

Research Article

Hierarchical Feature Extraction Assisted with Visual Saliency for Image Quality Assessment

Ruizhe Deng, Yang Zhao, and Yong Ding

College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China

Correspondence should be addressed to Yong Ding; dingy@vlsi.zju.edu.cn

Received 14 June 2017; Revised 11 August 2017; Accepted 20 August 2017; Published 3 October 2017

Academic Editor: Shang-Hong Lai

Copyright © 2017 Ruizhe Deng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Image quality assessment (IQA) is desired to evaluate the perceptual quality of an image in a manner consistent with subjective rating. Considering the characteristics of hierarchical visual cortex, a novel full reference IQA method is proposed in this paper. Quality-aware features that human visual system is sensitive to are extracted to describe image quality comprehensively. Concretely, log Gabor filters and local tetra patterns are employed to capture spatial frequency and local texture features, which are attractive to the primary and secondary visual cortex, respectively. Moreover, images are enhanced before feature extraction with the assistance of visual saliency maps since visual attention affects human evaluation of image quality. The similarities between the features extracted from distorted image and corresponding reference images are synthesized and mapped into an objective quality score by support vector regression. Experiments conducted on four public IQA databases show that the proposed method outperforms other state-of-the-art methods in terms of both accuracy and robustness; that is, it is highly consistent with subjective evaluation and is robust across different databases.

1. Introduction

Image processing plays an indispensable role in our daily lives and numerous professional fields. However, there exist many factors that would potentially degrade image quality during image acquisition, compression, transmission, restoration, and other procedures [1]. Therefore, image quality assessment (IQA), which aims to automatically estimate the quality loss due to distortions, is crucial for various image processing systems in performance estimate and optimization [2]. Since human visual system (HVS) is the ultimate receiver of images, subjective IQA completed by human observers always reflects the perceptual quality of images faithfully, yet it is cumbersome, time-consuming, and unstable, resulting in its impracticability to be applied in real-time systems [3]. Thus, accurate and robust objective IQA methods that automatically evaluate the image quality are urgently needed. Generally, objective IQA methods can be divided into three classes according to the availability of the corresponding undistorted reference image, i.e., full reference (FR), reduced reference (RR), and no reference (NR) [1–3]. FR methods require full access to the distortion-free reference image;

conversely, NR methods have no access to the reference image, while RR methods make use of partial information about the reference image. In this paper, our work is confined to FR methods, which evaluate image quality by measuring the disparity between the reference and distorted images.

The early proposed FR methods, such as mean squared error and peak signal-to-noise ratio, are calculated simply according to the pixel-wise intensity. Without regarding the properties of HVS, they are widely criticized for not correlating well with subjective ratings [4]. Later explored methods that try to simulate the function of HVS, such as noise quality measure [5], are blamed for high computational complexity with small performance gain. The popular structural similarity method (SSIM) [6], along with its improved versions [7, 8], calculates the structural information fidelity to measure the quality degradation. Moreover, natural image statistics are also introduced for feature extraction [9]. Great efforts have been made to develop more comprehensive features to accurately quantify human's subjective perception upon image quality. The implementation of more complicated feature extraction schemes can be found in more recent works [10, 11].

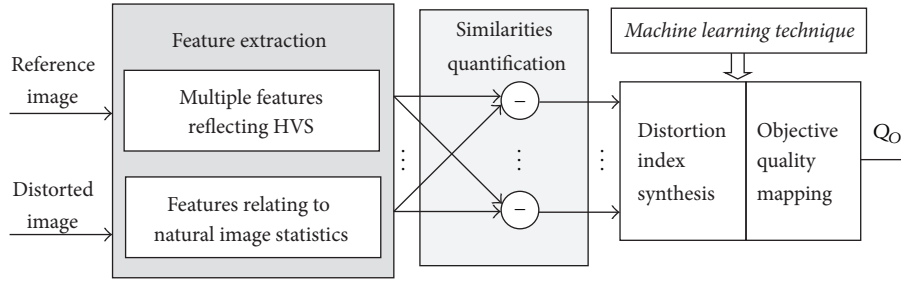


FIGURE 1: General framework of full reference IQA.

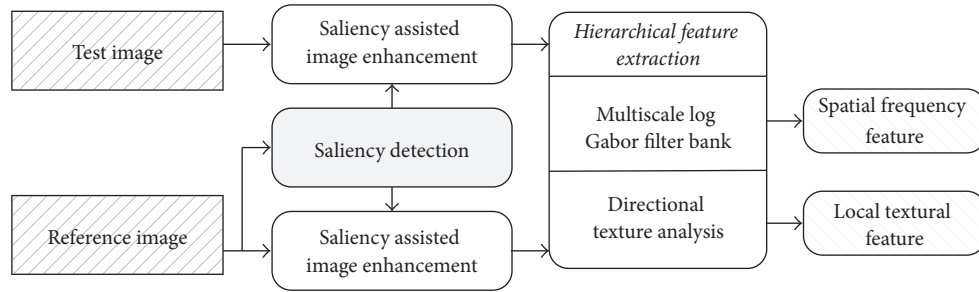


FIGURE 2: Framework of the saliency-assisted hierarchical feature extraction.

Numerous studies have confirmed that the desirable image features for IQA should be relevant to image quality closely and correlate well with human's subjective sensation [5]. Since each of the quality indexes only reveals image quality from a certain aspect, a constructive solution is to employ a comprehensive description that includes features representing quality in different aspects, which is exactly what we attempt to achieve based on the knowledge of the hierarchical properties of HVS. Specifically, different areas in visual cortex show different interests in images, and the primary and secondary visual cortex (areas V1 and V2), occupying the two largest parts in visual cortex and which are most responsible for the generation of vision, are sensitive to spatial frequency information and texture information of images, respectively [12, 13]. Based on this fact, log Gabor filters and local pattern analysis are employed in this paper to extract the two kinds of features, respectively. Moreover, the feature extraction and further implementation stages are conducted with the assistance of visual saliency information.

The contribution of this paper lies in the following: (1) a comprehensive indicator of image quality is proposed based on extracting hierarchical quality-aware features that are attractive to HVS with the assistance of visual saliency maps; (2) by statistically quantifying the difference between the quality indicators of distorted images and the corresponding reference images, an effective quality assessment method is developed, which is proved accurate and robust on multiple databases.

2. Proposed Method

The proposed method follows a three-step framework as shown in Figure 1. Firstly, features assumed to well represent image quality are extracted from the tested and reference images. Then, the similarities between these features are quantified and regarded as indices that reveal the quality of the tested image. The final step is to build a function that synthesizes the quality indices into an objective quality score Q_o .

The framework of feature extraction is shown in Figure 2, which can be divided into two stages. In the first stage, the visual important regions of the test and reference images are enhanced with the assistance of saliency map. The second stage is hierarchical feature extraction. As is explained and shown in Figure 2, the spatial frequency and local texture information are captured separately as the lower- and higher-level features. Each procedure involved will be elaborated in the sections below.

2.1. Saliency-Assisted Visual Regions Highlight. Visual attention is usually represented by visual saliency in computer vision; higher saliency value denotes more attention it receives. Distortions in the regions with more attention have a larger impact on subjective sensation than that in less attractive regions [14]. Intuitively, visual saliency is intrinsically related to IQA, since both depend on the behavior of HVS. Thus, researchers have been trying to integrate saliency models into IQA and made a significant progress [15, 16]. In

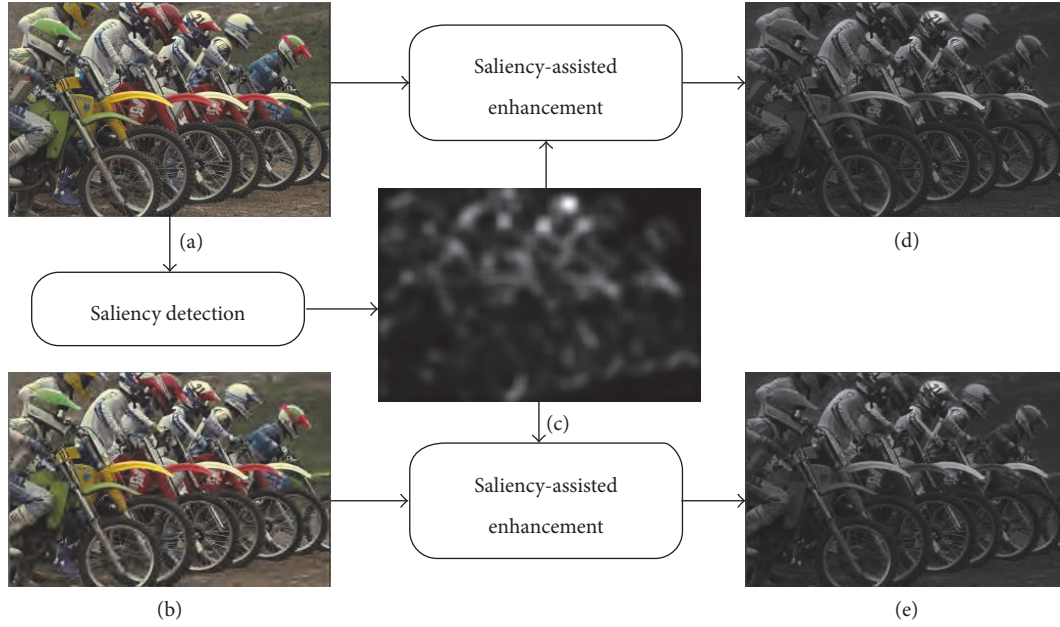


FIGURE 3: Image enhancement with the assistance of saliency map.

this paper, we use visual saliency to enhance the images by highlighting the more important regions in an image.

The saliency map is computed based on the reference image since the significant information for capturing saliency maps may be damaged in a distorted image. Among various approaches to construct saliency maps, the spectral residual (SR) visual saliency model [17] is adopted in this paper owing to its robustness and low complexity. The obtained saliency map is then combined with the images by pixel-wise multiplication,

$$I_E(x, y) = I_O(x, y) \times SM(x, y), \quad (1)$$

where SM is the saliency map, I_O is the original image, and I_E is the enhanced image. Subsequent feature extraction is operated on I_E . Figure 3 shows the effect of saliency-assisted enhancement, (a) is the reference image “bikes,” (b) is its distortion version contaminated by JPEG 2000 compression (JP2K), both taken from the LIVE database [18], (c) is the saliency map computed based on the intensity of the reference image (a), and (d) and (e) are the enhanced reference and distorted images. It can be observed that the luminance in more salient regions are relatively brighter than other regions.

2.2. Hierarchical Quality-Aware Feature Extraction. As is known, the hierarchical properties of the visual cortex are very complex. Each hierarchy is interested in different kinds of visual features; among them, the primary and secondary visual cortex areas (V1 and V2) are the first receivers of visual signal from eyes, occupy the two largest parts in the visual cortex, and are most responsible for the generation of early vision [12]. Therefore, we emphasize the importance of V1 and V2 in IQA problem. Since V1 is sensitive to simple

features like edge, bar, local frequency, etc., while V2 tends to be attracted to higher-level features like local texture and shape information [13], in this paper, the spatial frequency features and local texture features are captured and integrated for a comprehensive description of image quality. Concretely, the spatial frequency features are represented by the energy maps deriving from a log Gabor filter bank, and the local texture features are denoted as the coding results from local directional texture analysis.

To begin with, Gabor filters are widely used to capture spatial frequency information from images for the multiscale and multidirection properties similar with HVS. However, there exists apparent flaws with Gabor filters for the purpose of IQA application [19]. Firstly, the DC component makes the filter response depend largely on the gray scale of the smooth regions, which is undesirable because stronger responses are expected to occur in complicated regions rather than smooth bright regions. Secondly, the maximum bandwidth of Gabor filters is limited to approximately 1 octave. Thirdly, they are insufficient to cover broad spectral information with maximal spatial localization. As an alternative to Gabor filters, log Gabor filters remove the DC component and have a Gaussian transfer function on logarithmic axis, enabling it to capture information on a broader bandwidth [20]. The filters are defined in a polar coordinate; its radial and angular responses are defined as

$$\begin{aligned} LG(f) &= \exp\left(-\frac{[\log(f/f_0)]^2}{2\sigma_f^2}\right), \\ LG(\theta) &= \exp\left(-\frac{(\theta - \theta_0)^2}{2\sigma_\theta^2}\right), \end{aligned} \quad (2)$$

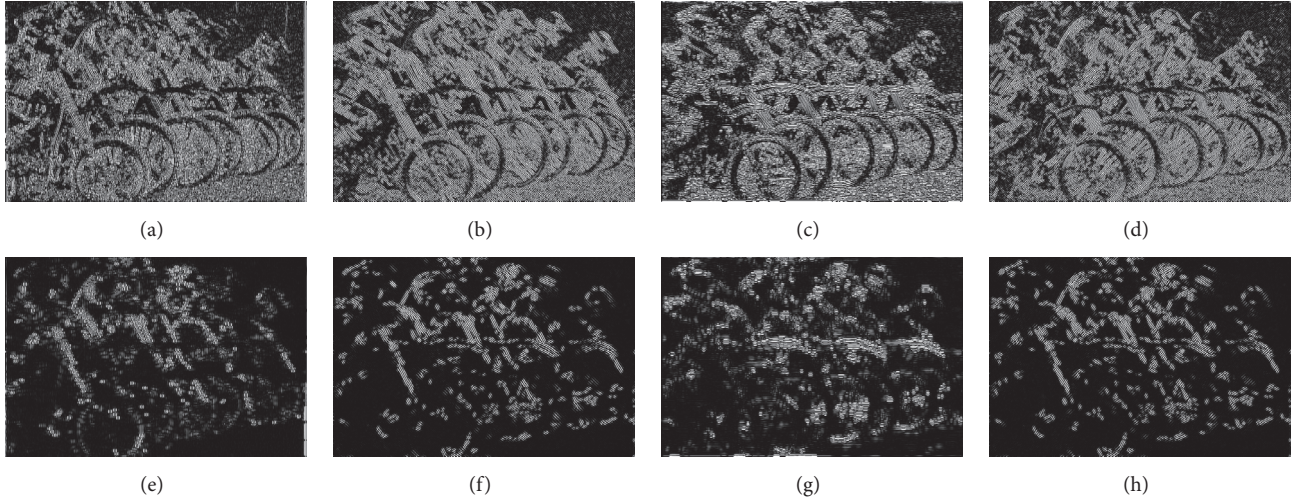


FIGURE 4: The magnitude maps of the log Gabor filter responses of an enhanced original image and its distorted version.

where (f, θ) represent the polar coordinates, f_0 and θ_0 denote the center frequency and orientation angle of the filter, and σ_f and σ_θ denote the scale and angular bandwidths, respectively. The overall frequency response is calculated as the product of the two components. With different f_0 and θ_0 , a log Gabor filter can be defined in various scales and orientations. In this paper, a 6-scale and 4-direction filter bank is involved for multiscale and multidirection feature extraction. Moreover, a fast Fourier transform (FFT) is applied to speed up the process. After filtering the redundant information in frequency domain, an inverse FFT is operated to project the responses back to spatial domain. For each filter of specific scale and orientation, the response is composed of a complex matrix, the magnitude of which is regarded as the extracted spatial frequency feature.

Given the enhanced reference image and its distorted version, shown as Figures 3(d) and 3(e), Figure 4 shows the log Gabor filter responses of them at a certain scale in four directions; that is, 0° , 45° , 90° , and 135° , (a–d) are response magnitude maps of the enhanced reference image and (e–h) are that of the distorted image. It shows that the distortion causes obvious damage on the log Gabor filter responses, indicating that log Gabor filters can act as effective indicators of image quality.

The log Gabor filter bank gives a quality description from the prospective of low-level spatial frequency information; however, it is unable to capture the higher-level texture and shape information, which is of great significance for V2 [13]. For compensation, local texture features are extracted to comprehensively describe image quality. Specifically, the first-order local tetra pattern analysis in directional perspective is adopted [21], because it captures more detailed and discriminative information than simple local pattern analysis tools like local binary pattern and is only moderately complex in computation.

Given a pixel g in an image, let I_g , I_h , and I_v denote the intensity of g and its right and below adjacent pixels, respectively; then the directions of the pixels are defined as 1,

2, 3, or 4 according to the relationship between the intensities of target pixel and its horizontal and vertical neighbors,

$$\text{Dir}(g) = \begin{cases} 1, & I_h \geq I_g, I_v \geq I_g \\ 2, & I_h \geq I_g, I_v < I_g \\ 3, & I_h < I_g, I_v < I_g \\ 4, & I_h < I_g, I_v \geq I_g. \end{cases} \quad (3)$$

In this way, the image is coded to a texture map with each item equal to an integer ranged from 1 to 4. Finally, based on the resulting direction value of pixels in image, four texture maps TM_i ($i = 1, 2, 3, 4$) are calculated as the texture features,

$$\text{TM}_i(g) = \sum_{j=1}^7 (2^j \times \text{sgn}(g_j)), \quad (4)$$

$$\text{sgn}(g_j) = \begin{cases} 1, & \text{Dir}(g_j) = i \\ 0, & \text{Dir}(g_j) \neq i, \end{cases}$$

where g_j ($j = 0$ to 7) denotes the j th pixel out of eight neighboring pixels of g .

Apparently, the distortion that an image suffers also have impact on texture patterns, as shown in Figure 5, where (a–d) are the patterns calculated from the enhanced reference image, (e–h) are calculated from the distorted image, and the involved reference and distorted images are the same with that used in Figure 4. The extracted patterns (a–d) are texture clear, while the texture information in (e–h) is apparently disrupted.

2.3. Difference Measurement of Extracted Features. It is no doubt that the inherently quality-aware information contained in extracted features will be damaged by distortion, and the difference between the features extracted from the enhanced reference and distorted images is considered as

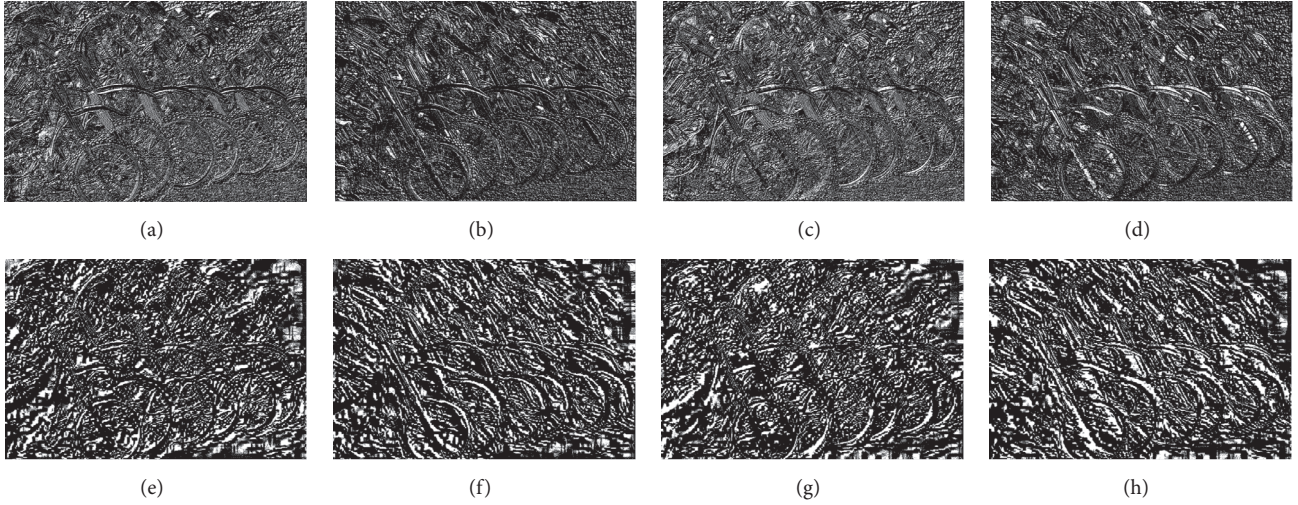


FIGURE 5: The texture maps of an enhanced reference image and its distorted version.

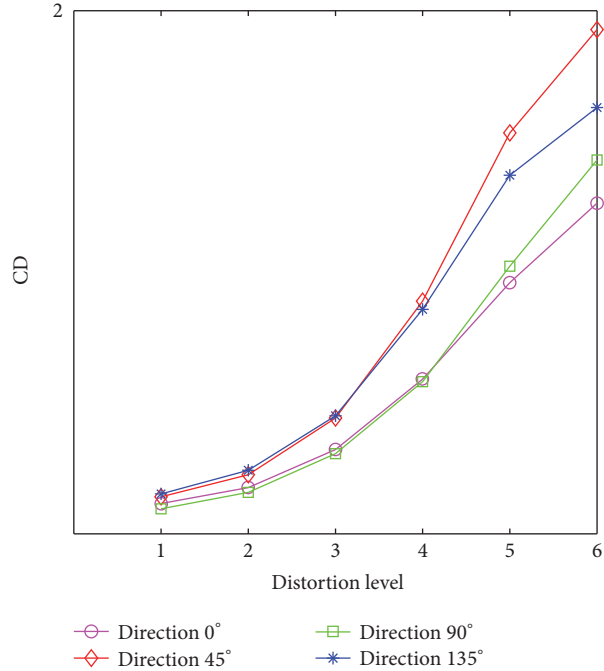


FIGURE 6: Scatter map of distortion level versus CD in different directions.

quality indices. In this paper, chi-square distance (CD) is used to measure the difference between the spatial frequency features extracted by log Gabor filters,

$$CD(LG_{f,\theta}^D, LG_{f,\theta}^R) = \frac{1}{M} \sum_{k=1}^M \frac{(LG_{f,\theta}^D(k) - LG_{f,\theta}^R(k))^2}{LG_{f,\theta}^D(k) + LG_{f,\theta}^R(k)}, \quad (5)$$

where the inputs $LG_{f,\theta}$ denote the log Gabor response at frequency f and orientation θ , superscripts D and R denote distorted and reference images, respectively, and M is the total number of items in a magnitude map.

For validation of the effectiveness of CD measurement as a quality index, Figures 6 and 7 exhibit the relation between the distortion level and CD in different directions and scales. And the reference image Figure 3(a) and its distorted images contaminated by JP2K at six levels are used to calculate the CD. The vertical axis represents CD, and the horizontal axis represents distortion level, which is determined by differential mean opinion score (DMOS). Note that, for an image in database, its mean opinion score (MOS) or differential mean opinion score (DMOS) is the subjective image quality score assigned to it obtained from experiments conducted on human observers. The features

involved in Figure 6 are the responses of log Gabor filters in four directions at the first scale, that is, 0° , 45° , 90° , and 135° , marked with different color. The features involved in Figure 7 are the log Gabor filter responses at six scales in direction 0° , with a view to demonstrate the distinction among different scales. Intuitively, with the filter scale increasing, the relative CD decreases rapidly, which is mainly because responses in higher scales abandon a lot of detailed information and the difference between the responses of reference and distorted images is thus relatively subtler. It can be clearly seen from Figures 6 and 7 that severer distortion always results in larger CD, which agrees with the human visual perception.

In addition, cosine similarity (CS) is utilized to quantify the difference between the local texture features of the reference and distorted images,

$$CS(TM_i^D, TM_i^R) = \frac{\sum_{l=1}^N (TM_i^D(l) \cdot TM_i^R(l))}{\sqrt{\sum_{l=1}^N TM_i^D(l) \cdot \sum_{l=1}^N TM_i^R(l)}}, \quad (6)$$

where the inputs TM_i denote pattern of direction i , the superscripts D and R denote distorted and reference images, respectively, and N is the total number of items in TM.

Similarly, Figure 8 further illustrates the validity of CS measurement as a quality index. Figure 8 shows a monotonic relationship between distortion level and CS; severer distortion always results in smaller CS, which demonstrate that the extracted local texture features are of great effectiveness to indicate image quality and the CS is an efficient quality index.

2.4. Objective Quality Mapping by Support Vector Regression.

For multifeature extraction based methods, a necessary operation is to construct a regression function that projects the calculated quality indices to an objective quality score. Different regression techniques such as general regression neural networks (GRNN) [22], multiple kernel learning (MKL) [23], and SVR [24] can be used to learn the regression model. In this paper, the SVR technique is adopted for its high performance on high-dimensional regression [3]. Specifically, ε -SVR is employed, and a LIBSVM package is utilized for the implementation [25].

Given a training set $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_k, y_k)\}$, where \mathbf{x}_i , $i = 1, \dots, k$, is the feature vector of the i th image in the training set of size k and y_i is the subjective quality score (MOS/DMOS), we try to find a function F that has the deviation of at most ε from y_i for all the training data with the constraint of flatness, that is, seeking a small \mathbf{w} [25]. The regression function can be represented as

$$F(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) + b, \quad (7)$$

where φ is a nonlinear function used to map the input feature vector \mathbf{x} into a high-dimensional space, \mathbf{w} is the weight vector, and b is the bias term.

Appropriate \mathbf{w} and b should be found to satisfy the following constraint:

$$|y_i - F(\mathbf{x}_i)| \leq \varepsilon. \quad (8)$$

By introducing the slack variables ξ_i and ξ_i^* , \mathbf{w} and b can be obtained by solving the following optimization problem [24]:

$$\begin{aligned} & \text{minimize} \quad \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^k (\xi_i + \xi_i^*) \\ & \text{subject to} \quad \mathbf{w}^T \varphi(\mathbf{x}_i) - (y_i - b) \leq \varepsilon + \xi_i \\ & \quad \quad \quad y_i - b - \mathbf{w}^T \varphi(\mathbf{x}_i) \leq \varepsilon + \xi_i^* \\ & \quad \quad \quad \xi_i, \xi_i^* \geq 0, \end{aligned} \quad (9)$$

and the constant C determines the trade-off between \mathbf{w} and the slack variables.

As shown in [24], \mathbf{w} can be calculated by

$$\mathbf{w} = \sum_{i=1}^{n_{sv}} (\eta_i^* - \eta_i) \varphi(\mathbf{x}_i), \quad (10)$$

where η_i^* and η_i are the Lagrange multipliers used in the Lagrange function optimization and n_{sv} is the number of support vectors. For data points satisfying (8), the corresponding η_i^* and η_i will be zero, and the training data with nonzero η_i^* and η_i are support vectors used to find \mathbf{w} . Combining (7) and (10), the regression function can be written as

$$F(\mathbf{x}) = \sum_{i=1}^{n_{sv}} (\eta_i^* - \eta_i) \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}) + b, \quad (11)$$

and then the radius basis kernel function K can be defined as

$$K = \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}) = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}\|^2}, \quad (12)$$

where γ defines the width of the kernel.

In addition, a fivefold cross-validation scheme is adopted for the training and testing procedure. Concretely, images in a database are divided randomly into five nonoverlapping sets, four of them are used for training and the remaining one for testing, that is, 80% for training and 20% for testing. The cross-validation procedure is repeated 1,000 times and the experimental results presented in this paper are the averaged value.

3. Experimental Results and Discussions

3.1. Experiment Setup. The performance of the proposed method is examined on four large-scale databases, including LIVE [18], TID2008 [26], TID2013 [27], and CSIQ [2]. Each database contains distorted images contaminated by various types of distortions at different levels, and each distorted image is assigned a subjective quality score (MOS/DMOS). The basic information of the four public databases is introduced in Table 1.

To evaluate the performance of IQA methods in terms of accuracy, monotonicity, and consistency, four commonly used metrics are calculated, including Pearson Linear Correlation Coefficient (PLCC), Spearman Rank-Order Correlation Coefficient (SROCC), Kendall Rank-Order Correlation Coefficient (KROCC), and Root Mean Squared Error

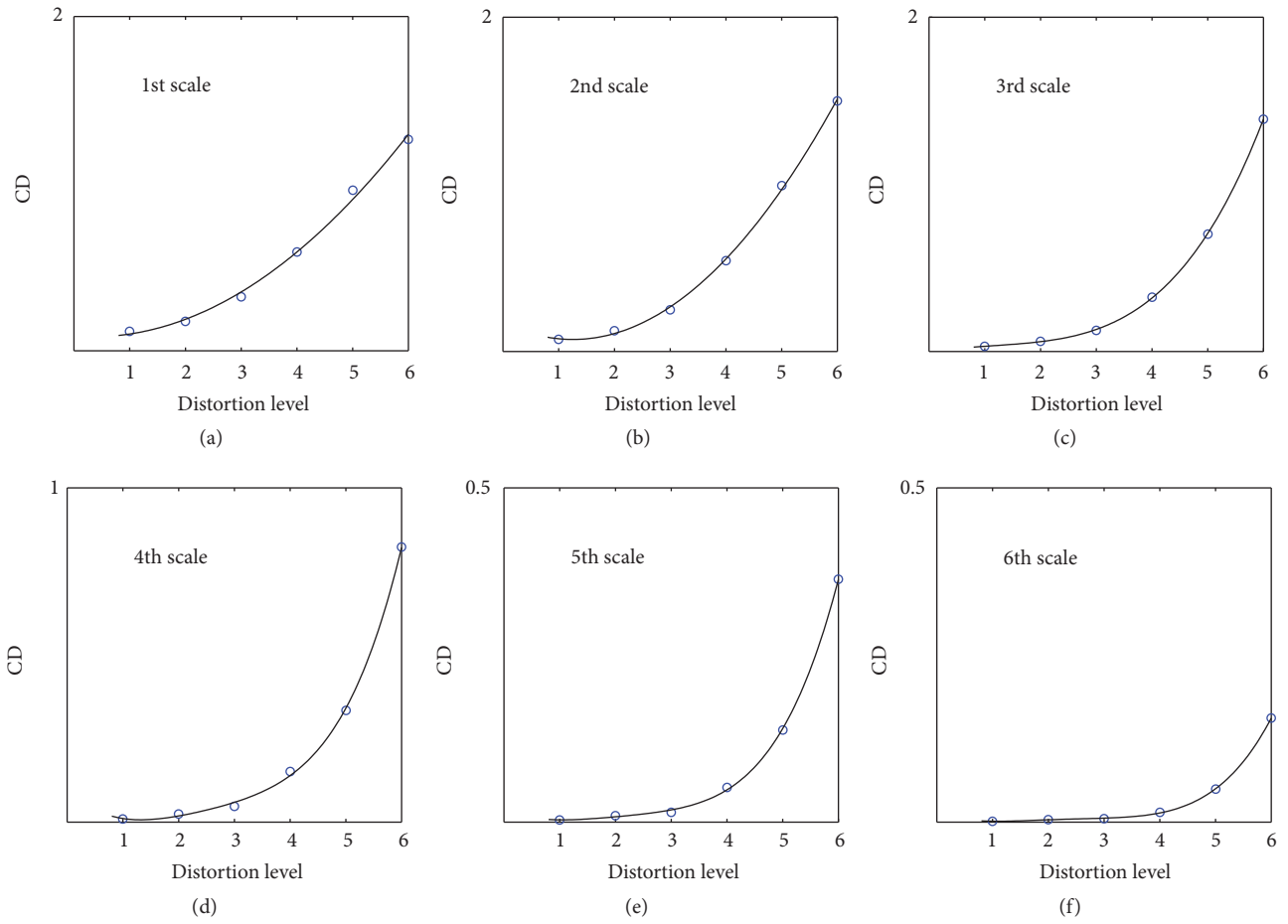


FIGURE 7: Scatter maps of distortion level versus CD at different scales.

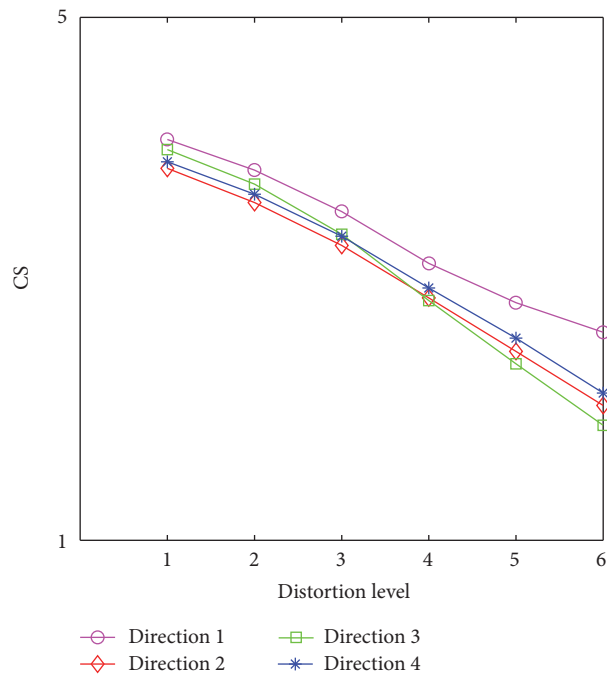


FIGURE 8: Scatter map of distortion level versus CS in different directions.

TABLE 1: Basic information of four databases.

Database	Reference images	Distorted images	Distortion types	Resolution	Observers
LIVE	29	779	5	Various	161
CSIQ	30	866	6	512 × 512	35
TID2008	25	1700	17	512 × 384	838
TID2013	25	3000	24	512 × 384	971

(RMSE) [8]. Note that a better IQA method is supposed to achieve higher PLCC, SROCC, KROCC, and lower RMSE values. PLCC indicates the linear coherency between the objective and subjective scores. For the i th image in a database of size N , given its subjective quality score S_i and predicted objective quality score O_i , PLCC and RMSE can be computed as

$$\text{PLCC} = \frac{\sum_{i=1}^N (O_i - \bar{O})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^N (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^N (S_i - \bar{S})^2}}, \quad (13)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - S_i)^2}.$$

The nonparametric rank-based correlation metrics, SROCC and KROCC, which test the coherency between the rank orders of the scores to measure the prediction monotonicity, are given by

$$\text{SROCC} = 1 - \frac{6 \sum_{i=1}^N (O_i - S_i)^2}{N(N^2 - 1)}, \quad (14)$$

$$\text{KROCC} = \frac{N_c - N_d}{(1/2)N(N-1)},$$

where N_c and N_d are the numbers of concordant and discordant pairs in the dataset, respectively.

Particularly, since the numerical ranges of the objective and subjective scores are different, and the relationship between the subjective and objective scores may not be linear due to the nonlinear quality rating of human observers, before the calculation of linear correlation metrics PLCC and RMSE, a nonlinear mapping function should be involved for a fair comparison of IQA methods as suggested in [18]. The five-parameter logistic regression function is given by

$$y = \beta_1 \left(0.5 - \frac{1}{1 + e^{\beta_2(x - \beta_3)}} \right) + \beta_4 \times x + \beta_5, \quad (15)$$

where x and y are the objective scores before and after the mapping and β_1 to β_5 are parameters obtained numerically by a nonlinear regression process in MATLAB optimization toolbox to maximize the correlations between subjective and objective scores. Since the four databases adopt different schemes to quantify subjective quality scores, β_1 to β_5 of different databases are shown in Table 2.

TABLE 2: Numerical values of parameters.

	β_1	β_2	β_3	β_4	β_5
LIVE	2.6762	1.1525	50.9363	1.0024	1.0566
TID2008	7.8189	2.6964	8.5277	0.9833	-3.1470
TID2013	-2.7263	4.6724	40.4188	1.1185	-4.7654
CSIQ	3.6708	0.9432	16.4166	0.9426	2.0596

3.2. Performance Comparison. In this section, we present the performance of the proposed method in comparison to existing methods including MAD [2], SSIM [6], VIF [9], LLM [10], MCSD [11], VSI [15], FSIMvs [16], GMSD [28], IFS [29], LCSIM [30], and GLD [31] on four public databases. The overall performance comparison results on LIVE database are presented in Table 3, with the best performance for each metric being highlighted in boldface, which shows that the proposed method outperforms other listed methods in terms of almost every performance metric. Apparently, the proposed method acquires quite large leading margin on LIVE database. The performance comparison results on TID2008 and TID2013 databases are listed in Table 4, where the best result is also highlighted in boldface. Obviously, the proposed method remains an advantage on the databases, with two exceptions appearing in the SROCC as the state-of-the-art method LLM performs slightly better than our method, yet the disparity between the two methods are quite small. Table 5 shows the performance comparison results on CSIQ database, from which we can clearly find that the proposed method exhibits the best and most stable performance among all listed methods.

Obviously, the performance of the proposed method shows evident superiority on most metrics and maintains a high level across all databases. By contrast, the previous methods are either generally inaccurate in predicting subjective evaluation or incapable to stay high level on all databases. For example, the predicting accuracy of classical methods SSIM and VIF is worse than most of the state-of-the-art methods, while the novel methods LCSIM and LLM perform very well on certain database but failed to be as competitive on other databases. Thus, it can be confirmed that the proposed method is accurate and robust.

In addition, since the performance of the proposed training-based method slightly varies during each time of experiments, a t -test with significant level at 10% is carried out on PLCC and SROCC to show whether a performance disparity is significant; the results are shown in Table 6,

TABLE 3: Performance comparison on LIVE database.

		SSIM	VIF	MAD	VSI	GMSD	IFS	MCSD	LCSIM	GLD	LLM	Our
LIVE	PLCC	0.9449	0.9604	0.9683	0.9482	0.9600	0.9586	0.9675	0.9757	0.9521	0.9578	0.9820
	SROCC	0.9479	0.9636	0.9675	0.9524	0.9603	0.9599	0.9668	0.9749	0.9631	0.9608	0.9759
	KROCC	0.7963	0.8282	0.8421	0.8058	0.8269	0.8254	0.8407	0.8600	0.8297	0.8230	0.8735
	RMSE	8.9455	7.6137	6.9073	8.6816	7.6211	7.7764	6.9079	5.9821	8.3667	7.7678	4.3403

TABLE 4: Performance comparison on TID databases.

		SSIM	FSIM _{VS}	VIF	MAD	VSI	GMSD	IFS	MCSD	LCSIM	GLD	LLM	Our
TID2008	PLCC	0.7732	0.8929	0.8084	0.8306	0.8762	0.8651	0.8810	0.8844	0.8965	0.8909	0.8971	0.9086
	SROCC	0.7749	0.8941	0.7491	0.8340	0.8979	0.8750	0.8903	0.8911	0.8892	0.8849	0.9077	0.9053
	KROCC	0.5768	0.7133	0.5860	0.6445	0.7123	0.6880	0.7009	0.7133	0.7053	0.7030	0.7368	0.7391
	RMSE	0.8511	0.5997	0.7899	0.7468	0.6466	0.6737	0.6349	0.6263	0.5945	0.6099	0.5982	0.5580
TID2013	PLCC	0.7895	NA	0.7720	0.8267	0.9000	0.8542	0.8791	0.8589	0.8651	NA	0.9068	0.9080
	SROCC	0.7417	NA	0.6769	0.7807	0.8965	0.8045	0.8697	0.8015	0.8086	NA	0.9037	0.9022
	KROCC	0.5588	NA	0.5147	0.6035	0.7183	0.6331	0.6785	0.6385	0.6292	NA	0.7209	0.7359
	RMSE	0.7608	NA	0.7880	0.6975	0.5404	0.6444	0.5909	0.6349	0.6218	NA	0.5277	0.5182

TABLE 5: Performance comparison on CSIQ database.

		SSIM	VIF	MAD	VSI	GMSD	IFS	MCSD	LCSIM	GLD	LLM	Our
CSIQ	PLCC	0.8613	0.9277	0.9502	0.9279	0.9543	0.9576	0.9560	0.9704	0.9515	0.9000	0.9704
	SROCC	0.8756	0.9195	0.9466	0.9423	0.9571	0.9581	0.9592	0.9624	0.9549	0.9050	0.9673
	KROCC	0.6907	0.7393	0.7970	0.7857	0.8122	0.8158	0.8171	0.8323	0.8108	0.7238	0.8425
	RMSE	0.1334	0.0980	0.0818	0.0979	0.0791	0.0757	0.0770	0.0634	0.0810	0.1232	0.0630

TABLE 6: Significance testing results of PLCC and SROCC.

		SSIM	VIF	MAD	VSI	GMSD	IFS	MCSD	LCSIM	GLD	LLM
PLCC	LIVE	1	1	1	1	1	1	1	1	1	1
	TID2008	1	1	1	1	1	1	1	1	1	1
	TID2013	1	1	1	1	1	1	1	1	NA	1
	CSIQ	1	1	1	1	1	1	1	1	0	1
SROCC	LIVE	1	1	1	1	1	1	1	1	1	1
	TID2008	1	1	1	1	1	1	1	1	1	-1
	TID2013	1	1	1	1	1	1	1	1	NA	-1
	CSIQ	1	1	1	1	1	1	1	1	1	1

where 0 denotes there is no significance difference between the comparing method and the proposed one and 1 or -1 represents a significant superiority or inferiority of the proposed method. The results demonstrate that except for inferiority comparing to LLM of SROCC on TID2008 and TID2013 and the nondistinguishable difference with LCSIM of PLCC on CSIQ, the proposed method shows significant superiority over other methods, indicating that the predicting accuracy of the proposed method maintains a relatively high level on these databases.

In order to test the computational cost, the running time of each IQA method processing a tested image of size 512 × 512 is listed in Table 7. All experiments are performed on a PC with Intel i5-6500 3.2 GHz CPU and 8 G RAM. Since the source codes of some methods are not openly accessible, five of the compared methods are involved in efficiency comparison. The operating system is Windows 10 and the software platform is MATLAB R2016b. As indicated in Table 7, SSIM is the most efficient method and runs much faster than others. Based on multiscale contrast similarity deviation

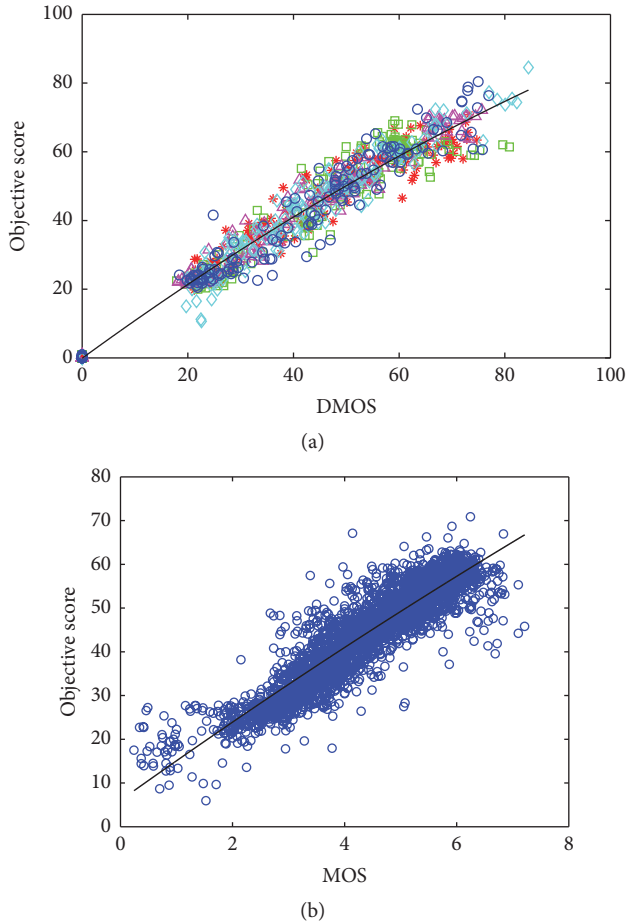


FIGURE 9: Scatter maps of objective and subjective scores on LIVE (a) and TID2013 (b) databases.

TABLE 7: Running time of the competing IQA methods.

	SSIM	VIF	MAD	VSI	GMSD	MCSD	Our
Time (s)	0.0108	0.8067	0.9866	0.2161	0.0175	0.0315	0.5859

and gradient magnitude similarity deviation quantification, the computational complexities of MCSD and GMSD are much low, so that they are so efficient compared with other IQA methods. The proposed method exploits properties of HVS to extract hierarchical features that comprehensively indicate image quality. Thus, the proposed method, which exceeds VIF and MAD in efficiency, has a relatively high complexity. However, it can be improved further through code optimization and parallelization in the future work.

To further demonstrate the effectiveness of the proposed method, scatter maps of subjective ratings versus objective scores on LIVE and TID2013 databases are given in Figure 9, where each point represents an image in database. Intuitively, the fitted curves show that the subjective scores display a substantial correlation with the objective scores, and the point cluster is closely around the fitted curve, which illustrates

that the proposed method is quite consistent with human perceptual rating.

4. Conclusion

According to the hierarchical property of visual cortex, a novel full reference IQA method is proposed in this paper. Specifically, log Gabor filters and local pattern analysis are employed to extract the hierarchical features that well reflect image quality. Moreover, the feature extraction is assisted with visual saliency maps since visual attention has great impact upon human evaluation of image quality. The experimental results show that the proposed method achieves outstanding performance in terms of prediction accuracy as well as robustness across different databases.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] J. Wu, W. Lin, G. Shi, L. Li, and Y. Fang, "Orientation selectivity based visual pattern for reduced-reference image quality assessment," *Information Sciences*, vol. 351, pp. 18–29, 2016.
- [2] 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, pp. 1–21, 2010.
- [3] 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.
- [4] Y. Zhang and D. M. Chandler, "No-reference image quality assessment based on log-derivative statistics of natural scenes," *Journal of Electronic Imaging*, vol. 22, no. 4, pp. 1–23, 2013.
- [5] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Transactions on Image Processing*, vol. 9, no. 4, pp. 636–650, 2000.
- [6] 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.
- [7] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proceedings of the 37th Asilomar Conference on Signals, Systems and Computers*, pp. 1398–1402, Pacific Grove, Calif, USA, November 2003.
- [8] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–1198, 2011.
- [9] 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.
- [10] H. Wang, J. Fu, W. Lin, S. Hu, C. J. Kuo, and L. Zuo, "Image quality assessment based on local linear information and distortion-specific compensation," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 915–926, 2017.
- [11] T. Wang, L. Zhang, H. Jia, B. Li, and H. Shu, "Multiscale contrast similarity deviation: An effective and efficient index for

- perceptual image quality assessment,” *Signal Processing: Image Communication*, vol. 45, pp. 1–9, 2016.
- [12] B. A. Olshausen and D. J. Field, “How close are we to understanding V1?” *Neural Computation*, vol. 17, no. 8, pp. 1665–1699, 2005.
- [13] N. Kruger, P. Janssen, S. Kalkan et al., “Deep hierarchies in the primate visual cortex: what can we learn for computer vision?” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1847–1871, 2013.
- [14] J. Li, L.-Y. Duan, X. Chen, T. Huang, and Y. Tian, “Finding the secret of image saliency in the frequency domain,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 12, pp. 2428–2440, 2015.
- [15] L. Zhang, Y. Shen, and H. Li, “VSI: a visual saliency-induced index for perceptual image quality assessment,” *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4270–4281, 2014.
- [16] J. Y. Lin, T. J. Liu, W. Lin, and C.-C. J. Kuo, “Visual-saliency-enhanced image quality assessment indices,” in *Proceedings of the 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA 2013*, pp. 1–4, Kaohsiung, Taiwan, November 2013.
- [17] X. D. Hou and L. Q. Zhang, “Saliency detection: a spectral residual approach,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 2280–2287, Minneapolis, MN, USA, June 2007.
- [18] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, “A statistical evaluation of recent full reference image quality assessment algorithms,” *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.
- [19] E. Walia and V. Verma, “Boosting local texture descriptors with Log-Gabor filters response for improved image retrieval,” *International Journal of Multimedia Information Retrieval*, vol. 5, no. 3, pp. 173–184, 2016.
- [20] D. J. Field, “Relations between the statistics of natural images and the response properties of cortical cells,” *Journal of the Optical Society of America A*, vol. 4, no. 12, pp. 2379–2394, 1987.
- [21] S. Murala, R. P. Maheshwari, and R. Balasubramanian, “Local tetra patterns: a new feature descriptor for content-based image retrieval,” *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2874–2886, 2012.
- [22] C. Li, A. C. Bovik, and X. Wu, “Blind image quality assessment using a general regression neural network,” *IEEE Transactions on Neural Networks*, vol. 22, no. 5, pp. 793–799, 2011.
- [23] X. Gao, F. Gao, D. Tao, and X. Li, “Universal blind image quality assessment metrics via natural scene statistics and multiple kernel learning,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 12, pp. 2013–2026, 2013.
- [24] A. J. Smola and B. Scholkopf, “A tutorial on support vector regression,” *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.
- [25] C. Chang and C. Lin, “LIBSVM: a Library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27:1–27:27, 2011.
- [26] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, “TID2008 - a database for evaluation of full-reference visual quality assessment metrics,” *Advances of Modern Radioelectronics*, vol. 10, pp. 30–45, 2008.
- [27] N. Ponomarenko, L. Jin, O. Ieremeiev et al., “Image database TID2013: peculiarities, results and perspectives,” *Signal Processing: Image Communication*, vol. 30, pp. 57–77, 2015.
- [28] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, “Gradient magnitude similarity deviation: a highly efficient perceptual image quality index,” *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684–695, 2014.
- [29] H.-W. Chang, Q.-W. Zhang, Q.-G. Wu, and Y. Gan, “Perceptual image quality assessment by independent feature detector,” *Neurocomputing*, vol. 151, no. 3, pp. 1142–1152, 2015.
- [30] M. Oszust, “Full-reference image quality assessment with linear combination of genetically selected quality measures,” *PLoS ONE*, vol. 11, no. 6, article e0158333, 2016.
- [31] A. Saha and Q. M. J. Wu, “Full-reference image quality assessment by combining global and local distortion measures,” *Signal Processing*, vol. 128, pp. 186–197, 2016.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

