PAPER

Image Quality Assessment Based on Multi-Order Local Features Description, Modeling and Quantification

Yong DING[†a)], ***Member***, **Xinyu ZHAO**[†], **Zhi ZHANG**[††], ***and*** **Hang DAI**[†††], ***Nonmembers***

SUMMARY Image quality assessment (IQA) plays an important role in quality monitoring, evaluation and optimization for image processing systems. However, current quality-aware feature extraction methods for IQA can hardly balance accuracy and complexity. This paper introduces multi-order local description into image quality assessment for feature extraction. The first-order structure derivative and high-order discriminative information are integrated into local pattern representation to serve as the quality-aware features. Then joint distributions of the local pattern representation are modeled by spatially enhanced histogram. Finally, the image quality degradation is estimated by quantifying the divergence between such distributions of the reference image and those of the distorted image. Experimental results demonstrate that the proposed method outperforms other state-of-the-art approaches in consideration of not only accuracy that is consistent with human subjective evaluation, but also robustness and stability across different distortion types and various public databases. It provides a promising choice for image quality assessment development.

key words: *image quality assessment, image quality degradation, local pattern representation, feature extraction, visual perception*

1. Introduction

With the widespread use of digital images, images are usually affected by a wide variety of distortions during acquisition, compression, storage, transmission and reproduction, resulting in perceptual quality degradation [1]–[5]. Quantification and assessment for such degradation of image quality play an important role in image processing system. Image quality assessment (IQA) can provide feedback for processing system quality monitoring, evaluation and optimization. It has become a hot topic [1]–[7].

IQA is categorized into subjective and objective assessment. Subjective assessment by human observers can provide accurate results. However, it may be influenced by several critical factors including the environment conditions, motivations and mood of observers. Moreover, subjective assessment is a laborious, expensive, time-consuming and non-repeatable process [8], [9]. Due to the weaknesses of subjective assessment, recent researches are focused on developing objective metrics to predict image quality automatically that are consistent with the subjective evaluation [1]–[3], [6]–[8].

Traditional objective assessment methods like peak of signal-to-noise ratio (PSNR) and mean squared error (MSE) are widely used. They can be easily computed and have clear physical meanings but are challenged because of low accuracy [10]. Over the last decades, various effective IQA methods have been explored in literature [11]–[20], which are conducted in two directions. One is that IQA is implemented by reflecting the way in which human beings perceive images, as images are ultimately viewed by human beings. Some meaningful human visual system (HVS) oriented objective methods are proposed by taking advantage of the known characteristics of HVS including multi-channel decomposition, just notable difference, contrast sensitivity functions, spatial and frequency error pooling, contrast and luminance masking, and so on. The well-known method is visual information fidelity (VIF) [11] which incorporates the HVS model into information fidelity criterion (IFC) [12]. IFC and VIF achieve better perceptual consistency in quality prediction for different types of distortions [13]. However, their complexity is so high that limits their applications. Besides, some interesting methods based on other properties of HVS have been proposed [14]–[21]. Visual signal-to-noise ratio (VSNR) quantifies the visual fidelity based on near-threshold and supra-threshold properties of human vision [15]. Weighted multi-scale methods are explored to evaluate the image quality at multiple resolutions cohering with human perception [16], [17]. A four-stage perceptual approach is proposed in which visual features are extracted by 2-D Gabor filter and post-processed by a divisive normalization transform to reflect the nonlinear mechanism of HVS [18]. Divisive normalization masking models are introduced into image quality assessment [19]. Instead of simulating the functional components of the lower level HVS, the methods in [20], [21] tend to include higher level aspects of HVS, such as visual attention, in objective assessment. However, since there are too many characteristics of the human visual system (HVS) unexplored at present, it is difficult to model the complex and rigorous HVS very well relying on the limited understanding, which in turn does not work well in IQA. Furthermore, computational efficiency becomes a significant issue resulting from the extremely complicated modeling.

State-of-the-art methods turn to the other direction that manages to capture the statistical properties (features) from an image and map them to the perceptual quality [3], [7],

Manuscript received June 3, 2016.
Manuscript revised February 5, 2017.
Manuscript publicized March 16, 2017.

[†]The authors are with the College of Information Science and Electronic Technology, Zhejiang University, Hangzhou, China, 310027.
[††]The author is with the Department of Electrical and Information Technology, Lund University, Lund, SE-22100, Sweden.
[†††]The author is with the Department of Computer Science, University of York, York, YO10 5GH, UK.
a) E-mail: dingy@vlsi.zju.edu.cn

DOI: 10.1587/transinf.2016EDP7244

[16], [17]. Such feature-based methods are inclined to effectively extract quality-aware features and have achieved notable success [4], [5], [13]. Among these feature-based methods, structural similarity index (SSIM) [10] is quite attractive and popular owing to its simplicity and excellent performance, which is based on the hypothesis that HVS is highly adapted for extracting structural information from images. In SSIM, the luminance, contrast and structure comparisons obtained from the reference and distorted images are involved for perceptual quality estimation [10], [22]. More recently, there are several extensive methods to improve its performance further, such as gradient-based structural similarity [23], complex wavelet structural similarity (CW-SSIM) index [24], content-partitioned structural similarity index [25], and information content weighted SSIM (IW-SSIM) [17]. However, it should be noticed that SSIM based methods are less competitive in measuring the quality of blurred images. A recent study demonstrates that gradient information that captures both contrast and structure of an image allows more emphasis on distortions around the edge regions [8], [23]. With the gradient similarity, the change of contrast and structure in images are captured [8]. Based on the fact that HVS understands an image mainly according to its low-level features, a feature similarity index uses phase congruency as the primary feature and employs the image gradient magnitude as the secondary feature [26].

From the analysis above, how to extract features effectively and map them to perceptual quality is the key issue to feature-based IQA method development. With respect to feature extraction, local description has received a lot of attention in recent years [27]–[30]. Local features perform very well in many image applications including image retrieval, object recognition and texture analysis. Such success inspires us to introduce the local feature description into image quality assessment for quality-aware feature extraction.

The main contributions of this paper include: (1) we are the first to propose a multi-order local feature description method for IQA that the first-order structure derivative and high-order discriminative information are integrated to serve as the quality-aware features (Sect. 3.1); (2) we propose a new algorithm to exploit the optimal combination of orders and optimal weighting factors for feature description in IQA (Sect. 3.2 and 3.4); (3) we conduct experiments on various distortions and different databases. The experimental results show that the proposed method outperforms state-of-the-art methods in consideration of three aspects: accuracy, stability and computational complexity, which make it very attractive and competitive for real-time applications where the computing and battery-life resource constraints should be considered carefully (Sect. 4).

The rest of this paper is organized as follows. Section 2 provides some background work and a brief review about local feature description. Sections 3 presents a dedicated method based on multi-order local features description and optimization process. Section 4 reports the experimental comparisons on different public databases with state-of-the-art methods and discussions. Finally, the conclusions are drawn in Sect. 5.

2. Related Work

Local feature detection and description have received a lot of attention in recent years [27]–[30]. With the success in effectively representing the spatial structure information of images, the importance of local descriptors such as local binary pattern (LBP) [31] and local derivative pattern (LDP) [32] has been well recognized in different application areas (for example, the object recognition [33]). We will first give a brief overview of LBP and LDP and then discuss the introduction of local features into IQA.

2.1 Local Binary Pattern

Derived from a general definition of texture in a local neighborhood, LBP is designed for texture description which is defined as a grayscale invariant texture measure to model images [31]. It has been successfully applied in face recognition, texture classification, and background modeling. The wide applications of LBP operator result from its invariance to monotonic gray-level changes and computational efficiency. The idea of LBP is to divide an image into sub-regions which include the compositions of micro-patterns [34]. By thresholding neighboring pixels with a central pixel in 3×3 local region, the LBP operator labels each pixel of an image and concatenates them into a binary string. The basic LBP can be formally represented as

$$LBP(I_C) = \sum_{i}^{N} sgn(I_i - I_C) \times 2^i \tag{1}$$

where I_C is the gray value of the central pixel, I_i are the gray value of the neighboring pixels surrounding the central pixel in a given local sub-region with i being the index and N being the number of the neighboring pixels of I_C as shown in Fig. 1. And $sgn(x)$ is a binary coding function which encodes the co-occurrence at different neighboring pixels as

$$sgn(x) = \begin{cases} 1 & \forall x \geq T \\ 0 & \forall x < T \end{cases} \tag{2}$$

where T is a threshold to determine the relationship between the central pixel and its neighboring pixels. Figure 2 shows an example of LBP for local structures and discriminating information extraction with T being set to zero.

I_1	I_2	I_3
I_8	I_C	I_4
I_7	I_6	I_5

Fig. 1 I_C and the neighbouring pixels

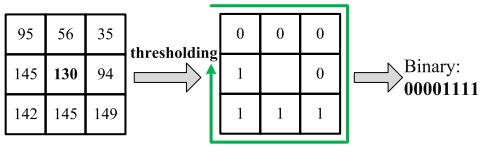

Fig. 2 Example of LBP operator

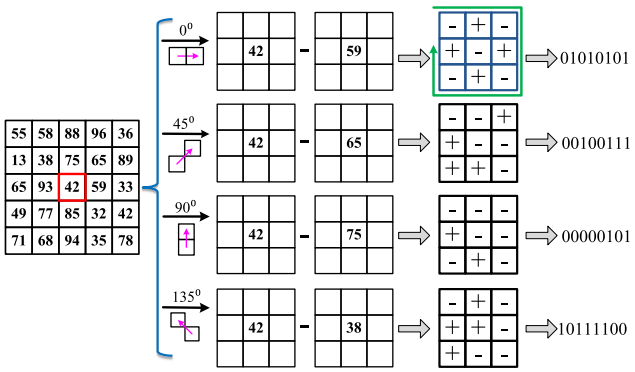

Fig. 3 Example of second-order LDP operator

2.2 Local Derivative Pattern

LBP can be considered as the nondirectional first-order circular derivative local pattern which is used to concatenate binary comparative results for generating the micro-patterns [35]. Since the first-order derivative pattern fails to combine the relationship of neighborhoods, to capture more detailed discriminative information, the high-order derivative descriptor is proposed in [32]. The notable characteristic of LDP is that it captures local information in 4 directions (0°, 45°, 90°and 135°) and concatenates the results into a 32-bit binary sequence.

Similar to the definition of LBP, the second-order LDP which captures the changes of first-order derivative information along direction α among local neighbors is defined as

$$LDP_{\alpha}^{(2)}(I_C) = \sum_{i}^{N} sgn(G_{\alpha,i} \times G_{\alpha,C}) \times 2^i \tag{3}$$

where $G_{\alpha,i}$ is the gradient along α direction that is calculated by a central pixel and its 4 neighbouring pixels, for example, $G_{0°,C} = I_4 - I_C$, $G_{45°,C} = I_3 - I_C$, $G_{90°,C} = I_2 - I_C$ and $G_{135°,C} = I_1 - I_C$, where I_C is the gray value of the central pixel, I_4, I_2, I_3 and I_1 are the gray value of the right, up, up-right and up-left neighbouring pixels, respectively. Thus, the second-order local pattern information along 4 directions is

$$LDP^{(2)}(I_C) = \{LDP_{\alpha}^{2}(I_C) | \alpha = 0°, 45°, 90°, 135°\} \tag{4}$$

An example of the second-order LDP computation is illustrated in Fig. 3. It is convenient to extend the second-order LDP to form the n^{th}-order pattern which describes the gradient trend changes in a local region of directional n^{th}-order derivative as below, pattern information along 4 directions is

$$LDP^{(n)} = \{f(G_{\alpha}^{n-1}(I_C), G_{\alpha}^{n-1}(I_1)), f(G_{\alpha}^{n-1}(I_C), G_{\alpha}^{n-1}(I_2)), \ldots, f(G_{\alpha}^{n-1}(I_C), G_{\alpha}^{n-1}(I_8)) | \alpha = 0°, 45°, 90°, 135°\} \tag{5}$$

where $G_{\alpha}^{n-1}(I_C)$ is the $(n-1)^{th}$-order derivative of the current pixel along α direction, and which encodes the $(n-1)^{th}$-order gradient transitions into binary patterns is defined as,

$$f(G_{\alpha}^{n-1}(I_C), G_{\alpha}^{n-1}(I_i)) = \begin{cases} 0 & (G_{\alpha}^{n-1}(I_C), G_{\alpha}^{n-1}(I_i)) > 0 \\ 1 & (G_{\alpha}^{n-1}(I_C), G_{\alpha}^{n-1}(I_i)) \le 0 \end{cases} \tag{6}$$

The high-order local patterns provide a stronger discriminative capability in describing detailed texture information than the first-order local pattern as used in LBP. The higher order is, the more details that can be extracted from an image by the local pattern operator [32], [35]. However, over-detailed patterns tend to be noise instead of identity information [32].

2.3 Local Feature for IQA

The local structural primitives in the early vision stage are crucial to represent image semantic information in late HVS process [36]. Structural information conveys the main visual contents of an image, and structural degradation will directly impact on image perception. Therefore, local feature description has a significant potential in the application of image quality assessment. In [37], LBP is introduced for the calculation of spatial distribution structural distribution to develop a novel full reference IQA method. However, the performance of LBP is bounded to some extent as a local structure descriptor since it lacks of magnitude information. Therefore, [36] presents the generalized local binary for quality-aware feature extraction. These local structure descriptor based methods achieve good evaluation results as well as computational simplicity. However, they only use the first-order derivative pattern for feature extraction which results in comparatively low accuracy of image quality quantification. As is well known, in image quality perception, HVS is more sensitive to detailed information than background in a scene [38]. That is to say, the high-order information (high-order characteristics, co-occurrence statistics etc.) is significant for image quality analysis and evaluation. In this paper, we introduce LDP into IQA to capture more detailed discriminative information. In the proposed method, the distributions of local pattern representation can be used as a feature descriptor. And the divergence between the features modeled by spatially enhanced histogram is taken as a measure of image quality degradation. Moreover, the optimal combination of orders for feature description is explored.

3. Image Quality Assessment Based on Multi-Order Local Description

3.1 Overview of the Proposed Framework

The proposed method is based on the hypothesis that visual perception is highly adapted for extracting structural information from a scene. Figure 4 illustrates the overall framework of the proposed image quality assessment method based on multi-order local description. Briefly, it includes three phases: feature extraction phase (Sect. 3.2) and similarity quantification, quality mapping phase (Sect. 3.3) and weighting factors optimization (Sect. 3.4). In feature extraction phase, we introduce multi-order local descriptor to obtain more detailed structure discrimination information. Then such first- and high order information is combined and encoded to construct the local pattern representation. In the similarity quantification and quality mapping phase, spatial histogram is employed to model the distributions of the local representation. Subsequently, similarity of the distributions between the reference and distorted images is quantified by a measure of chi-square distance. Finally, the similarity of all the histogram bins is synthesized with a pooling strategy to generate an objective distortion index.

3.2 Quality-Aware Features Extraction

There are three stages in feature extraction for image quality assessment: (1) pre-processing; (2) feature extraction including the first- and high-order local information; (3) representation combination. Figure 5 is an example of feature extraction based on multi-order local pattern description.

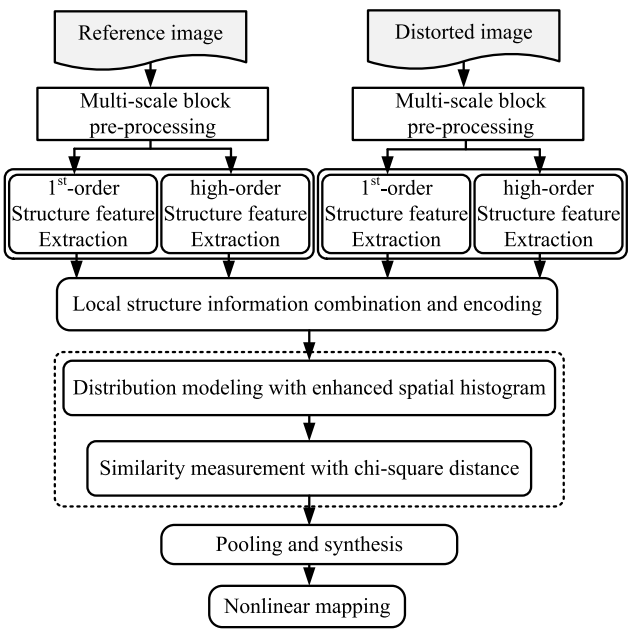

Fig. 4 Framework of the proposed method

To overcome the shortcoming of traditional local pattern methods, such as small spatial support area being not able to capture larger scale structure, low robustness against variations in position information and pixel-wise comparison being sensitive to noise, in our framework, there is a multi-scale block pre-processing procedure. The multi-scale block processing is used to extend the local region for local pattern description in which computations are done based on average values of block sub-regions, instead of individual pixels. It can make the subsequent operations very flexible to deal with the processing in different scales. In addition, the pre-processing has an advantage that it makes the subsequent process more computationally efficient. As shown in Fig. 5, the image is divided into patches with the fixed size of 30×30, and the total number of these patches are denoted as M. Then we down-sampled the 30×30 patch using a 3×3 filter so that pixels in each non-overlapped 3×3 sub-patch were averaged arithmetically. As a result, we get a down-sampled patch with the size of 10×10.

After the pre-processing, with thresholding, the first-order structure features $f^{(1)}$ are extracted from local 3×3 regions by the distribution of LBP operator as shown in Eq. (1). Then, for more detailed discriminative information extraction, the higher-order local information along 4 directions ($\alpha = 0°, 45°, 90°$ and $135°$) is captured with the distributions of Eq. (3) and its extension. Furthermore, the high-order features $f^{(2)}, \ldots$ and $f^{(n)}$ are obtained by Eq. (4) and its extension where n is the order of local pattern descriptors. The high-order local patterns provide a stronger discriminative capability in describing detailed texture information than the first-order local pattern. It can be visually observed from Fig. 5 that as the order increases, more and more details are extracted from the image.

In most cases, a single measure cannot provide enough information about the amount and spatial structure. Better

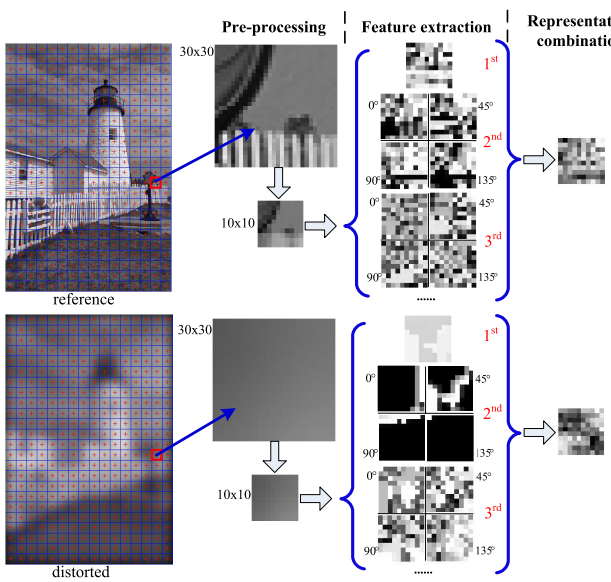

Fig. 5 Example of feature extraction

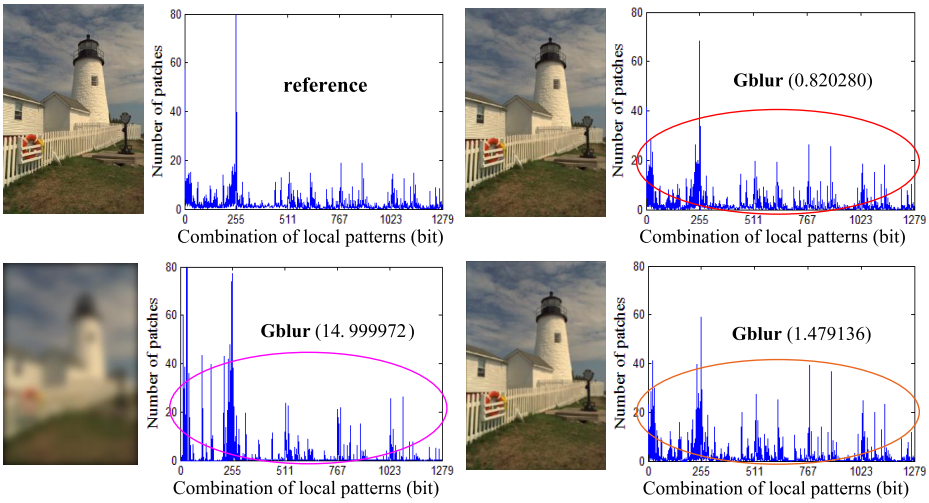

Fig. 6 Illustration of distribution divergence

structure discrimination should be obtained by considering joint occurrences of two or more features [39]. In this paper, the codes of local first- and high-order are combined and encoded into a binary string to construct the local pattern representation and the distributions of them are weighted with different factors,

$$f = \{\lambda_1 \times f^{(1)} + \lambda_2 \times f^{(2)} + \cdots + \lambda_n \times f^{(n)}\} \tag{7}$$

where λ_1, λ_2, ... and λ_n are the weighting factors for $f^{(1)}$, $f^{(2)}$, ... and $f^{(n)}$, respectively. In order to achieve the best performance, the weighting factors (λ_1, λ_2, ..., λ_n) of these different order components should be optimized, which will be introduced in Sect. 3.4.

3.3 Similarity Quantification and Quality Mapping

In image quality assessment domain, the appearance of an interested region can be well characterized by the distribution of its local representation. As well known that histogram, a global representation of the image pattern, is invariant to translation and rotation [40]. However, it is not sufficient due to its weakness in encoding spatial distributions. For the purpose of enhancing the discrimination ability, spatial histogram is an alternative because it is more robust against variations in pose or illumination than other methods [32], [34]. In this paper, we employ spatial histogram, which uses spatial templates to preserve information of both the appearance and the spatial relations in local regions. Except for the first-order features $f^{(1)}$, the higher-order local features $f_\alpha^{(n)}$ are extracted from a direction α ($\alpha = 0°$, $45°$, $90°$ or $135°$) with the n^{th}-order operator. And then a histogram of the multi-order features is combined by all these features with the size of $K \times D$, where K is the total number of binary sequences encoded from local first- and high-order representation and D is denoted as the distribution of K in a patch. That is, when n is set to 2, local first-order representation and local second-order representation are concatenated into 40-bit binary sequences and the value of K is 5×2^8. As mentioned above, the images are divided into M patches that are corresponding to M histograms. Subsequently, these M histograms are concatenated to generate the spatially enhanced histogram with the size of $M \times K \times D$. Thus, the spatial histograms $f(m, \alpha, n)$ are obtained as

$$\begin{aligned} f(m, \alpha, n) = \{ & f_\alpha^{(n)}(R_m) | m = 1, \ \ldots, \ M; \\ & \alpha = 0°, 45°, 90°, 135°; \\ & n = 1, \ \ldots, \ N\} \end{aligned} \tag{8}$$

where M is the total number of patches, and N refers to the highest order in operator. In the meanwhile, $f_\alpha^{(n)}(R_m)$ is the histogram extracted from the n^{th}-order operator in a direction α and computed independently within the m^{th} patch.

Thus, we can get three different levels of local description: structure information histograms, discriminative information histograms along four directions and concatenation of regional histograms for global image representation. As examples, Fig. 6 exhibits the joint histogram distributions of a reference image and a series of Gauss blur (Gblur) distorted images in which the first- and second-order local information is employed as features. Intuitively, the histograms of distorted images are quite different from that of the reference image. Moreover, the histograms vary with the degree of distortion. For instance, chi-square distance (introduced below) is used to show the differences numerically. The distances between the histogram of reference image and the histogram of the image filtered by a circular-symmetric 2-D Gaussian kernel of standard deviation 14.99972, 1.479136 and 0.820280 are 2.2521, 0.7119 and 0.2109, respectively. It is reasonable that such divergence between histograms may be taken as a measure for assessing image quality. That is, the quantification of this divergence can be regarded as a good approximation of perceived distortion in image quality.

Many similarity measures for histogram matching have been proposed, such as histogram intersection,

log-likelihood measure and chi-square distance [31]. Since the chi-square distance is simple in operation and our experimental results demonstrate its good performance outperforming the other methods, in this paper, it is use for measuring the similarity between two histograms in corresponding patches of the reference and distorted images. In this way, distortion index of an image is estimated as the dissimilarity.

$$Index = \sum_{m=1}^{M} \frac{w_m}{\sum_{m=1}^{M} w_m} \times \left(\frac{1}{K} \sum_{k=1}^{K} \frac{(R_{m,k} - D_{m,k})^2}{R_{m,k} + D_{m,k}}\right) \quad (9)$$

where M is the total number of patches, K is the total number of binary sequences encoded from local first- and high-order representation. $R_{m,k}$ and $D_{m,k}$ are distribution of local representation as shown in Eq. (7) of the m^{th} patch in the reference and distorted images, respectively. And w_m is the weight of different patch for pooling strategy to imitate that HVS is more sensitive to the irregular and erratic parts in an image. Since HVS does not pay an equivalent attention to different regions in an image, each region may not bear the same importance as others. For example, the weights of different patches in pooling strategy can be generated with a visual saliency map produced by a successful saliency detection model based on natural statistics [41]. The motivation behind such a saliency-based pooling approach is that visual attention is attracted to distinctive saliency regions in an image, and thus more weights should be given.

3.4 Weighting Factors Optimizaiton

The Pearson linear correlation coefficient (PLCC) and the Spearman rank-order correlation coefficient (SRCC) are used to measure the correlation between the predicted quality score and the subjective quality score. To remove any nonlinearity due to a subjective rating process and to facilitate method comparisons in a common analysis space, the objective distortion index obtained by Eq. (9) is mapped to a predicted image quality via a five-parameter logistic function [42],

$$Q(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + exp[\beta_2(x - \beta_3)]}\right) + \beta_4 x + \beta_5 \quad (10)$$

where x is the image distortion index obtained by Eq. (9), the parameters $\beta_1, \beta_2, \beta_3, \beta_4$, and β_5 are determined by minimizing the sum of squared differences between the mapped predicted image quality Q and the subjective quality score for optimization. Then PLCC and SRCC are defined as

$$PLCC = \frac{\sum_{i=1}^{n}(Q_i - \overline{Q})(S_i - \overline{S})}{\sqrt{\sum_{i=1}^{n}(Q_i - \overline{Q})^2}\sqrt{\sum_{i=1}^{n}(S_i - \overline{S})^2}} \quad (11)$$

where n is the number of the distorted images, Q_i and S_i are the predicted quality score and the human subjective quality score, respectively. $\overline{Q}$ and $\overline{S}$ are the mean of Q and S, respectively.

$$SRCC = 1 - \frac{6\sum_{i=1}^{n}(Q_i - S_i)^2}{n(n^2 - 1)} \quad (12)$$

The difinations of variables are the same as Eq. (11). In this paper, a user-defined performance metric in terms of PLCC and SRCC is introduced to serve as the optimal target for optimizing the combination of orders.

$$ST = \gamma_1 \times PLCC + \gamma_2 \times SRCC \quad (13)$$

where γ_1 and γ_2 are user-defined weights. $PLCC$ and $SRCC$ are indicators of prediction accuracy and monotonicity, respectively. Therefore, ST is the weighted summation of prediction accuracy and monotonicity. The procedure is executed to maximum such a certain performance target for optimization. And when the performance starts to degrade, the training process will stop.

As shown in Algorithm 1, the features input refer to the distributions with the weights normalized of multi-order local representations from all distorted images and the corresponding reference images. When the total number of distorted images is Z, the size of these parameters, as $D^{(n)}$, is determined to $M \times K \times Z$. In the meanwhile, the input parameters include the maximum order n and the subjective scores DMOS. When the maximum order is 1, there are only first-order features involved in the following calculation. We use Eq. (9) to get Index that is the raw objected image quality: with the structure features weights normalized, the first mean operator get the dissimilarity between corresponding patches from reference image and distorted image, and the next mean operator is used to get chi-square distance in the whole image, which is the raw objective predicted quality. In this case, in order to get ST_{opt}, the distortion index is put into Eq. (10), Eq. (11), Eq. (12) and Eq. (13) successively. Specially, Matlab function corr() is employed to calculate PLCC and SRCC. And when the order increased, the procedure is divided into three steps. The first step is to use Eq. (7) to get the features that contain structure information from different orders with weighting factors $\lambda_1, \lambda_2, \ldots, \lambda_n$ combined in W. And the next step is aimed at ST, which could refer above. The last step is comparison containing two levels. The first level is to find out the parameters to obtain the largest ST_{temp} in each order. And the next comparison of ST_{temp} among all input orders is to acquire the optimized parameter ST_{opt}. In the meanwhile, optimal weights are represented by λ vector.

Such binary pattern representation combines local texture information and high-order sensitive characteristic of HVS successfully which can reflect various distinctive spatial relationships in a local region. As shown in Fig. 5, the combined representation of the distorted image is quite different from that of the reference image. It may be potential to serve as effective features for image quality measurement. However, over-detailed patterns with high computing cost tend to be sensitive to noise. Considering the balance of computing cost and identification accuracy, it is reasonable to use limited high-order patterns for detailed information extraction. In Sect. 4, we will discuss the effectiveness of local representation for image quality assessment and determine the optimal orders and weighting factors by using Algorithm 1 on different databases and distortions.

Algorithm 1 Pseudocodes for optimal proportions-Grid_search()

Input: the first-, second-, ..., n^{th}-order features in distorted and reference images $D^{(1)}$, $D^{(2)}$, ..., $D^{(n)}$ and $R^{(1)}$, $R^{(2)}$, ..., $R^{(n)}$, respectively, the detected maximum order n and subjective scores DMOS.
Output: Optimal weights λ
Initialize: $ST_{opt} \leftarrow 0$, $\lambda \leftarrow$ zeros$(1, n)$
for $i \leftarrow 1 : n$ **do**
  if i==1 **then**
    $R \leftarrow R^{(1)}$
    $D \leftarrow D^{(1)}$
    Index←mean(mean($\frac{(R-D)^2}{(R+D)}$,2),1) //Matlab fuction mean ()
    Q ← logistic_fitting(Index) //logistic function fitting
    PLCC,SRCC ← corr(Q,DMOS) //Matlab function corr ()
    $ST_{opt} \leftarrow \gamma_1$×PLCC+$\gamma_2$×SRCC
    $\lambda(1, 1) \leftarrow 1$
  else
    $l \leftarrow [1 : 100]$
    $ST_{temp} \leftarrow 0$
    $\lambda_{temp} \leftarrow$ zeros$(1, n)$
    temp←combnk(l,i) //Matlab function combnk ()
    for p←1:size(temp,1) **do**
      W←[W; perms(temp(p,:))]
    end for
    W←[W; ones(1,i)] //produce combination of weights
    W_N←size(W,1)
    for $j \leftarrow 1$:W_N **do**
      $R \leftarrow$ zeros(size($R^{(1)}$))
      $D \leftarrow$ zeros(size($D^{(1)}$))
      for $t \leftarrow 1 : i$ **do**
        $R \leftarrow R^{(t)}$×W$(j, t) \div sum$(W$(j, :)) + R$
        $D \leftarrow D^{(t)}$×W$(j, t) \div sum$(W$(j, :)) + D$
      end for
      Index←mean(mean($\frac{(R-D)^2}{(R+D)}$,2),1)
      PLCC, SRCC ← corr(logistic_fitting(Index),DMOS)
      $ST \leftarrow \gamma_1$×PLCC+γ_2×SRCC
      if $ST > ST_{temp}$ **then**
        $ST_{temp} \leftarrow ST$
        $\lambda_{temp}(1, 1 : i) \leftarrow W(j, :)$
      end if
    end for
    if $ST_{temp} > ST_{opt}$ **then**
      $ST_{opt} \leftarrow ST_{temp}$
      $\lambda \leftarrow \lambda_{temp}$
    end if
  end if
end for
return λ

4. Experimental Results and Discussions

4.1 Database and Criteria for Evaluation

Experiments are conducted on four comprehensive databases including LIVE [43], CSIQ [44], TID2008 (TID) [45] and TID 2013 [46] (with enough samples, distortion types and observers), which have been widely used to evaluate the performance of various IQA methods. Each database consists of hundreds of color images contaminated by a variety of distortion types as well as their corresponding reference images. The characteristics of these databases are listed in Table 1. These distortions reflect a broad range of image impairments that might occur in real-world applications, from smoothing to structured distortion, image dependent distortions, and random noise. And for each image, databases provide subjective evaluation results by human subjects under controlled conditions, differential mean opinion scores (DMOS), which are obtained by psychometric tests. LIVE database is typical. In the database, there are 779 distorted images developed from a set of 29 high resolution color images that are quite representative in the content, structure, and lighting condition, and there are five types of distortions at different distortions levels including JP2K, JPEG, WN, Gblur and FF listed in Table 2.

Table 1 Databases for performance evaluation

Database name	Reference images	Distorted images	Distortion types	Subjects number
LIVE	29	779	5	161
TID2008	25	1700	17	838
TID2013	25	3000	24	917
CSIQ	30	866	6	35

Table 2 LIVE database

Distortion type		Distorted images	Meanings
JPEG	♯1	87	JPEG compression
	♯2	88	
JP2K	♯1	87	JPEG2000 compression
	♯2	82	
WN		145	white noise contamination
Gblur		145	Gauss blur
FF		145	fast fading wireless

Three criteria are employed for quantitative performance evaluations, i.e., PLCC, SRCC and RMSE (root mean square error). SRCC measures the prediction monotonicity of an IQA method, which only computes the ranking of samples. PLCC is an indicator of prediction accuracy. And RMSE is an index of prediction consistency with subjective scores.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Q_i - \overline{Q})^2} \tag{14}$$

Where n is the number of the distorted images, Q_i is the predicted quality score and $\overline{Q}$ is the mean of Q.

Both PLCC and RMSE are computed by the regression analysis methodology shown as Eq. (10). The larger values of PLCC and SRCC as well as the smaller value of RMSE mean better performance of IQA method.

For performance comparison, the proposed method is compared with some representative methods including VIF [1], MAD [4], MS-SSIM [4], GSM [8], SSIM [10], NSER [14], IW-SSIM [17], FSIMc [26], NR-GLBP [36], SILD [47] and GMSD [48].

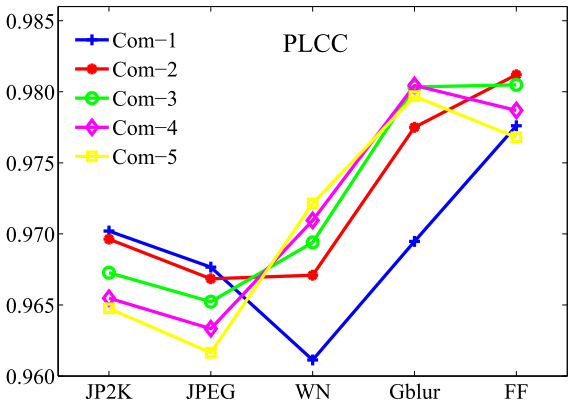

Fig. 7 PLCC comparison of different combinations

Table 3 Databases for performance evaluation

	JP2K	JPEG	WN	Gblur	FF
Com-1	0.9702	0.9677	0.9611	0.9695	0.9776
Com-2	0.9696	0.9668	0.9671	0.9775	0.9812
Com-3	0.9673	0.9652	0.9694	0.9803	0.9805
Com-4	0.9655	0.9633	0.9710	0.9805	0.9787
Com-5	0.9648	0.9616	0.9721	0.9797	0.9768

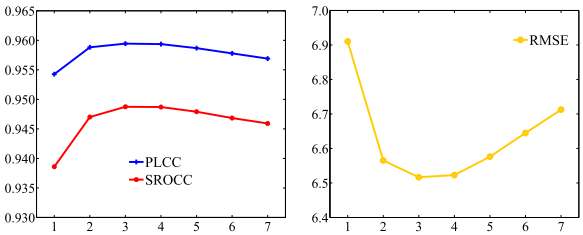

Fig. 8 Comparisons of different orders on the whole database

4.2 Parameters Optimization

There are some parameters should be chosen to optimize the performance of the proposed method including the size of multi-scale block in the preprocessing, the patch size of image division, the component of local structure information and their proportions in Eq. (7), and the weights of different patches in Eq. (9) for chi-square distance calculation. To make a good trade-off between computing cost and accuracy, in this paper, the multi-scale block is denoted as 3×3, patch size of image division is set to 30×30.

Regarding to the component of local structure information, how many orders of local patterns are involved in the combined local pattern representation is investigated in the extensive experiments. For individual distortion dataset, Fig. 7 shows the performance comparison of different components in local representation in terms of PLCC, where 'Com-1' represents there is only the first-order pattern in the local representation, 'Com-2' is a combination of the first- and second-order patterns, 'Com-3' is a combination of the first-, second- and third-order patterns, 'Com-4' is a combination of the first-, second-, third-, and fourth-order patterns, and 'Com-5' is a combination of the first-, second-, third-, fourth-, and fifth-order patterns. In LDP operator [32], with the first-order pattern representing the first-order gradient directions information, the change of derivative directions among local neighborhoods is extracted by the second-order pattern, and the turning point in a given direction is encoded. The more detailed relationship in local neighbors is abtained from the higher-order LDP patterns. Moreover, it is proved that compared with the first-order operator, higher-order operator extracts additional directional information in [49]. In this case, we combined these patterns to reserve more detailed discriminative information. It can be observed from Table 3 and Fig. 7 that 'Com-1', 'Com-2', 'Com-3', 'Com-4', and 'Com-5' revealed inconsistency as for evaluating different diction types. For JP2K and JPEG, 'Com-1' is the best; for WN, Gblur and FF, the champion is 'Com-5', 'Com-4' and 'Com-2', respectively. It is worth mentioning that the method with combinations of 'Com-3' and 'Com-4' achieve better performance improvements over that with only the first-order pattern, validating that the high-order information can improve the performance significantly. However, this trend will not continue when the fourth or much higher order information is involved in the local representation as shown in Fig. 8 which is conducted on the whole database. The reason of the performance drops when it reaches to the fourth order is that high-order local patterns are sensitive to noise which may make the structure features in an image indiscernible. From above, to balance the identification accuracy and feature length, we select a combination of 'Com-3' for detailed discriminative information extraction in image quality assessment. After proportions opmizations with Algorithm 1, the proportions of the first-, second- and third-order local patterns are confirmed as 0.52, 0.13 and 0.35.

4.3 Performance Comparison on Different Databases

After the parameters exploration, to provide quantitative measure on the performance of our proposed method, there are some comprehensive comparisons between the state-of-art methods and proposed method. To validate the performance and robustness of IQA schemes on different databases, the comparison on all types of distortions in the four databases is demonstrated in Table 4. The best results have been highlighted in boldface for each database.

From Table 4, it is obvious that with lower RMSE and higher PLCC and SRCC on LIVE and TID2013 databases, the proposed method achieves the best performance, and it performs better than the other methods except GMSD and FSIMc on TID2008. Besides, the proposed method has a comparable performance with CMSD on CSIQ database. In accordance with neurons in the early visual areas, local feature description extracts image features over small local regions [50]. The first-order operator that extracts edges information [51] and high-order operator conveying more

Table 4 Performance comparison on different databases

Database	Criteria	PSNR	SSIM	MS-SSIM	IW-SSIM	FSIMc	VIF	GSM	GMSD	Proposed
LIVE	PLCC	0.8723	0.9449	0.9409	0.8900	0.9503	0.9598	0.9437	0.9678	**0.9760**
	SRCC	0.8756	0.9479	0.9513	0.8901	0.9599	0.9632	0.9554	0.9605	**0.9653**
	RMSE	13.3597	8.9454	9.2593	7.3413	7.2002	7.6670	9.0376	7.6200	**4.9604**
TID2008	PLCC	0.5309	0.7732	0.8473	0.8579	0.8762	0.8090	0.8462	**0.8788**	0.8636
	SRCC	0.5245	0.7749	0.8549	0.8559	0.8840	0.7496	0.8554	**0.8907**	0.8563
	RMSE	1.1372	0.8511	0.7127	0.6895	0.6468	0.7888	0.7151	**0.6404**	0.6705
TID2013	PLCC	0.6748	0.7895	0.8367	0.8319	**0.8769**	0.7760	0.8229	0.8590	0.8732
	SRCC	0.6869	0.7417	0.7853	0.7779	0.8510	0.6776	0.8019	0.8044	**0.8530**
	RMSE	0.9149	0.7608	0.6789	0.6880	0.5959	-	0.7044	0.6346	**0.5988**
CSIQ	PLCC	0.8276	0.8612	0.8983	0.9144	0.9192	0.9227	0.8979	**0.9541**	0.9260
	SRCC	0.8389	0.8756	0.9121	0.9213	0.9310	0.9195	0.9126	**0.9570**	0.9205
	RMSE	0.1474	0.1334	0.1153	0.1063	0.1034	0.0980	0.1156	**0.0786**	0.0997
Average	PLCC	0.7264	0.8422	0.8808	0.8736	0.9057	0.8669	0.8777	**0.9142**	0.9097
	SRCC	0.7315	0.8350	0.8759	0.8613	**0.9065**	0.8275	0.8813	0.9032	0.8988
	RMSE	3.8898	2.6727	2.6916	2.2063	2.1366	-	2.6432	2.2434	**1.5824**

knowledge of texture regions [52] are combined to generate more sufficient accurate quality-aware features. The experimental results reveal that local feature description has several properties that favor its usage in detailed discriminative information extraction for image quality assessment.

The robustness of the proposed scheme on different databases is also demonstrated in Table 4. Note that the SSIM and VIF perform well on LIVE database but not on TID2008 and TID2013 databases. Likewise, GSM is good for TID2008 and TID2013 but is relatively poor on CSIQ database. Overall, FSIMc, GMSD, and the proposed scheme give more consistent and stable performances across all the four databases in comparison with the other schemes. Among the three schemes, GMSD performs slightly less well on TID2013 and LIVE; FSIMc only gives a comparable good performance on all the databases, whereas the proposed one performs slightly less well on TID2008 and CSIQ.

4.4 Performance Evaluation on Individual Distortion Type

To further examine the performance and robustness of the IQA schemes on each distortion type, a thorough performance evaluation of our method is conducted on LIVE database. As shown in Eq. (10), a nonlinear mapping between objective scores obtained by the proposed objective method and subjective quality ratings, is established by fitting a logistic curve. Figure 10 exhibits the subjective ratings of perception versus predicted values for each type of distortion, where each point represents one distorted image. And Fig. 9 presents the scatter plots of the objective scores for the entire LIVE database of the proposed method versus subjective DMOS. If the predicted score reflects the DMOS faithfully, scatter plots should be close to the fitted curve. As a result, the more scatter plots gather around the fitted curve, the better the method is. It is evident that the scatter plots in Fig. 9 and Fig. 10 are quite close to the fitted curve for each of the distortion sets as well as for the entire database, indicating that our proposed method is consistent well with human perception.

Moreover, detailed comparisons in terms of PLCC, SRCC and RMSE on different distortion types in LIVE database are listed in Table 5. As illustrated in Table 5,

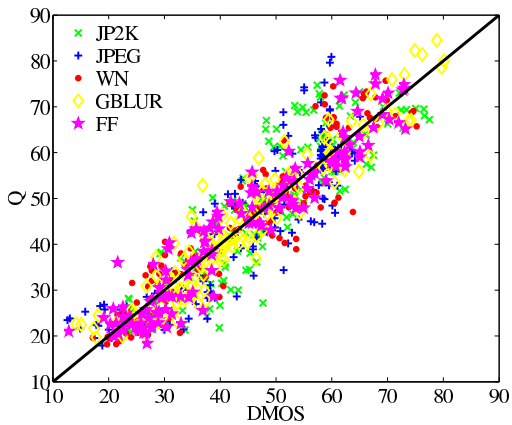

Fig. 9 Scatter plots of DMOS versus objective value on entire database

Table 5 Performance comparison with representative methods

	Methods	JP2K	JPEG	WN	Gblur	FF
PLCC	SSIM	0.936	0.928	0.964	0.873	0.943
	MS-SSIM	0.958	0.943	0.986	0.958	0.935
	NSER	0.965	0.981	0.984	0.955	0.913
	GSM	0.956	0.984	0.960	0.945	0.933
	VIF	0.963	0.942	**0.988**	0.974	0.883
	SILD	0.960	**0.986**	0.987	0.965	0.949
	Proposed	**0.974**	0.972	0.985	**0.975**	**0.981**
SRCC	SSIM	0.932	0.910	0.963	0.894	0.941
	MAD	0.938	0.949	0.971	0.899	0.886
	MS-SSIM	0.954	0.911	0.978	0.954	0.935
	FSIMc	**0.972**	**0.984**	0.972	0.971	0.952
	NSER	0.957	0.977	0.980	0.949	0.919
	GSM	0.970	0.978	0.977	0.951	0.940
	VIF	0.968	0.984	**0.984**	**0.972**	0.965
	SILD	0.954	0.979	0.981	0.961	0.950
	GMSD	0.971	0.978	0.974	0.957	0.942
	Proposed	0.967	0.946	0.973	0.956	**0.967**
RMSE	SSIM	6.424	6.509	5.135	5.861	8.430
	MS-SSIM	4.671	5.353	**2.671**	4.528	5.872
	VIF	4.963	5.372	4.360	**3.990**	6.855
	Proposed	**3.642**	**4.466**	2.739	4.358	**3.848**

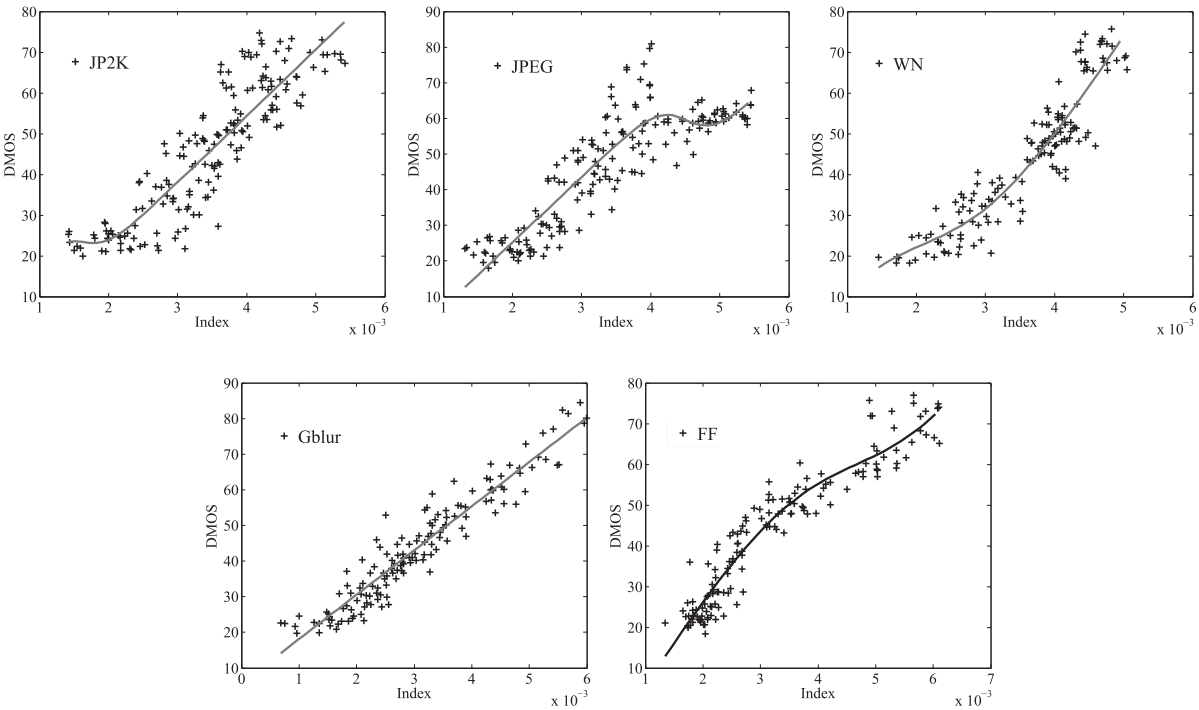

Fig. 10 Scatter plots of DMOS versus objective value across different distortions

Table 6 Execution time comparison (in second per image)

Method	SSIM	MS-SSIM	SVR	IW-SSIM	GSM	FSIM	VIF	Proposed
Time	0.1067	0.3176	22.0830	1.3522	0.2370	1.6181	3.8190	0.5429

FSIMc, SILD, VIF and the proposed method achieve more competitive performance than MS-SSIM and SSIM. However, the factors used to scale each saliency region in FSIMc are generated based on the subjective study carried out, thus, it may not be easily applied to general situations. By contrast, in the proposed method, the weights in pooling strategy and the factors for feature fusion are generated adaptively and automatically by a saliency detection model and a Grid-search Process. Furthermore, the quality-aware representation in our approach combines local gradient information and derivative information with direction information. When local image information is distorted, the corresponding distribution of the quality-aware representation will be altered. Compared to NSER, GSM and SILD which obtain less prominent results for some distortion types, our scheme proves to be more effective across the distortion types. Overall, the experimental results show that the proposed method achieves comparatively higher PLCC and SRCC while lower RMSE values demonstrating its better performance not only within a given distortion type but also across different distortion types. For an example, our method maintains 2.18% to 11.68% higher prediction accuracy and 25.6% to 54.4% higher prediction consistency than SSIM. It should be noted that VIF achieves a noticeable performance at the expense of its much high computational cost, which can be found in the computational efficiency comparison presented in the next Section.

4.5 Computational Efficiency Comparison

In real-time applications, it is desired that IQA methods should not only have high accuracy but also have low computational complexity. Especially for the cases of computing and battery-life resource constraints, computing efficiency has become a critical issue for IQA development. To examine the computational complexity of different methods, we measured the average execution time required to evaluate an image of size 512×512 on a PC (3GHz Intel E5700 CPU with 2 GB RAM).

Table 6 reports the comparison of required time in seconds per image of the proposed method with that of some representational methods. In experiments, all the methods are implemented with MATLAB. Since there is no code optimization in our MATLAB implementations, the results in Table 6 are only rough estimates. It can be seen from Table 6 that VIF takes more time than the other methods resulting from its multi-scale wavelet decomposition and complex training procedure. The proposed method, by contrast, which takes 0.5429 second per image, only about 14.2% of the time taken by VIF, is more computationally efficient than VIF. Furthermore, it is faster than IW-SSIM, FSIM and SVR since no multi-scale decomposition or Laplacian

pyramid decomposition or support vector regression are involved in the proposed method. On the other hand, although SSIM and MS-SSIM take less average execution time, they are not competitive approaches when taking all the performance terms (Listed in Tables 4 and 5) into consideration. In a word, our method achieves much high computational efficiency as it exacts the local structural features with a few operations in a small neighborhood and without any training procedure.

Above all, considering the performance shown in Tables 4 to 6, resulting from its computational simplicity, the proposed method is a good trade-off of accuracy, stability and computational complexity, making it very attractive and competitive for practical applications. These experimental results confirm that local feature description has great power that favors its usage in quality-aware features extraction for image quality assessment.

5. Conclusion and Future Work

In this paper, local description is introduced into image quality assessment for quality-aware features extraction. With integrating the first-order structure derivative and high-order discriminative information into local pattern representation, a novel IQA scheme is presented in which perceptual quality degradation is estimated by quantifying the divergence between distributions of the local pattern representation. Comprehensive experimental evaluations on different databases demonstrate that the proposed method achieves better performances than the state-of-the-art methods in consideration of accuracy, stability and complexity, which makes it qualified to apply in practical systems. It is further confirm that introducing multi-order local structure representation into IQA is much meaningful. It provides a good choice for image quality assessment development.

To improve the accuracy of image quality prediction further, future work includes investigating a more sophisticated weight generation mechanism based on the analysis on characteristics of all types of distortions in the pooling stage and exploring advanced statistical image models for local information extraction.

Acknowledgments

The authors would like to acknowledge the supports from National High Technology Research and Development Program of China (2015AA016704c), National Science & Technology Support Program of China (2013BAH03B01) and Zhejiang Provincial Natural Science Foundation (LY14F020028, LY15F040001).

References

[1] D.M. Chandler, "Seven challenges in image quality assessment: past, present, and future research," ISRN Sign. Proc., vol.2013, pp.905685:1–53, Feb. 2013.
[2] M. Oszust, "Decision fusion for image quality assessment using an optimization approach," IEEE Signal Process. Lett., vol.23, no.1, pp.65–69, Jan. 2016.
[3] K. Ma, T. Zhao, K. Zeng, and Z. Wang, "Objective quality assessment for color-to-gray image conversion," IEEE Trans. Image Process., vol.24, no.12, pp.4673–4685, Dec. 2015.
[4] E.C. Larson and D.M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," J. Electron. Imag., vol.19, no.1, pp.011006:1011006:21, Jan. 2010.
[5] W. Lin and C.-C.J. Kuo, "Perceptual visual quality metrics: A survey," J. Vis. Commun. and Imag. Repr., vol.22, no.4, pp.297–312, May 2011.
[6] L.R.V. Pato, S. Vandenberghe, B. Vandeghinste, and R. Van Holen, "Evaluation of fisher information matrix-based methods for fast assessment of image quality in pinhole spect," IEEE Trans. Med. Imag., vol.34, no.9, pp.1830–1842, Sept. 2015.
[7] W.L. Wu, X.B. Gao, D.C. Tao, and X.L. Li, "Blind image quality assessment via deep learning," IEEE Trans. Neural Netw. Learn. Syst., vol.26, no.6, pp.1275–1286, June 2015.
[8] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," IEEE Trans. Image Process., vol.21, no.4, pp.1500–1512, April 2012.
[9] T.J. Liu, K.H. Liu, H.H. Liu, and S.C. Pei, "Comparison of subjective viewing test methods for image quality assessment," IEEE International Conf. on Imag. Process. (ICIP), pp.27–30, 2015.
[10] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process., vol.13, no.4, pp.600–612, April 2004.
[11] H.R. Sheikh and A.C. Bovik, "A visual information fidelity approach to video quality assessment," The First International Workshop on Video Processing and Quality Metrics for Consumer Electronics, pp.23–25, 2005.
[12] H.R. Sheikh, A.C. Bovik, and G. De Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," IEEE Trans. Image Process., vol.14, no.12, pp.2117–2128, Dec. 2005.
[13] H.R. Sheikh, M.F. Sabir, and A.C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," IEEE Trans. Image Process., vol.15, no.11, pp.3440–3451, Nov. 2006.
[14] M. Zhang, X. Mou, and L. Zhang, "Non-shift edge based ratio (NSER): an image quality assessment metric based on early vision features," IEEE Signal Process. Lett., vol.18, no.5, pp.315–318, May 2011.
[15] D.M. Chandler and S.S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," IEEE Trans. Image Process., vol.16, no.9, pp.2284–2298, Sept. 2007.
[16] D.V. Rao and L.P. Reddy, "Contrast weighted perceptual structural similarity index for image quality assessment," 2009 Ann. IEEE India Conf. (INDICON), pp.1–4, Dec. 2009.
[17] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," IEEE Trans. Image Process., vol.20, no.5, pp.1185–1198, May 2011.
[18] Y. Ding, Y. Zhang, X. Wang, X. Yan, and A.S. Krylov, "Perceptual image quality assessment metric using mutual information of gabor features," Sci. China-Inf. Sci., vol.57, no.3, May 2014.
[19] V. Laparra, J. Muñoz-Marí, and J. Malo, "Divisive normalization image quality metric revisited," J. Opt. Soc. Amer. A, Opt. Imag. Sci. Vis., vol.27, no.4, pp.852–864, April 2010.
[20] A.K. Moorthy and A.C. Bovik, "Visual importance pooling for image quality assessment," IEEE J. Sel. Topics Signal Process., vol.3, no.2, pp.193–201, April 2009.
[21] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: based on eye-tracking data," IEEE Trans. Circuits Syst. Video Technol., vol.21, no.7, pp.971–982, July 2011.
[22] H. Zhang, Y. Ding, P.W. Wu, X.T. Bai, and K. Huang, "Image quality assessment by quantifying discrepancies of multifractal spectrums," IEICE Trans. Inf. & Syst., vol.E97-D, no.9, pp.2453–2460, Sept. 2014.
[23] C. Yang, K. Kuang, G. Chen, and S. Xie, "Gradient-based structural

similarity for image quality assessment," IEEE International Conf. on Imag. Process. (ICIP), pp.2929–2932, Oct. 2006.
[24] M.P. Sampat, Z. Wang, S. Gupta, A.C. Bovik, and M.K. Markey, "Complex wavelet structural similarity: A new image similarity index," IEEE Trans. Image Process., vol.18, no.11, pp.2385–2401, Nov. 2009.
[25] C. Li and A.C. Bovik, "Content-partitioned structural similarity index for image quality assessment," Sign. Proc., Imag. Commun., vol.25, no.7, pp.517–526, Aug. 2010.
[26] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," IEEE Trans. Image Process., vol.20, no.8, pp.2378–2386, Aug. 2011.
[27] R. Lan, Y. Zhou, and Y.Y. Tang, "Quaternionic local ranking binary pattern: A local descriptor of color images," IEEE Trans. Image Process., vol.25, no.2, pp.566–579, 2016.
[28] S. Ren, X. Cao, Y. Wei, and J. Sun, "Face alignment via regressing local binary features," IEEE Trans. Image Process., vol.25, no.3, pp.1233–1245, 2016.
[29] A. Satpathy, X. Jiang, and H. Eng, "Face alignment via regressing local binary features," IEEE Trans. Image Process., vol.23, no.5, pp.1953–1964, 2014.
[30] S.R. Dubey, S.K. Singh, and R.K. Singh, "Local wavelet pattern: A new feature descriptor for image retrieval in medical ct databases," IEEE Trans. Image Process., vol.24, no.12, pp.5892–5903, 2015.
[31] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Trans. Pattern Anal. Mach. Intell., vol.24, no.7, pp.971–987, July 2002.
[32] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor," IEEE Trans. Image Process., vol.19, no.2, pp.533–544, Feb. 2010.
[33] S.F. Razavi, H. Sajedi, and M.E. Shiri, "Integration of colour and uniform interlaced derivative patterns for object tracking," IET Imag. Proc., vol.10, no.5, pp.381–390, May 2016.
[34] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol.28, no.12, pp.2037–2041, Dec. 2006.
[35] K.-C. Fan and T.-Y. Hung, "A novel local pattern descriptorlocal vector pattern in high-order derivative space for face recognition," IEEE Trans. Image Process., vol.23, no.7, pp.2877–2891, July 2014.
[36] M. Zhang, C. Muramatsu, X. Zhou, T. Hara, and H. Fujita, "Blind image quality assessment using the joint statistics of generalized local binary pattern," IEEE Signal Process. Lett., vol.22, no.2, pp.207–210, Feb. 2015.
[37] J. Wu, W. Lin, and G. Shi, "Image quality assessment with degradation on spatial structure," IEEE Signal Process. Lett., vol.21, no.4, pp.437–440, April 2014.
[38] M.F. Costa, M.T.S. Barboni, and D.F. Ventura, "Psychophysical measurements of luminance and chromatic spatial and temporal contrast sensitivity in duchenne muscular dystrophy," Psych. Neur., vol.4, no.1, pp.67–74, Jan./June 2011.
[39] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," Patt. Rec., vol.29, no.1, pp.51–59, Jan. 1996.
[40] R. Khan, C. Barat, D. Muselet, and C. Ducottet, "Spatial histograms of soft pairwise similar patches to improve the bag-of-visual-words model," Comp. Vis. and Imag. Und., vol.132, pp.102–112, March 2015.
[41] L. Zhang, M.H. Tong, T.K. Marks, H. Shan, and G.W. Cottrell, "Sun: A bayesian framework for saliency using natural statistics," J. Vis., vol.8, no.7, pp.1–20, Dec. 2008.
[42] VQEG, "Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment Phase II," [Online], http://www.vqeg.org/, Aug. 2003.
[43] H.R. Sheikh, Z. Wang, L. Cormack, and A.C. Bovik, "Live image quality assessment database release," [online]. http://live.ece.utexas.edu/research/quality
[44] E.C. Larson and D.M. Chandler, "Consumer subjective image quality database," [online]. http://vision.okstate.edu/index.php?loc=csiq
[45] N. Ponomarenko and K. Egiazarian, "Tampere image database," [online]. http://www.ponomarenko.info/tid2008.htm
[46] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.C.J. Kuo, "Color image database tid2013: Peculiarities and preliminary results," 4th European Workshop on Vis. Inf. Process. (EUVIP), pp.106–111, June 2013.
[47] H. Qi, S. Jiao, W. Lin, L. Tang, and W. Shen, "Content-based image quality assessment using semantic information and luminance differences," Elec. Lett., vol.50, no.20, pp.1435–1436, Sept. 2014.
[48] W. Xue, L. Zhang, X. Mou, and A.C. Bovik, "Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index," IEEE Trans. Image Process., vol.23, no.2, pp.684–695, 2014.
[49] S. Murala, R.P. Maheshwari, and R. Balasubramanian, "Local Tetra Patterns: A New Feature Descriptor for Content-Based Image Retrieval," IEEE Trans. Image Process., vol.21, no.5, pp.2874–2886, 2012.
[50] N. Kruger, P. Janssen, S. Kalkan, M. Lappe, A. Leonardis, J. Piater, A.J. Rodriguez-Sanchez, and L. Wiskott, "Deep hierarchies in the primate visual cortex: What can we learn for computer vision?," IEEE Trans. Pattern Anal. Mach. Intell., vol.35, no.8, pp.1847–1871, 2013.
[51] S. Murala, R.P. Maheshwari, and R. Balasubramanian, "Local tetra patterns: A new feature descriptor for content-based image retrieval," IEEE Trans. Image Process., vol.21, no.5, pp.2874–2886, 2012.
[52] K. Ghosh, S. Sarkar, and K. Bhaumik, "Understanding image structure from a new multi-scale representation of higher order derivative filters," Image and Vision Computing, vol.25, no.8, pp.1228–1238, 2007.

Yong Ding received B.S. degree and M.S. degree from School of Electronic Science & Applied Physics, Hefei University of Technology, Hefei, P.R. China in 1997 and 2000, respectively. Then, he received Ph.D. degree from College of Electronic Science & Engineering, Nanjing, P.R. China in 2008. From 2000 to 2006, he was a Senior Engineer in R & D Center of Hisense. From 2006 to 2008, he was a Senior Project Leader of Architecture Design Department in Ominivision. And from 2009, he has been an Associate Professor of Institute of VLSI Design in Zhejiang University. In 2011, he was a Visiting Scholar at Laboratory of Mathematical Methods of Image Processing of Moscow Lomonosov State University, Russia. From Feb. 2015 to Feb. 2016, he has been a Visiting Scholar at Illinois Institute of Technology, USA. His research interests concentrate on image objective quality assessment, digital video processing and associated SoC architectures. Up to now, he has authored more than 50 papers at Journals in these fields of research. And, he has made several plenary or invited talks on international conferences. Besides, he holds more than 30 Chinese patents.

Xinyu Zhao received the Bachelor degree in School of Physics, Shandong University, China, in June 2014. She is currently pursuing the Master degree at the Institute of VLSI Design, Zhejiang University, China. Her current research interests include natural image quality assessment.

Zhi Zhang received the B.S. degree in electrical engineering and the Ph.D. degree in circuits and systems from Zhejiang University, Hangzhou, China, in 2006 and 2012, respectively, and the Ph.D. degree in electronic and computer systems from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 2013. From 2012 to 2014, he was an engineer for LTE physical layer algorithms and simulations, in ST-Ericsson and Ericsson BMOD, Lund, Sweden. He is currently a postdoctoral researcher at the department of electrical and information technology, Lund University, Sweden. He is the reviewer of IEEE Transactions on Industrial Electronics, IEEE Communications Letters, IEEE Internet of Things Journal, International Journal of Distributed Sensor Networks, Mobile Information Systems, etc. His research interests include Internet of Things, energy efficient video network, high efficiency WLAN, remote wireless energy transfer and harvesting, wireless sensor network, RFID, LTE, etc.

Hang Dai received the B.S. degree in electrical engineering, Shandong University, China, in July 2012, and the master degree in circuits and systems from Zhejiang University, China, in May 2016. Currently, he is pursuing the Ph.D. degree at the Department of Computer Science, University of York, UK.