IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 19, 2021, accepted January 6, 2022, date of publication January 11, 2022, date of current version February 4, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3141914

Blind Image Quality Assessment of Screen Content Images via Fisher Vector Coding

YONGQIANG BAI, ZHONGJIE ZHU, CONGHUI ZHU, AND YUER WANG
Ningbo Key Laboratory of DSP, Zhejiang Wanli University, Ningbo 315100, China

Corresponding author: Zhongjie Zhu (zhongjiezhu@hotmail.com)

This work was supported in part by the Natural Science Foundation of China under Grant 61671412; in part by the Natural Science Foundation of Zhejiang Province under Grant LY21F010014 and Grant LY19F010002; in part by the Natural Science Foundation of Ningbo, China, under Grant 2018A610053 and Grant 202003N4323; in part by the General Scientific Research Project of Zhejiang Education Department under Grant Y201941122; in part by the Ningbo Municipal Projects for Leading and Top Talents under Grant NBLJ201801006; in part by the Fundamental Research Funds for Zhejiang Provincial Colleges and Universities; and in part by the School-Level Research and Innovation Team of Zhejiang Wanli University.

ABSTRACT This paper presents an effective blind image quality assessment (BIQA) method for screen content images (SCIs) based on Fisher vector encoding, with the hypothesis that local statistics will be altered with the change of distortions, and can be characterized by the fusion of statistical models and direction vectors. Firstly, a specific Gaussian mixture model (GMM) is generated from a corpus of pristine SCIs to simulate the local distribution of SCIs in spatial domain. Then, discriminative features are generated to characterize the quality of test image with Fisher vector coding and generated GMM. Finally, support vector regression is adopted to learn the mapping between discriminative features and subjective opinion scores. To validate the performance of our method, extensive experiments are conducted on three public SCI databases and the results well confirm its superiority over the existing relevant BIQA method of SCIs.

INDEX TERMS Blind image quality assessment, screen content image, Gaussian mixture model, fisher vector coding.

I. INTRODUCTION

As an important medium for human-computer interaction, Screen content images (SCIs) are extensively used in remote desktop, cloud computing, online education and virtual screen sharing and so on [1]–[3]. Because they consist of artificial image and natural images, the visual perception mechanism of SCIs is different from traditional natural images [4]–[6]. Meanwhile, as reference image is not available in most cases, designing effective blind image quality assessment (BIQA) method for SCIs is in an urgent demand.

Up to now, there have been some prior efforts devoted to BIQA of SCIs, including feature extraction methods and deep learning methods. For these feature extraction methods, Gu *et al.* extracted 13 perceptual-inspired features with the free energy-based brain theory and structural degradation model [7]. And then, Gu *et al.* extracted four types of features descriptive, i.e., picture complexity, screen content statistics, global brightness quality, and sharpness of details, to improve efficiency [8]. Min *et al.* integrated block-based corner and edge features through a multi-scale weighting framework [9]. Lu *et al.* extracted orientation features and structure features by orientation selectivity mechanism for quality prediction [10]. Fang *et al.* incorporated luminance and texture features with both local and global feature representation [11]. Zheng *et al.* employed gray level co-occurrence matrix-based local features (i.e., entropy, contrast, and local phase coherence) and BRISQUE features [12], which were derived from the distribution of normalized luminance and products of neighboring normalized luminance in spatial domain [13]. In addition to these quality-aware features extracted from spatial domain, Yang *et al.* represented the texture features of SCIs by means of sparse coding, which were characterized by local histogram of oriented gradient features [14]. And also, Wu *et al.* leveraged sparse representation to extract the local structural feature and the global feature from the rough and smooth regions, luminance statistical feature and local binary pattern property, respectively [15]. The key step of these traditional methods is to manually extract quality-aware features

The associate editor coordinating the review of this manuscript and approving it for publication was Byung-Gyu Kim.

via analyzing the characteristics of SCIs, and demonstrated moderate performance on the legacy benchmark databases. With the development of neural network and related technology, many excellent deep learning models are designed to characterize advanced semantic information of SCIs. Zuo *et al.* proposed the first convolutional neural network for SCIs by considering the visual differences between textual and pictorial image patches [16]. Chen *et al.* designed a naturalization module with an upsampling layer and a convolutional layer for the quality prediction of SCIs [17]. Jiang *et al.* treated the image patches with different strategies to modify the convolutional neural network [18]. Yang *et al.* integrated the contour and edge information with L-moment distribution estimation to design an adaboosting back-propagation neural network [19]. These deep learning methods automatically capture the high-level features of SCIs, but lack intuition and interpretability due to neural network characteristics, and prone to underfitting or overfitting results considering the scale of the database. In a word, all these methods were not attempted in terms of statistical characteristics of SCIs. The main reason is that, the artificial part in SCIs destroys the natural scene statistics (NSS) features [20], which is widely adopted for BIQA of natural images [21], [22]. Hence, how to find particularly reliable statistical features, which can be used to characterize the intrinsic quality variations of SCIs, is still a difficult problem to be solved seriously and thoroughly.

To solve this problem, this paper proposed a BIQA method of SCIs based on Gaussian mixture model (GMM) and Fisher vector coding (FVC), with the hypothesis that local statistics will be altered with the change of distortions, and can be characterized by the fusion of statistical models and direction vectors. Firstly, GMM is adopted as the mainstream of generative model for modeling the local features [23]. Then, discriminative features are generated to characterize the quality of test image with Fisher vector coding and generated GMM. Finally, the above features should be employed directly for quality regression. The main contributions are as follows:

(1) The proposed method can efficiently and theoretically characterize the effect of distortions on quality degradation of SCIs. On the one hand, the patch trained GMM is adopted to simulate and approximate the local statistical model of SCIs, due to the variable composition of the artificial and natural parts. On the other hand, the quality-aware features can be more accurately characterized with the dimension elevation and direction extraction of the FVC. And experimental results on three public databases confirmed this hypothesis and demonstrated the efficacy of the proposed method.

(2) The localized representation of statistical properties is the precondition and foundation of quality prediction for SCIs. The artificial portion of SCIs destroys the natural scene statistics of natural scenes, and it is also impractical to obtain a single and accurate statistical model due to the variable composition in SCIs. For this, the strategy of local representation and multi-mode fusion is adopted by training the GMM with each patch. That is, the corpus of SCIs is first divided into many patches to simulate the local and variable composition in SCIs, and then the corresponding Gaussian models are constructed and combined as the target GMM. Hence, the local features of SCIs can be modeled preliminarily and used to characterize the quality-aware features of degraded images.

(3) The change trend of statistical model is the critical step to characterize image degradation. Among them, component and direction are two important factors. For each degraded image, the FVC is adopted in this paper to extract the quality-aware features. As a coding method derived from the Fisher kernel, the FVC can represent powerfully the local feature by dimension elevation, almost without additional calculations. And the obtained vectors contain not only component information but also the information of mean and variance. Meanwhile, FVC is essentially a partial derivative of Gaussian distribution, so we can directly obtain the direction of change on the basis of Gaussian distribution. Hence, the extracted features with FVC can more fully represent the distortion characteristics of SCIs, and then predict its objective image quality by quality regression.

The paper is organized as follows. Section II illustrates the proposed method in detail. Experimental results are shown and analyzed in Section III. Finally, Section IV concludes the paper.

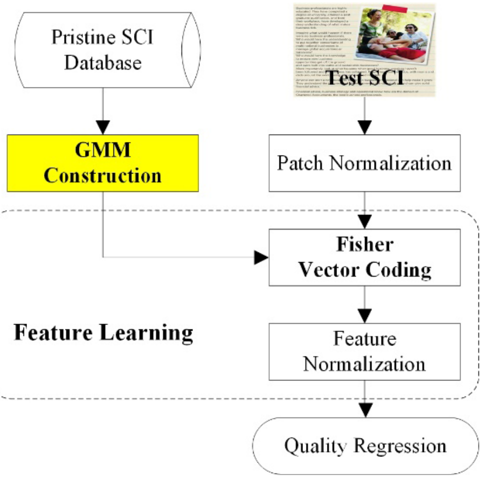


FIGURE 1. Framework of the proposed method.

II. METHODOLOGY

Considering that the artificial part in SCIs destroys the NSS feature of nature images, we build the quality-aware image features inspired by FVC. The framework of the proposed method is shown in Fig. 1 below. In general, the proposed method consists of three parts: (1) GMM Construction, which is to generate the target GMM from pristine SCI database offline. (2) Feature Learning, which is to generate the discriminative features with FVC and generated GMM. (3) Quality Regression, which performs via support vector regression (SVR) based on the combined statistics differences. Detailed process is described as follows.

A. GMM CONSTRUCTION

The FVC is one of the most powerful local feature encoding and image representation generation methods [24], [25]. But when implement the FVC, how to specify the distribution $P = (\mathrm{X}\,|\lambda)$ is extremely important for this method, as the artificial part in SCIs destroys the NSS features. Generally speaking, the distribution resembles a Gaussian distribution only within a local region of the feature space, and intuitively each Gaussian distribution can be seen as a feature prototype. For example, assume an SCI image is simply composed of four components of Gaussian distribution as shown in Fig. 2, where p_i denotes i-th Gaussian distribution with each prior. Obviously, the desired GMM can be formed by linear combination of these components, and it can theoretically approximate any type of probability density distribution, if there are enough components. Hence, a certain number of Gaussian distributions are generated as a measurement anchor in this paper, to accurately depict the whole feature space of the pristine SCI database.

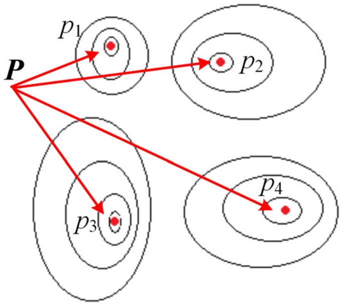


FIGURE 2. Simplified schematic diagram of local probability distribution of SCI image.

Specifically, for these collected SCIs, the raw image patches, with $n \times n$ size in gray-scale domain, are all normalized with divisive normalization transform, which has been widely used in BIQA domain to mimic the early nonlinear processing in human visual system and then reduce redundancies [26], [13].

$$\hat{x}(i,j) = \frac{x(i,j) - \alpha}{\beta + \delta} \tag{1}$$

where, $x(i,j)$ and $\hat{x}(i,j)$ are the raw and normalized image patches, respectively, and (i,j) are the indices sampled on a regular grid over the entire image. α and β are the local mean and standard deviation of each patch, and parameter δ is the constant set as 10 here to prevent instability when the denominator approaches zero. In addition, a zero-phase component analysis whitening process [27] is applied to further remove the linear correlations between each patch.

Subsequently, we consider these normalized image patches $\hat{P}(i,j)$ as local features, and choose VLFeat open-source library to implement the GMM training [28]. For each image, T normalized patches are extracted: $\mathrm{X} = [\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_T] \in R^D$, where each column corresponds to one patch. The constructed GMM for X can be described as $p(X\,|\omega, \mu, \sigma^2)$, where ω, μ and σ^2 denote the prior, mean, and covariance of each component in GMM. Among them, $\omega_i \geq 0, sum(\omega_i) = 1$. As the GMM model can theoretically fit any type of distribution, it is particularly suitable to solve the situation of containing multiple different distributions in SCIs.

Besides, K clusters of GMM are constructed to capture various distortion characteristics. Note that, (1) the process is carried out offline; (2) the constructed GMM is no longer required to be updated and can be directly used as the target GMM for feature learning of the test image.

B. FEATURE LEARNING

With the constructed GMM, the tested image can be coded with the FVC to extracted the quality-aware feature. In essence, the FVC calculates the partial derivative of Gaussian distribution for each local component, and characterize an image with the gradient vector of likelihood function. After the process of FVC, the dimension of image features is greatly elevated, and the obtained vectors contain not only component information but also the information of mean and variance for each component, which can be better used to describe the inherent information of SCI. Meanwhile, the fisher vector, as the partial derivative, contains the direction change of each Gaussian distribution component. So even if we meet two SCIs with the same components, their content and degradation still can be effectively distinguished based on the direction information in the obtained gradient vectors.

Assuming that $\hat{x}_t$ is drawn independent identical distribution from the distribution $P(\mathrm{X}\,|\lambda)$, the sample $X = [\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_T]$ can be described by the gradient vector of the likelihood function with regard to the model parameter λ in the Fisher kernel as follows.

$$\mathrm{G}_\lambda^\mathrm{X} = \nabla_\lambda \log P(\mathrm{X}\,|\lambda) = \sum_{t=1}^{T} \nabla_\lambda \log P(\hat{x}_t\,|\lambda) \tag{2}$$

Let F is called Fisher information matrix and is defined as $\mathrm{F} = E[\mathrm{G}_\lambda^\mathrm{X} \mathrm{G}_\lambda^{\mathrm{X}T}]$, the Fisher kernel is then defined as $K(\mathrm{X},\mathrm{Y}) = \mathrm{G}_\lambda^{\mathrm{X}T} \mathrm{F}^{-1} \mathrm{G}_\lambda^\mathrm{X}$, which measures the direct distance between two samples (that is the similarity between two samples), and generates the distortion impact on degradation as the desired features. Meanwhile, the Fisher kernel as a Kernel function can also simplify mapping space operation, and use an equivalent low-dimensional calculation to avoid the dimensional disaster caused by high-dimensional space operation.

As a result, two samples can be directly compared by the linear kernel of their corresponding normalized gradient vectors which are often called Fisher vectors. Let $\mu_\lambda(\mathrm{X}) = P(\mathrm{X}\,|\lambda)$, Fisher vectors of an image can be described as follows.

$$\mathrm{G}_{\lambda}^{\mathrm{X}} = \sum_{t=1}^{T} \mathrm{F}_{\lambda}^{-1/2} \mathrm{G}_{\lambda}^{\mathrm{X}} = \sum_{t=1}^{T} \mathrm{F}_{\lambda}^{-1/2} \nabla_{\lambda} \log u_{\lambda}(\hat{x}_t) \tag{3}$$

For the generated GMM with the parameter $\lambda = \{\omega_k, \mu_k, \sigma_k, k = 1, \ldots, K\}$, then

$$\mu_{\lambda}(x) = \sum_{k=1}^{K} \omega_k \mu_k(x) \tag{4}$$

where, $\sum_{k=1}^{K} \omega_k = 1$, and $\mu_k(x)$ can be calculated by

$$\mu_k(x) = \frac{\exp\{-\frac{1}{2}(x-\mu_k)^T \sigma_k^{-1}(x-\mu_k)\}}{(2\pi)^{\mathrm{D}/2}\, |\sigma_k|^{1/2}}.$$

Subsequently, the quality-aware features can be obtained with normalized Fisher vectors as follows.

$$\mathrm{G}_{\omega,k}^{\mathrm{X}} = \frac{1}{T\sqrt{\omega_k}} \sum_{t=1}^{T} \gamma_t(k)(x_t - \omega_k) \tag{5}$$

$$\mathrm{G}_{\mu,k}^{\mathrm{X}} = \frac{1}{T\sqrt{\omega_k}} \sum_{t=1}^{T} \gamma_t(k)(\frac{x_t - \mu_k}{\sigma^k}) \tag{6}$$

$$\mathrm{G}_{\sigma,k}^{\mathrm{X}} = \frac{1}{T\sqrt{2\omega_k}} \sum_{t=1}^{T} \gamma_t(k)[\frac{(x_t - \mu_k)^2}{\sigma_k^2} - 1] \tag{7}$$

where, $\gamma_t(k)$ is the probability that generated by the k-th Gaussian component for x_t, that is $\gamma_t(k) = P(k\,|x_t, \lambda)$.

And then, both the quality-aware features are concatenated to a single long quality-aware feature: $f = [\mathrm{G}_{\mu,k}^{\mathrm{X}\top}, \mathrm{G}_{\sigma,k}^{\mathrm{X}\top}]$, $k = 1, 2, \ldots, K$. Furthermore, there are some similar contents in SCIs and similar quality scores in subjective opinion scores, and these similarities will increase image feature similarity, then severely decrease the contribution of other important dimensions and hurt the overall feature effectiveness. Hence, element wise signed power normalization is adopted on the aggregated features to alleviate the corruption caused by these similarities [29]. Specifically, each local feature $\hat{f}$ can be obtained as follows.

$$\hat{f} = sign(f)\,|f|^{\lambda} \tag{8}$$

where, λ is the parameter to control the inhibition degree on the frequent components and f is one of the feature values. Finally, the entire quality-aware features, which are used to quality regression, can be denoted by $F = [\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_K,] \in R^{D\times K\times 2}$.

C. QUALITY REGRESSION

After feature learning, the quality evaluation is achieved using SVR to create a fair comparison with the state-of-the-art BIQA methods. Specifically, SVR model is first learned using a set of training SCIs, then the trained SVR model is used to evaluate the quality of testing SCIs. Here, SVR with radial basis function kernel is adopted as the mapping function from normalized features to subjective quality scores by using LIBSVM package [30].

In this paper, the patch size D is set to 7×7 and the cluster number K is set to100 with experience, so that quality-aware representation provides a vector of dimensionality $D \times K \times 2 = 9800$ (that is F) in total for each input SCI. And the practical effect of each feature vector will be detailedly illuminated in the next section. Meanwhile, the database is divided into training and testing subsets randomly for 1000 times, with 80% as the training dataset and the rest as testing dataset, and the median performance across 1000 times is reported.

III. EXPERIMENTAL RESULTS

A. TESTING DATABASES AND EVALUATION METHODOLOGY

To test the effectiveness of the proposed method, some comparison experiments are conducted on three public SCI databases, i.e., screen content image quality assessment database (SIQAD) [5], screen content database (SCD) [31] and screen content image database (SCID) [32]. For SIQAD, it includes 20 reference SCIs and 980 distorted versions corrupted by seven degradation types, i.e., Gaussian Noise (GN), Gaussian Blur (GB), Motion Blur (MB), Contrast Change (CC), JPEG, JPEG 2000 (J2K) and Layer Segmentation based Coding (LSC), each of which includes seven levels. For SCD, it includes24 reference SCIs and 492 distorted versions corrupted by two distortion types, that is, Screen Content Compression (SCC) and High Efficiency Video Coding (HEVC). For SCID, it consists of 40 reference SCIs and 1800 distorted versions corrupted by nine degradation types, i.e., GN, GB, MB, CC, Color Quantization with Dithering (CQD), JPEG and J2K, HEVC and SCC, each of which includes five levels.

Meanwhile, for the sake of fairness, three commonly-used criteria, i.e., Pearson's linear correlation coefficient (PLCC), Spearman's rank order correlation coefficient (SRCC) and root mean squared error (RMSE), are adopted to evaluate the performance although there are some limitations for these three criteria [33], [34]. PLCC, SRCC, and RMSE estimate the prediction of accuracy, monotonicity, and consistency, respectively. In general, a better perceptual quality prediction metric is expected to have higher PLCC and SRCC values, and lower RMSE value. Furthermore, to remove the nonlinearity of objective quality predictions, a nonlinear logistic regression process with five parameters is applied in the implement as follows [35].

$$f(x) = \beta_1(\frac{1}{2} - \frac{1}{1 + e^{(\beta_2(x-\beta_3))}}) + \beta_4 x + \beta_5 \tag{9}$$

where $(\beta_1, \ldots, \beta_5)$ are the parameters to be fitted, x and $f(x)$ denote the original and the fitted quality scores, respectively.

B. OVERALL PERFORMANCE COMPARISON

In this subsection, we compare the proposed method with the following state-of-the-art BIQA methods for SCIs:

TABLE 1. Performance indices of the proposed and compared methods on three databases.

		SIQAD			SCD			SCID	
	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
BQMS	0.7549	0.7223	9.3042	0.4264	0.3761	1.6701	0.6487	0.6138	10.7787
SIQE	0.7906	0.7625	8.7650	0.7168	0.7012	1.5470	0.6371	0.6034	10.9202
UCA	0.6892	0.6925	/	/	/	/	/	/	/
OSM	0.8306	0.8007	7.9331	0.7068	0.6804	1.5301	/	/	/
NRLT	0.8442	0.8202	7.5957	**0.9227**	**0.9156**	**0.8091**	0.8377	0.8178	7.7265
HRFF	0.8520	0.8320	7.4150	/	/	/	/	/	/
TFSR	0.8618	0.8354	7.4910	/	/	/	0.8017	0.7840	8.8041
CLGF	0.8331	0.8107	7.9172	/	/	/	0.6978	0.6870	10.1439
PICNN	0.8960	**0.8970**	6.7900	/	/	/	0.8270	0.822	8.0130
SIQA-DF	**0.9000**	0.8880	**6.2422**	/	/	/	**0.8514**	**0.8507**	**7.0687**
ABPNN	0.8529	0.8336	7.2817	/	/	/	0.7147	0.6920	10.3988
Proposed	**0.9014**	**0.8915**	**6.1684**	**0.9239**	**0.9198**	**0.8434**	**0.8681**	**0.8550**	**7.0170**

feature extraction methods (BQMS [7], SIQE [8], UCA [9], OSM [10], NRLT [11], HRFF [12], TFSR [14], and CLGF [15]) and deep learning methods (PICNN [17], SIQA-DF [18], and ABPNN [19]). Meanwhile, these methods are conducted on three SCI databases: SIQAD, SCD, and SCID. The experimental results are shown in Table 1, where the top two results in each case are highlighted with boldface and '/' indicates that the value is not available.

From Table 1, we have the following observations. Compared with the feature extraction methods, our method can get better performance with 5% improvement for PLCC in SIQAD database. Furthermore, the performance of our method can be further improved, especially in SCID database. The main reason is that, the manual features cannot accurately characterize and measure the intrinsic quality variations of SCIs, due to the excessive subjectivity and independence caused by the limitations of research progress of visual perception and personal preferences. Meanwhile, the deep learning methods show more excellent performance than these feature extraction methods, as they can automatically capture the high-level and pertinent features for SCIs depending on the neural network and database scale. By the comparison in Table 1, our method shows the similar performance with the deep learning methods, as the statistical model can more objectively reveal the internal quality characteristics of the image, and the patch trained GMM, dimension elevation and direction extraction of the FVC are more suitable for the content generation and distribution peculiarity of SCIs with the previous analysis. Besides, the target GMM is trained off-line and are not altered for the follow-ups, the proposed method has relatively low complexity and is more conducive to practical applications.

C. PERFORMANCE COMPARISON ON INDIVIDUAL DISTORTION TYPE

To comprehensively evaluate the three types of indices to predict the quality degradations of SCIs corrupted by different distortion types, we conduct the performance experiment on three SCI databases as mentioned above in this section. To be specific, Table 2 -4 show the PLCC, SRCC, RMSE results of different distorted types for the proposed and compared methods on SIQAD respectively, and the top two results in each case are highlighted with boldface. Obviously, the proposed method has obvious advantages for all distortion types considering both the sub-features, especially for the CC, JPEG, J2K and LSC. Similar to NSS of natural images, the results demonstrate that the statistical model can more effectively and objectively reveal the internal law of image degradation for SCIs, after local optimization and vector coding. Meanwhile, the proposed method shows the better prediction of consistency, by comparing the average validity

TABLE 2. PLCC results of different distortion types for the proposed and compared methods on SIQAD.

	GN	GB	MB	CC	JPEG	J2K	LSC
BQMS	0.837	0.756	0.724	0.721	0.765	0.791	0.843
SIQE	0.878	0.914	0.784	0.686	0.724	0.734	0.733
UCA	/	/	/	/	/	/	/
OSM	/	/	/	/	/	/	/
NRLT	0.913	0.895	0.899	0.813	0.793	0.685	0.723
HRFF	0.902	0.890	0.874	0.826	0.763	0.754	0.770
TFSR	**0.929**	**0.937**	**0.924**	0.656	0.833	0.835	0.807
CLGF	0.858	0.908	0.861	0.744	0.660	0.746	0.558
PICNN	0.910	0.919	0.889	0.826	**0.829**	**0.852**	0.836
SIQA-DF	0.912	0.924	0.890	**0.844**	**0.829**	0.828	**0.858**
ABPNN	0.914	0.923	**0.895**	0.777	0.801	0.798	0.791
Proposed	**0.938**	**0.908**	0.889	**0.909**	**0.903**	**0.905**	**0.927**

TABLE 3. SRCC results of different distortion types for the proposed and compared methods on SIQAD.

	GN	GB	MB	CC	JPEG	J2K	LSC
BQMS	0.835	0.763	0.718	0.726	0.766	0.792	0.827
SIQE	0.852	0.917	0.835	0.687	0.744	0.724	0.734
UCA	/	/	/	/	/	/	/
OSM	/	/	/	/	/	/	/
NRLT	0.897	0.881	0.892	0.707	0.770	0.676	0.698
HRFF	0.872	0.863	0.850	0.687	0.718	0.744	0.740
TFSR	**0.914**	**0.931**	**0.915**	0.650	**0.838**	**0.835**	0.795
CLGF	0.848	0.915	0.869	0.572	0.678	0.768	0.584
PICNN	0.902	0.916	0.880	0.699	0.823	0.834	**0.872**
SIQA-DF	0.901	0.910	0.880	0.728	0.812	0.816	0.858
ABPNN	0.910	**0.922**	**0.887**	**0.747**	0.777	0.778	0.759
Proposed	**0.917**	0.893	0.880	**0.903**	**0.889**	**0.879**	**0.900**

TABLE 4. RMSE results of different distortion types for the proposed and compared methods on SIQAD.

	GN	GB	MB	CC	JPEG	J2K	LSC
BQMS	8.162	8.839	9.240	9.211	8.587	8.416	7.834
SIQE	8.142	6.424	8.078	9.157	6.478	7.673	6.316
UCA	/	/	/	/	/	/	/
OSM	/	/	/	/	/	/	/
NRLT	6.311	6.917	6.452	7.843	5.872	6.544	5.786
HRFF	6.267	6.783	6.466	6.874	5.862	6.501	5.473
TFSR	**5.311**	**5.214**	**5.527**	10.501	**5.254**	**5.638**	5.622
CLGF	/	/	/	/	/	/	/
PICNN	6.201	5.870	**5.772**	7.012	5.470	5.992	**4.673**
SIQA-DF	6.115	5.768	5.791	**6.747**	**5.384**	5.812	**4.462**
ABPNN	5.975	5.732	6.714	8.068	6.801	6.554	5.456
Proposed	**5.092**	**5.502**	5.972	**5.408**	5.598	**5.686**	5.321

TABLE 5. Performance indices of the proposed with two types of distortion types on SCD.

	HEVC	SCC	ALL
PLCC	0.917	0.888	**0.924**
SRCC	0.911	0.881	**0.920**
RMSE	0.822	1.020	**0.843**

TABLE 6. Performance indices of the proposed with two types of distortion types on SCID.

	GN	GB	MB	CC	CQD
PLCC	0.959	0.967	0.939	0.936	0.941
SRCC	0.953	0.953	0.933	0.930	0.936
RMSE	3.516	3.011	4.137	4.253	3.893
	JPEG	J2K	HEVC	SCC	ALL
PLCC	0.930	0.927	0.852	0.856	**0.868**
SRCC	0.924	0.923	0.855	0.851	**0.855**
RMSE	3.860	3.752	5.154	5.085	**7.017**

and fluctuation magnitude for all distorted types. For example, the feature extraction methods, especially for SIQE and TFSR, show excellent performance to handle GN, GB, and MB, but their performance obviously poor for other distortion types.

Furthermore, the same experiments are conducted on the other public SCI databases, i.e., SCD and SCID, and the relative results are shown in Tables 5 and 6. These experimental results show the similar performance of the proposed method and other methods with SIQAD. In conclusion, it is clear that the proposed method can more precisely and steadily evaluate and reflect various degenerations, which further verifies the effectiveness and robustness of the proposed method.

D. CROSS-DATABASE VALIDATION

Cross-database validation is conducted to verify the generalizability of the proposed method here. Considering that SIQAD and SCID are the representative and largest databases, respectively, and both of them contains 6 distortion types (i.e., GN, GB, MB, CC, JPEG, and J2K), so both of them are adopted as the training and testing databases, respectively. In this paper, one database is trained with these 6 distortion types, and the other is used to test the performance with the trained model. Here entire samples of both databases are adopted for model training and testing, which can reduce dependence on the scale of the database and further verify the generalizability of the proposed method [36].

Table 7 shows the cross-database results for each type of distortion, in which (a) means that the model is trained with SIQAD and tested with SCID, and (b) is the opposite. First, both cross-database performances are similar to each other, indicates that the proposed model has the advantages of high generalization ability, regardless of database size and complexity. Second, the cross-database performance is marginally worse than the in-database performance, which is also a common problem of existing methods. The primary reason is that different fusion rules in each database may generate complex degradation mechanisms. Third, the cross-database performance has decreased for the proposed method but it still achieves a satisfactory performance and stability for most distortion types, except for the JPEG and J2K type as they belong to complex composite compression distortion.

Besides, the cross-database performance is worse than the deep learning methods as shown in Table 1, but it is still comparable with them considering its interpretability. Thus, cross-database results demonstrate that the proposed method can achieves good prediction accuracy and generalization.

E. PARAMETER SETTING

In this subsection, comparison experiments are conducted to validate the influence of the parameter setting on three databases. Here, the sensitivity of cluster number K is discussed by enumerating some certain values in proper interval around the determined value. The corresponding results are shown in Table 8 and the results with determined value in each case are highlighted with boldface.

As mentioned in Section II, parameter K defines the cluster number and directly determines the dimensionality of feature vector. Obviously, larger K value leads to bigger dimensionality of feature vector and greater computational cost. As shown in Table 8, the performance increases first and then decreases slightly with the increase of the value of K. Considering the balance between algorithm performance and computation, K

TABLE 7. Cross validation results of the proposed method with six types of distortion on SIQAD and SCID.

Distortion	(a) Training with SIQAD			(b) Training with SCID		
	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
GN	0.9354	0.9229	4.4458	0.8921	0.8778	6.7410
GB	0.9442	0.9250	3.9131	0.8461	0.8457	7.1941
MB	0.9159	0.8965	4.8584	0.8378	0.8359	7.3102
CC	0.8324	0.7855	6.7832	0.8483	0.8478	7.0374
JPEG	0.7954	0.7790	7.0225	0.8533	0.8482	6.9572
J2K	0.6837	0.6880	7.7272	0.6411	0.0383	10.4801
Overall	0.8525	0.8552	6.2174	0.8478	0.8463	7.2710

TABLE 8. Comparison results with different values of cluster number (K) on different databases.

Cluster Number(K)		50	100	150
Dimensionality		4900	9800	14700
SIQAD	PLCC	0.8903	**0.9014**	0.8901
	SRCC	0.8763	**0.8915**	0.8779
	RMSE	6.3017	**6.1684**	6.3123
SCD	PLCC	0.9098	**0.9239**	0.9084
	SRCC	0.9043	**0.9198**	0.9042
	RMSE	0.8736	**0.8434**	0.8739
SCID	PLCC	0.8515	**0.8681**	0.8526
	SRCC	0.8382	**0.8550**	0.8392
	RMSE	7.2793	**7.0170**	7.2459

is set to 100 empirically in the implementation as default in this paper.

IV. CONCLUSION

With the Fisher vector coding (FVC), a novel blind image quality assessment (BIQA) method for screen content images (SCIs) is proposed in this paper. After constructing the target Gaussian mixture model (GMM) from the corpus of SCIs, discriminative features are generated to characterize the quality of test image with FVC. And then, support vector regression (SVR) is adopted to learn the mapping between discriminative features and subjective opinion scores. Extensive experiments are conducted on three public SCI databases to validate the performance of our method, and the results well confirm its superiority over the existing relevant BIQA method of SCIs.

REFERENCES

[1] T. Li, X. Min, H. Zhao, G. Zhai, Y. Xu, and W. Zhang, "Subjective and objective quality assessment of compressed screen content videos," *IEEE Trans. Broadcast.*, vol. 67, no. 2, pp. 438–449, Jun. 2021.

[2] T. Nguyen, X. Xu, F. Henry, R.-L. Liao, M. G. Sarwer, M. Karczewicz, Y.-H. Chao, J. Xu, S. Liu, D. Marpe, and G. J. Sullivan, "Overview of the screen content support in VVC: Applications, coding tools, and performance," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 10, pp. 3801–3817, Oct. 2021.

[3] W. Kuang, Y.-L. Chan, S.-H. Tsang, and W.-C. Siu, "Online-learning-based Bayesian decision rule for fast intra mode and CU partitioning algorithm in HEVC screen content coding," *IEEE Trans. Image Process.*, vol. 29, pp. 170–185, 2020.

[4] X. Min, K. Gu, G. Zhai, X. Yang, W. Zhang, P. L. Callet, and C. Chen, "Screen content quality assessment: Overview, benchmark, and beyond," *ACM Comput. Surveys*, vol. 54, no. 9, p. 187, Dec. 2022.

[5] H. Yang, Y. Fang, and W. Lin, "Perceptual quality assessment of screen content images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4408–4421, Aug. 2015.

[6] Y. Bai, M. Yu, Q. Jiang, G. Jiang, and Z. Zhu, "Learning content-specific codebooks for blind quality assessment of screen content images," *Signal Process.*, vol. 161, pp. 248–258, Aug. 2019.

[7] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "Learning a blind quality evaluation engine of screen content images," *Neurocomputing*, vol. 196, pp. 140–149, Jul. 2016.

[8] K. Gu, J. Zhou, J.-F. Qiao, G. Zhai, W. Lin, and A. C. Bovik, "No-reference quality assessment of screen content pictures," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 4005–4018, Aug. 2017.

[9] 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.

[10] N. Lu and G. Li, "Blind quality assessment for screen content images by orientation selectivity mechanism," *Signal Process.*, vol. 145, pp. 225–232, Apr. 2018.

[11] Y. Fang, J. Yan, L. Li, J. Wu, and W. Lin, "No reference quality assessment for screen content images with both local and global feature representation," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1600–1610, Apr. 2018.

[12] L. Zheng, L. Shen, J. Chen, P. An, and J. Luo, "No-reference quality assessment for screen content images based on hybrid region features fusion," *IEEE Trans. Multimedia*, vol. 21, no. 8, pp. 2057–2070, Aug. 2019.

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

[14] J. Yang, J. Liu, B. Jiang, and W. Lu, "No reference quality evaluation for screen content images considering texture feature based on sparse representation," *Signal Process.*, vol. 153, pp. 336–347, Dec. 2018.

[15] J. Wu, Z. Xia, H. Zhang, and H. Li, "Blind quality assessment for screen content images by combining local and global features," *Digit. Signal Process.*, vol. 91, pp. 31–40, Aug. 2019.

[16] L. Zuo, H. Wang, and J. Fu, "Screen content image quality assessment via convolutional neural network," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2082–2086.

[17] J. Chen, L. Shen, L. Zheng, and X. Jiang, "Naturalization module in neural networks for screen content image quality assessment," *IEEE Signal Process. Lett.*, vol. 25, no. 11, pp. 1685–1689, Nov. 2018.

[18] X. Jiang, L. Shen, Q. Ding, L. Zheng, and P. An, "Screen content image quality assessment based on convolutional neural networks," *J. Vis. Commun. Image Represent.*, vol. 67, Feb. 2020, Art. no. 102745.

[19] J. Yang, Z. Bian, J. Liu, B. Jiang, W. Lu, X. Gao, and H. Song, "No-reference quality assessment for screen content images using visual edge model and AdaBoosting neural network," *IEEE Trans. Image Process.*, vol. 30, pp. 6801–6814, 2021.

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

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

[22] Q. Jiang, F. Shao, G. Jiang, M. Yu, and Z. Peng, "Supervised dictionary learning for blind image quality assessment using quality-constraint sparse coding," *J. Vis. Commun. Image Represent.*, vol. 33, pp. 123–133, Nov. 2015.

[23] L. Liu, P. Wang, C. Shen, L. Wang, A. Hengel, C. Wang, and H. Shen, "Compositional model based Fisher vector coding for image classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2335–2348, Dec. 2017.

[24] X. Peng, C. Zou, Y. Qiao, and Q. Peng, "Action recognition with stacked Fisher vectors," in *Proc. 13th Eur. Conf. Comput. Vis.*, Sep. 2014, pp. 581–595.

[25] V. Lam, D. D. Le, S. Phan, S. Satoh, D. A. Duong, and T. D. Ngo, "Evaluation of low-level features for detecting violent scenes in videos," in *Proc. Int. Conf. Soft Comput. Pattern Recognit. (SoCPaR)*, Dec. 2013, pp. 213–218.

[26] S. Lyu and E. P. Simoncelli, "Nonlinear image representation using divisive normalization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.

[27] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, nos. 4–5, pp. 411–430, Jun. 2000.

[28] A. Vedaldi and B. Fulkerson. (2008). *VLFeat Open Source Library*. [Online]. Available: http://www.vlfeat.org/

[29] F. Perronnin, J. Sánchez, and T. Mensink, "Improving the Fisher kernel for large-scale image classification," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2010, pp. 143–156.

[30] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.

[31] S. Wang, K. Gu, X. Zhang, W. Lin, L. Zhang, S. Ma, and W. Gao, "Subjective and objective quality assessment of compressed screen content images," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 6, no. 4, pp. 532–543, Dec. 2016.

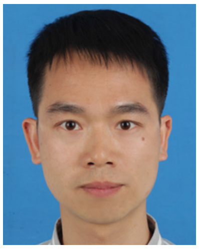
[32] Z. Ni, L. Ma, H. Zeng, Y. Fu, L. Xing, and K.-K. Ma, "SCID: A database for screen content images quality assessment," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, Nov. 2017, pp. 774–779.

[33] L. Krasula, K. Fliegel, P. Le Callet, and M. Klima, "On the accuracy of objective image and video quality models: New methodology for performance evaluation," in *Proc. Int. Conf. Quality Multimedia Exper. (QoMEX)*, Jun. 2016, pp. 1–6.

[34] A. Aldahdooh, E. Masala, O. Janssens, G. Van Wallendael, M. Barkowsky, and P. Le Callet, "Improved performance measures for video quality assessment algorithms using training and validation sets," *IEEE Trans. Multimedia*, vol. 21, no. 8, pp. 2026–2041, Aug. 2019.

[35] P. G. Gottschalk and J. R. Dunn, "The five-parameter logistic: A characterization and comparison with the four-parameter logistic," *Anal. Biochem.*, vol. 343, no. 1, pp. 54–65, Aug. 2005.

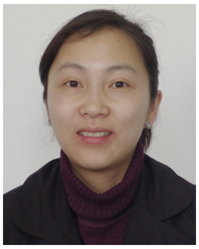
[36] Y. Bai, Z. Zhu, G. Jiang, and H. Sun, "Blind quality assessment of screen content images via macro-micro modeling of tensor domain dictionary," *IEEE Trans. Multimedia*, vol. 23, pp. 4259–4271, 2021.



YONGQIANG BAI received the B.S. and M.S. degrees from Zhengzhou University, China, in 2006 and 2009, respectively, and the Ph.D. degree from Ningbo University, China, in 2019. He is currently a Researcher with the College of Information and Intelligence Engineering, Zhejiang Wanli University, China. His research interests include data hiding, image & video processing, and artificial intelligence.

ZHONGJIE ZHU received the Ph.D. degree in electronics science and technology from Zhejiang University, China, in 2004. He is currently a Professor with the Faculty of Information and Intelligence Engineering, Zhejiang Wanli University, China. His research interests include video compression and communication, image analysis and understanding, data hiding, and 3D image signal processing.

CONGHUI ZHU graduated from the Zhejiang Business College, China, in 2019. He is currently studying with Zhejiang Wanli University, China. His research interests include image signal processing and artificial intelligence.



YUER WANG received the M.S. degree from Shanghai Fisheries University, China, in 2007. She is currently with Zhejiang Wanli University, China. Her research interests include digital video compression and signal processing, and information hiding.