. SRCC values under different parameter settings: (a) order number M, (b) training proportion.

V. CONCLUSION

Existing quality evaluation metrics for encrypted images are reference-based and have limited application range. In this work, we present a novel NR metric for automatically evaluating perceptual quality of encrypted images. As a pioneering work, we attempt to solve the quality evaluation problem by estimating visual recognizability and measuring the distortions in multiple image attributes. Specifically, statistical characteristics, self-correlation, and entropy are extracted to indicate visual recognizability, which reflects the general security of an image. Meanwhile, since naturalness, structure, texture are sensitive to perceptual quality, we extract a number of features to measure the distortions on these image attributes. Finally, a regression model is deployed to build the relationship between the feature space to quality scores. Extensive experiments have proved the effectiveness of the proposed metric on the IVC-SelectEncrypt database.

VI. ACKNOWLEDGMENT

The authors would like to thank the editors and the anonymous reviewers for their constructive comments, which greatly helped in improving this paper.

REFERENCES

[1] W. Stallings, *Cryptography and network security: principles and practices*. Pearson Education India, 2006.

[2] C. Wang, B. Zhang, K. Ren, and J. M. Roveda, "Privacy-assured outsourcing of image reconstruction service in cloud," *IEEE Transactions on Emerging Topics in Computing*, vol. 1, no. 1, pp. 166–177, 2013.

[3] Z. Shahid and W. Puech, "Visual protection of HEVC video by selective encryption of CABAC binstrings." *IEEE Transactions on Multimedia*, vol. 16, no. 1, pp. 24–36, 2014.

[4] Q. Sun and S.-F. Chang, "A secure and robust digital signature scheme for JPEG2000 image authentication," *IEEE Transactions on Multimedia*, vol. 7, no. 3, pp. 480–494, 2005.

[5] M. Grangetto, E. Magli, G. Olmo *et al.*, "Multimedia selective encryption by means of randomized arithmetic coding," *IEEE Transactions on Multimedia*, vol. 8, no. 5, pp. 905–917, 2006.

[6] A. Jolfaei, X.-W. Wu, and V. Muthukkumarasamy, "On the security of permutation-only image encryption schemes," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 2, pp. 235–246, 2016.

[7] Y. Mao and M. Wu, "A joint signal processing and cryptographic approach to multimedia encryption," *IEEE Transactions on Image Processing*, vol. 15, no. 7, pp. 2061–2075, 2006.

[8] B. Qin, R. H. Deng, S. Liu, and S. Ma, "Attribute-based encryption with efficient verifiable outsourced decryption," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 7, pp. 1384–1393, 2015.

[9] H. Hofbauer and A. Uhl, "Identifying deficits of visual security metrics for images," *Signal Processing: Image Communication*, vol. 46, pp. 60–75, 2016.

[10] T. Xiang, S. Guo, and X. Li, "Perceptual visual security index based on edge and texture similarities," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 5, pp. 951–963, 2016.

[11] L. Tong, F. Dai, Y. Zhang, and J. Li, "Visual security evaluation for video encryption," in *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 2010, pp. 835–838.

[12] Y. Yao, Z. Xu, and W. Li, "Visual security assessment for video encryption," in *Communications and Networking in China, 2008. Third International Conference on*. IEEE, 2008, pp. 1317–1322.

[13] W. Lin and C. C. Jay Kuo, "Perceptual visual quality metrics: A survey," *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, 2011.

[14] X. Min, K. Gu, G. Zhai, J. Liu, X. Yang, and C. W. Chen, "Blind quality assessment based on pseudo reference image," *IEEE Transactions on Multimedia*, vol. 20, no. 8, pp. 2049–2062, 2018.

[15] G. Yue, C. Hou, K. Gu, T. Zhou, and G. Zhai, "Combining local and global measures for DIBR-synthesized image quality evaluation," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 2075–2088, 2019.

[16] Q. Jiang, F. Shao, W. Gao, Z. Chen, G. Jiang, and Y.-S. Ho, "Unified no-reference quality assessment of singly and multiply distorted stereoscopic images," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1866–1881, 2019.

[17] K. Gu, J. Qiao, X. Min, G. Yue, W. Lin, and D. Thalmann, "Evaluating quality of screen content images via structural variation analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 10, pp. 2689–2701, 2018.

[18] H. Liu, N. Klomp, and I. Heynderickx, "A no-reference metric for perceived ringing artifacts in images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 4, pp. 529–539, 2010.

[19] G. Yue, C. Hou, K. Gu, S. Mao, and W. Zhang, "Biologically inspired blind quality assessment of tone-mapped images," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 3, pp. 2525–2536, 2018.

[20] L. Li, W. Xia, W. Lin, Y. Fang, and S. Wang, "No-reference and robust image sharpness evaluation based on multiscale spatial and spectral features," *IEEE Transactions on Multimedia*, vol. 19, no. 5, pp. 1030–1040, 2017.

[21] G. Yue, C. Hou, and T. Zhou, "Blind quality assessment of tone-mapped images considering colorfulness, naturalness, and structure," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 5, pp. 3784–3793, 2019.

[22] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[23] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[24] K. Gu, S. Wang, H. Yang, W. Lin, G. Zhai, X. Yang, and W. Zhang, "Saliency-guided quality assessment of screen content images," *IEEE Transactions on Multimedia*, vol. 18, no. 6, pp. 1098–1110, 2016.

[25] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, p. 011006, 2010.

[26] S. Wang, K. Gu, K. Zeng, Z. Wang, and W. Lin, "Objective quality assessment and perceptual compression of screen content images," *IEEE Computer Graphics and Applications*, vol. 38, no. 1, pp. 47–58, 2018.

[27] Y. Mao and M. Wu, "Security evaluation for communication-friendly encryption of multimedia," in *2004 International Conference on Image Processing*, vol. 1. IEEE, 2004, pp. 569–572.

[28] Y. Yao, Z. Xu, and J. Sun, "Visual security assessment for cipher-images based on neighborhood similarity," *Informatica*, vol. 33, no. 1, pp. 69–76, 2009.

[29] A. Jolfaei and A. Mirghadri, "A new approach to measure quality of image encryption," *International Journal of Computer and Network Security*, vol. 2, no. 8, pp. 38–44, 2010.

[30] A. C. Bovik, "Automatic prediction of perceptual image and video quality," *Proceedings of the IEEE*, vol. 101, no. 9, pp. 2008–2024, 2013.

[31] Q. Li, W. Lin, J. Xu, and Y. Fang, "Blind image quality assessment using statistical structural and luminance features," *IEEE Transactions on Multimedia*, vol. 18, no. 12, pp. 2457–2469, 2016.

[32] G. Yue, C. Hou, Q. Jiang, and Y. Yang, "Blind stereoscopic 3D image quality assessment via analysis of naturalness, structure, and binocular asymmetry," *Signal Processing*, vol. 150, pp. 204–214, 2018.

[33] G. Yue, C. Hou, K. Gu, and N. Ling, "No reference image blurriness assessment with local binary patterns," *Journal of Visual Communication and Image Representation*, vol. 49, pp. 382–391, 2017.

[34] Y. Zhou, L. Li, J. Wu, K. Gu, W. Dong, and G. Shi, "Blind quality index for multiply distorted images using bi-order structure degradation and nonlocal statistics," *IEEE Transactions on Multimedia*, vol. 20, no. 11, pp. 3019–3032, 2018.

[35] G. Yue, C. Hou, K. Gu, N. Ling, and B. Li, "Analysis of structural characteristics for quality assessment of multiply distorted images," *IEEE Transactions on Multimedia*, vol. 20, no. 10, pp. 2722–2732, 2018.

[36] L. Li, H. Zhu, G. Yang, and J. Qian, "Referenceless measure of blocking artifacts by Tchebichef kernel analysis," *IEEE Signal Processing Letters*, vol. 21, no. 1, pp. 122–125, 2014.

[37] S. Wang, C. Deng, B. Zhao, G.-B. Huang, and B. Wang, "Gradient-based no-reference image blur assessment using extreme learning machine," *Neurocomputing*, vol. 174, pp. 310–321, 2016.

[38] K. Gu, W. Lin, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "No-reference quality metric of contrast-distorted images based on information maximization," *IEEE Transactions on Cybernetics*, vol. 47, no. 12, pp. 4559–4565, 2017.

[39] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, and G. Zhai, "No-reference quality assessment of contrast-distorted images based on natural scene statistics," *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 838–842, 2015.

[40] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[41] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, 2012.

[42] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Processing: Image Communication*, vol. 29, no. 8, pp. 856–863, 2014.

[43] Y. Fang, J. Yan, L. Li, and J. Wu, "No reference quality assessment for screen content images," in *Multimedia & Expo Workshops (ICMEW), 2017 IEEE International Conference on*. IEEE, 2017, pp. 169–174.

[44] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Using free energy principle for blind image quality assessment," *IEEE Transactions on Multimedia*, vol. 17, no. 1, pp. 50–63, 2015.

[45] Q. Li, W. Lin, and Y. Fang, "No-reference quality assessment for multiply-distorted images in gradient domain," *IEEE Signal Processing Letters*, vol. 23, no. 4, pp. 541–545, 2016.

[46] H. Elkamchouchi and M. Makar, "Measuring encryption quality for bitmap images encrypted with rijndael and kamkar block ciphers," in *Radio science conference, 2005. NRSC 2005. Proceedings of the twenty-second national*. IEEE, 2005, pp. 277–284.

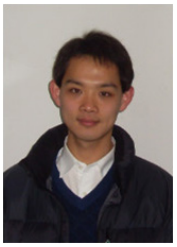
[47] D. L. Ruderman and W. Bialek, "Statistics of natural images: Scaling in the woods," in *Advances in neural information processing systems*, 1994, pp. 551–558.

[48] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1918–1927, 2005.

[49] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, 2011.
[50] M. S. Farid, M. Lucenteforte, and M. Grangetto, "Objective quality metric for 3D virtual views," in *2015 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2015, pp. 3720–3724.
[51] ——, "Evaluating virtual image quality using the side-views information fusion and depth maps," *Information Fusion*, vol. 43, pp. 47–56, 2018.
[52] X. Ran and N. Farvardin, "A perceptually motivated three-component image model-part I: Description of the model," *IEEE Transactions on Image Processing*, vol. 4, no. 4, pp. 401–415, 1995.
[53] P. Kovesi, "Image features from phase congruency," *Videre: Journal of Computer Vision Research*, vol. 1, no. 3, pp. 1–26, 1999.
[54] R. Hassen, Z. Wang, and M. M. Salama, "Image sharpness assessment based on local phase coherence," *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2798–2810, 2013.
[55] R. M. Haralick, K. Shanmugam *et al.*, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, no. 6, pp. 610–621, 1973.
[56] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
[57] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, p. 27, 2011.
[58] F. Autrusseau, T. Stuetz, and V. Pankajakshan, "Subjective quality assessment of selective encryption techniques," 2010. [Online]. Available: http://www.irccyn.ec-nantes.fr/~autrusse/Databases/SelectiveEncryption/.
[59] Video Quality Experts Group (VQEG), "Final report from the video quality experts group on the validation of objective models of video quality assessment," 2003. [Online]. Available: http://www.vqeg.org/
[60] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.

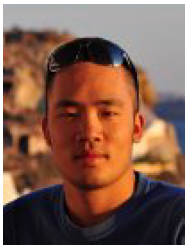
Guanghui Yue received the B.S. degree in communication engineering from Tianjin University, Tianjin, China, in 2014. He was a joint Ph.D. student with School of Computer Science and Engineering, Nanyang Technological University, Singapore, from September, 2017 to January, 2019. His research interests include bioelectrical signal processing, multimedia quality assessment, 3-D image visual discomfort prediction and pattern recognition.

Chunping Hou received the M.Eng. and Ph.D. degrees in electronic engineering from Tianjin University, Tianjin, China, in 1986 and 1998, respectively. Since 1986, she has been with the School of Electronic and Information Engineering, Tianjin University, Tianjin, China, where she is currently a Full Professor and the Director of the Broadband Wireless Communications and 3-D Imaging Institute. Her current research interests include 3-D image processing, 3-D display, wireless communication, and the design and applications of communication systems.



Ke Gu received the B.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015, respectively. He is currently a Professor with the Beijing University of Technology, Beijing, China. His research interests include environmental perception, image processing, quality assessment, and machine learning. He received the Best Paper Award from the IEEE Transactions on Multimedia (T-MM), the Best Student Paper Award at the IEEE International Conference on Multimedia and Expo (ICME) in 2016, and the Excellent Ph.D. Thesis Award from the Chinese Institute of Electronics (CIE) in 2016. He was the Leading Special Session Organizer in the VCIP 2016 and the ICIP 2017, and serves as a Guest Editor for the Digital Signal Processing (DSP). He is currently an Area Editor for Signal Processing Image Communication (SPIC), and an Associate Editor for the IEEE Access and the IET Image Processing (IET-IP). He is a Reviewer for 20 top SCI journals.

Tianwei Zhou received the B.S. degree in automation from Tianjin University, Tianjin, China, in 2014. She was a joint Ph.D. student with Department of Electrical & Computer Engineering, National University of Singapore from August, 2017 to August, 2018. Her current research interests include synchronization, networked control systems, event-triggered control, and image processing.



Hantao Liu (M' 11) received the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands, in 2011. He is currently an Assistant Professor with the School of Computer Science and Informatics, Cardiff University, Cardiff, U.K. His research interests include visual media quality assessment, visual attention modeling and applications, visual scene understanding, and medical image perception. He is currently serving as the Chair of the Interest Group on Quality of Experience for Multimedia Communications at the IEEE MMTC and an Associate Editor of the IEEE Transactions on Human-Machine Systems and the IEEE Transactions on Multimedia.