Journal of Visual Communication and Image Representation 49 (2017) 382–391

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

journal homepage: www.elsevier.com/locate/jvci

No reference image blurriness assessment with local binary patterns☆

Guanghui Yue[a,*], Chunping Hou[a], Ke Gu[b], Nam Ling[c]

[a] School of Electronic Information Engineering, Tianjin University, Tianjin, China
[b] Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
[c] Santa Clara University, Santa Clara, CA 95053, USA

ARTICLE INFO

Keywords:
Blurriness/sharpness
Image quality assessment (IQA)
No reference (NR)
Local binary pattern (LBP)

ABSTRACT

In this paper, we put forward an effective and efficient no reference image blurriness assessment metric on the basis of local binary pattern (LBP) features. In this proposal, we reveal that part of the LBP histogram bins present monotonously with the degree of blurriness. The proposed method contains the following steps. Firstly, the LBP maps of an input image are extracted with multiple radiuses. And then, the frequency of pattern histogram is analyzed before part of bins are chosen as the features. In addition, we also take the entropy of these bins as another feature. Finally, we learn the extracted features to predict the image blurriness score. Validation of the proposed method is conducted on the blurred images of LIVE-II, CSIQ, TID2008, TID2013, LIVE3D IQA Phase I and LIVE3D IQA Phase II. Experimental results demonstrate that compared with the state-of-the-art image quality assessment (IQA) methods, the proposed algorithm has notable advantage in correlation with subjective perception and computational complexity.

1. Introduction

During the acquisition, transmission and processing, digital images are inevitably suffered from diverse distortions, such as noise, compression and blur [1]. Image quality assessment (IQA) metrics aim to build mathematic models to evaluate the degree of image quality degradation [2]. In the past few years, IQA problem has been widely concerned. Generally speaking, IQA methods can be divided into two categories: subjective evaluation and objective evaluation. Subjective evaluation, which requires human observers, is effective, reliable and accurate to estimate perceptual quality. However, it is inconvenient, cost and time-consuming, and cannot be applied to online applications [1]. Compared with subjective evaluation, objective evaluation obtains more advantages, such as easy realization and low cost. So far, full reference IQA (FR-IQA) metrics are the most mature technology of objective evaluation with full access to reference image. In most cases, we cannot obtain the reference image. Thus, it is not practical to be widely applied in any occasions. As a result, the reduced reference IQA (RR-IQA) is gradually developed in case that only part of reference information are needed. However, we strongly concern about how to evaluate the quality of image in the absence of reference image. Nowadays, developing an efficient NR-IQA metric has become a hot spot and focus of research [3,4].

The objective IQA methods can be simply divided into two categories in terms of the difference of solved distortions, general-purposed method and distortion-specified method. The aim of general-purposed method is to solve any IQA problems without consideration of distortion categories. For example, Mittal et al. [5] designed a completely blind image quality method (named NIQE) for evaluating common distortions without relying on the type of them. The NIQE depends on the statistical features of space domain natural scene statistic (NSS) model. Zhang et al. [6] further improved the work and designed the quality-unaware IQA metric (named IL-NIQE). They obtained the quality of image patch through computing the Bhattacharyya-like distance from the learned multivariate Gaussian model. Finally, the image quality was obtained by average pooling. Freitas et al. [7] proposed a training-based NR-IQA method based on local ternary pattern (LTP) descriptors. However, compared to LBP, LTP is not rigidly invariant to gray-level transformation and less susceptible to noise in uniform regions. Hence, it is not competent to detect milder image degradation. In [8], the image was first filtered by the Laplacian of Gaussian (LOG) to form multi-scale subband images. Then, each subband image was encoded by LBP operator. Finally, the features were extracted from the joint LBP histograms and fed into support vector regression (SVR) network. Although these methods achieved considerable performance on common distortion types (i.e., White noise, Jpeg compression, Jpeg2000 compression and Gaussian blur), their performance might be declined when they meet other distortion type [9] or

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author.
E-mail address: yueguanghui_2014@tju.edu.cn (G. Yue).

http://dx.doi.org/10.1016/j.jvcir.2017.09.011
Received 25 October 2016; Received in revised form 24 July 2017; Accepted 22 September 2017
Available online 28 September 2017

other image category [10]. In summary, existing general-purposed methods cannot be effectively competent to each distortion category. Therefore, the distortion-specified method emerges as the times require. It is designed for specific distortion type, such as compressed, blurred and contrasted IQA problems.

In all the distorted categories, blur distortion occupies an important position and has received more and more attention. Currently, the measurement about image blurriness/sharpness can be classified into two categories, namely edge-based and transform-based methods. In the former case, the edge-based method is mainly based on the spread of edges and statistical characteristic of texture due to the fact that blurriness usually causes the variations of image content. Ferzli et al. [11] explored the mask effect of blurriness around edge and determined the just noticeable blur (JNB) in function of the local contrast. Then, the image sharpness was measured on edge spreads, which was determined by edge width and JNB. Besides, the cumulative probability of blurriness, defined as the summation of probability below blur threshold, can also represent the image degradation well [12]. With the increasing of blurriness and the spreading of the edges, more edges exceed the probability of blur threshold, P_{JNB}. Thus, the cumulative probability changes accordingly. Bahrami et al. [13] designed a fast sharpness metric by measuring the standard deviation of maximum local variation (MLV) distribution, which reflected the maximum intensity variation between the central pixel and its neighbors. In [14], perceptual sharpness index (PSI) was proposed to describe image sharpness by considering edge width and edge slope. In the latter case, transform-based metrics are mainly based on the hypothesis that the blurred image loses high frequency information compared with reference image. Vu and Chandler [15] proposed a fast image sharpness index in transform domain. In their method, the image quality was calculated as the weight average of log-energy of discrete wavelet transform (DWT) coefficients. In addition to the DWT coefficients, the moments of image reflects the structural information well. Thus, it can be reasonably applied to IQA problems [16]. In [17], the energy of image patch, which was calculated as the sum of squared non-DC Tchebichef moment values, was variance-normalized to represent the image blurriness with the guidance of saliency weight. Hassen et al. [18] utilized the hypothesis that blur disrupts the local phase coherence (LPC). Therefore, the image sharpness could be reflected by measuring the degradation of LPC strength. Gu et al. [2] designed a new sharpness metric via analysis of autoregressive (AR) parameters. They first calculated the energy difference and contrast difference in the locally estimated AR coefficients, and then quantified the image sharpness with percentile pooling to predict the overall score.

All aforementioned two categories of blurriness/sharpness methods are well-motivated and have achieved notable successfulness on most existing databases. Most of them are effective in capturing certain aspects about the impact of blur on perceived blurriness/sharpness. However, some of them ignore the following aspects: (1) there are large variations in the real-word images in terms of information and content complexity, thus the image's gradient strength is not only affected by the degree of blur but also depends on the sharp detail in the pristine images; (2) the computational complexity is relatively high due to the transform and gradient map calculation. To tackle these problems, we propose an effective and efficient blind image blurriness quality assessment method on the basis of local binary pattern (LBP).

The remaining sections of this paper are arranged as follows: Section 2 introduces the proposed LBP-based blurriness assessment method. Section 3 describes the experimental results and shows the analysis of the proposed method. Finally, the conclusions are drawn in Section 4.

2. Methodology

Fig. 1 depicts the diagram of the proposed blurriness quality assessment procedure. Two sub-procedures, the training procedure and testing procedure, consists of the proposed method. For training procedure, LBP descriptor is utilized to extract the image's structural information. Then, the histogram of LBP patterns is calculated to represent the image characteristic. Note that not all the histogram bins correlate well with the image blurriness, through the observation, we apply parts of histogram bins, which is monotonous to the degree of blurriness, as the features. Besides, we also regard the entropy of bins as another feature to reflect the blurriness. Finally, the support vector regression (SVR) procedure is adopted to learn a model from the feature vectors to subjective scores. For testing procedure, the histogram bins is computed first, then, the image's quality score is predicted by the trained model.

2.1. LBP algorithm

LBP, as an image descriptor, was first used for texture classification [19]. Due to the rotation invariant characteristic, it was further used in many areas, like face recognition [20] and human detection [21]. It considers the spatial relationship among pixels, and thus, reflects the structural and textural information to some certain. Since the structure and texture are distorted during quality degradation, LBP can be applied to IQA problem. Formally, given a pixel g_c, its LBP value can be expressed as the difference with the circularly symmetric neighborhood:

$$LBP_K = \sum_{i=0}^{K-1} \phi(g_i,g_c)\cdot 2^i \tag{1}$$

where $g_i, i = 1...,K$, denotes the gray value of circularly symmetric neighbor pixels. K is the number of the neighbors, which determines the quantization of the angular space and computational complexity. Generally speaking, the computational complexity greatly increases as K increases. $\phi(\cdot,\cdot)$ is the thresholding function and defined as:

$$\phi(g_i,g_c) = \begin{cases} 1, & \text{if } g_i - g_c \geqslant T \\ 0, & \text{if } g_i - g_c < T \end{cases} \tag{2}$$

T is the threshold, in this paper, we set it as zero. From Eq. (1), it can be intuitively observed that there are 2^K different types of LBP patterns. To reduce the number of LBP patterns, the uniformity measure can be applied due to the fact that certain patterns, defined as 'uniform' patterns, provide the major structural information with few spatial transitions [22]. A pattern is treated as uniform pattern if the circular sequence of bits contains no more than two transitions from zero to one, or one to zero. Fig. 2 depicts an example of uniform LBPs for 8-neighbor patterns. In this figure, the adjacent pixel is colored blue[1] if its intensity is larger than that of center pixel, otherwise, it is colored red. As can be seen from Fig. 2, there are 9 uniform patterns when K equals to 8. Additionally, the LBP value of uniform pattern doesn't change even the pattern is rotated. In other words, the unform pattern is invariant. Formally, the rotation invariant uniform LBP is expressed as:

$$LBP_K^{riu2} = \begin{cases} \sum_{i=0}^{K-1} \phi(g_i,g_c), & \text{if } \mu(LBP_K) \leqslant 2 \\ K+1, & \text{otherwise} \end{cases} \tag{3}$$

where the superscript $riu2$ denotes the rotation invariant uniform patterns. The uniform measure μ is calculated as the number of bitwise transitions:

$$\mu(LBP_K) = \sum_{i=1}^{K-1} \|\phi(g_i,g_c) - \phi(g_{i-1},g_c)\| + \|\phi(g_K,g_c) - \phi(g_0,g_c)\| \tag{4}$$

where $\|\cdot\|$ denotes l_2-norm operation. As can be seen in Eq. (4), the

[1] For interpretation of color in Fig. 2, the reader is referred to the web version of this article.

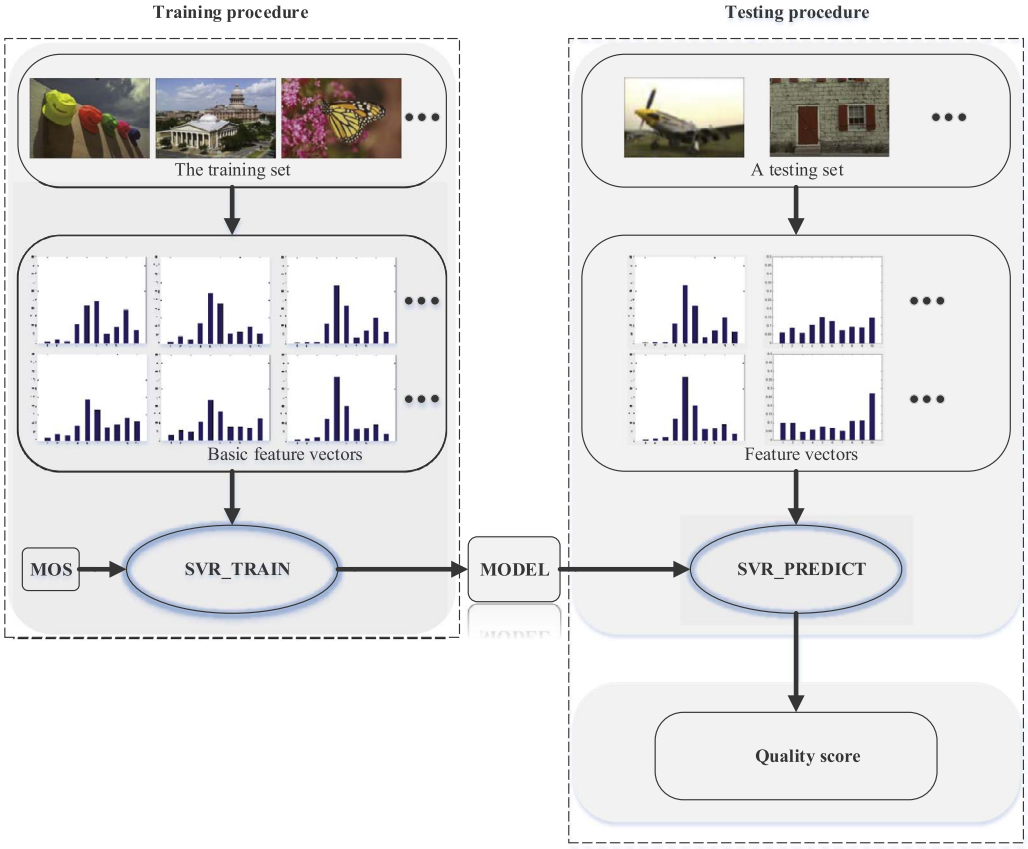

Fig. 1. Diagram of the proposed blurriness assessment method.

values of the adjacent two points in LBP pattern subtract each other, and then the difference values are summed up. With this operation, when K equals to 8, LBP patterns can form 9 uniform patterns (as shown in Fig. 2) and the other LBP patterns are labeled as the 10-th pattern. To be specific, two LBP patterns with $\mu = 0$ (as the first and ninth patterns in Fig. 2) form the first and second uniform patterns and 7×8 LBP patterns with $\mu = 2$ (seven different types (as the second to the eighth patterns in Fig. 2) multiply 8 orientations) form the third to ninth uniform patterns, while the other LBP patterns form the 10-th patterns. Generally, from Eqs. (3) and (4), we can obtain $K + 2$ LBP patterns ($K + 1$ uniform patterns and 1 nonuniform pattern). According to the LBP value of each pixel, the LBP patterns can be further expressed in form of structural histogram. As a result, an input image can be mapped into $K + 2$ bins.

2.2. Image blurriness assessment

Although LBP has been used in IQA problems [23], it is usually operated on transform domain or taken as the complementary feature. As a result, the computational complexity is high and the performance is limited. Through the statistic of a large amount of blurred images, we find that LBP histogram bins change regularly with the blurred degree, and thus, LBP can be directly employed in evaluating the blurriness of target image. In this article, we extract the LBP features directly on gray map and select part of histogram bins as the feature vectors to represent the image sharpness. To obtain a better performance, the LBP was calculated twice with different radiuses (as discussed in Part 3.4, Section 3). Fig. 3(a) shows a group of blurred images in TID2008 database [24]. The blurriness increases from left to right. Fig. 3(b) depicts

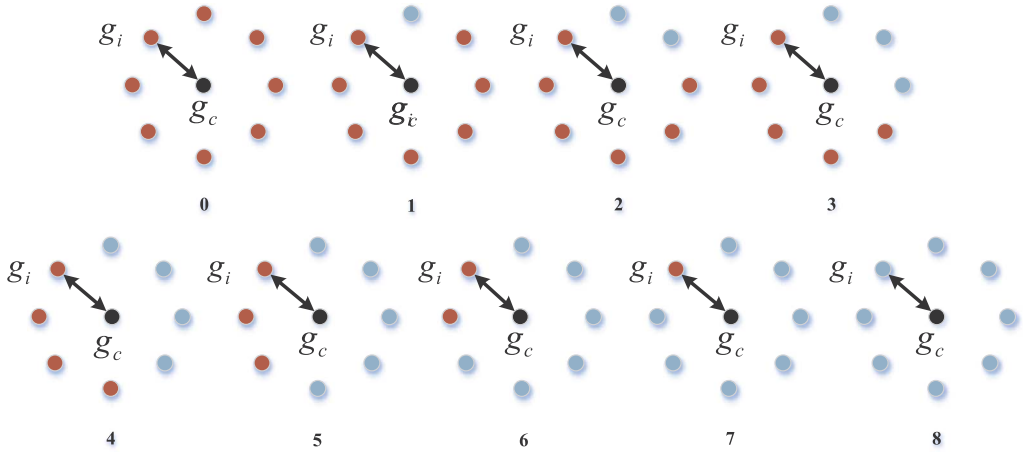

Fig. 2. Uniform LBPs for 8-neighbor patterns.

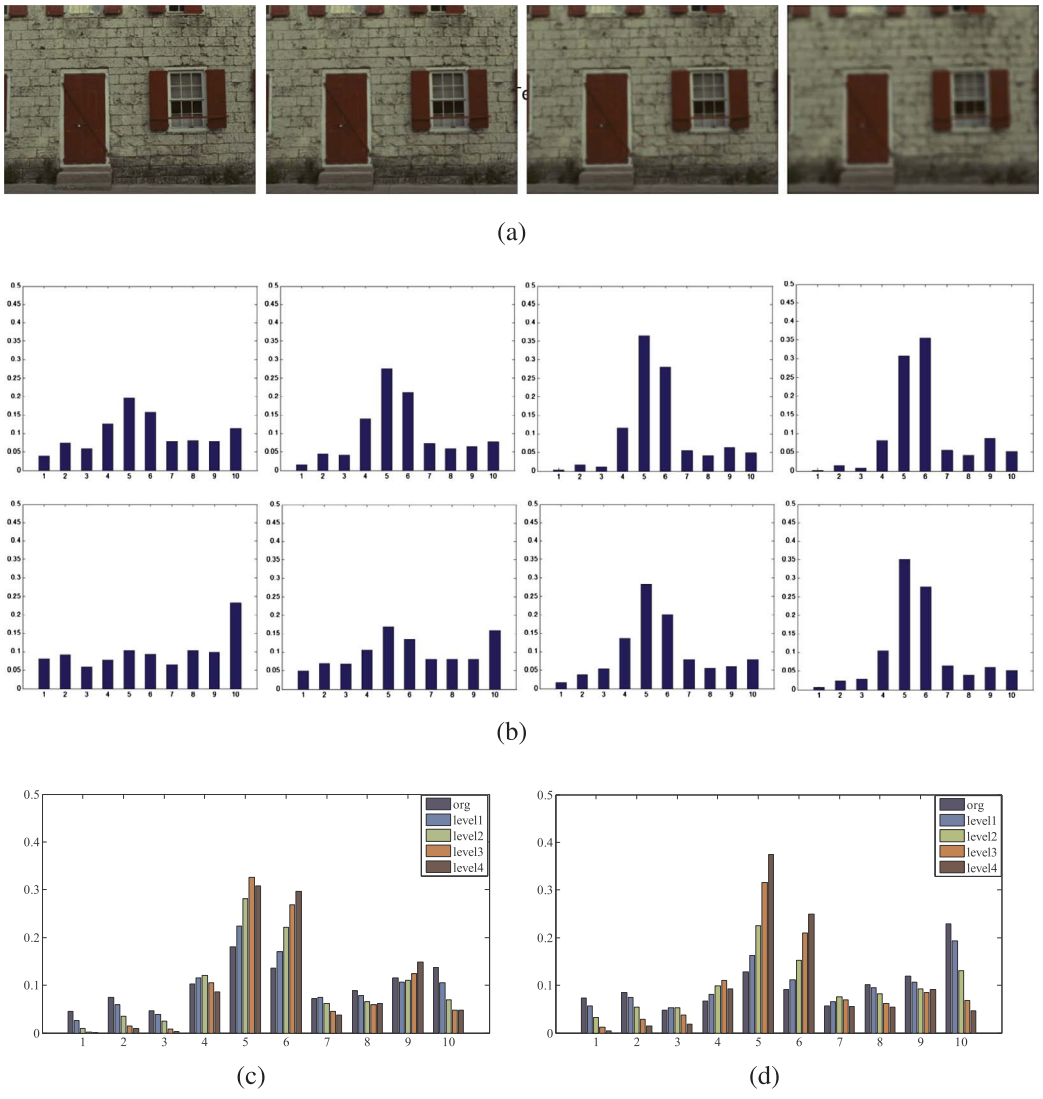

Fig. 3. LBP code distribution of reference image and blurred images in TID2008. (a) A group of blurred images in TID2008 database. The blurriness increases from left to right. (b) The corresponding LBP histograms. The radius value is one in the first row, while it is set as two in the second row. (c) and (d) are mean LBP histograms with radiuses of one pixel and two pixels, respectively. The org represents the reference image, while the others represent the blurred images. From level1 to level4, the image is more seriously blurred with increased Gaussian blur radius.

corresponding histogram of the 8-neighbors uniform LBP patterns with different radius values (one pixel and two pixels). We can find that, the histogram bins change regularly as the blurriness increases. To fully prove this observation, we analyzed the blurred image set with various blurred levels in TID2008 database. Fig. 3(c) and (d) shows the mean histograms of 8-neighbors uniform LBP patterns with same blurred level when the radius is one and two, respectively. In these figures, the "org" denotes the reference image; "level1-4" stand for the blurred images, where "level1" corresponds to the slightly blurred image, "level4" corresponds to the seriously blurred image. From the figures, we obtain the following important information: (1) not all the bins change regularly with the distortion level increased; (2) the changes of bins 1, 2, 3, and 7 in blurred image are monotonous to the blurriness (from level1 to level4), when the radius value is one; (3) while the bins 1, 2, 3, 5, 6, and 10 present a similar regularity when the radius value is two. In view of these observations, we have sufficient confidence that the LBP histogram bins can be competent to evaluate the image blurriness. Motivated by these, we utilize these observations to extract the feature vectors:

$$f_{LBP_{K,R},i} = \frac{1}{N} \cdot n(LBP^{riu2}_{K,R} i), \quad R \in \{1,2\} \\ i \in M \tag{5}$$

where i is the order of histogram bins, M is the set of selected bins. Specifically, the M contains the 1st, 2nd, 3rd, and 7th bins when radius is one and 1st, 2nd, 3rd 5th, 6th, and 10th bins when radius is two. $n(LBP^{riu2}_{K,R} i)$ is the number of rotation invariant uniform K-bit LBP pattern of type i when radius value is R. N is the total number of pixels in the image, it serves to normalize the metric so that $f_{LBP_{K,R},i} \in [0,1]$. In addition, we also find that with the degree of blurriness increasing, the distribution of bins is more complex compared to the uniform distribution. Thus, the entropy might be changed with the degree of blurriness increasing. In order to explore the regularity, we compute the entropy of bins in the same blurriness level on TID 2008, as shown in Fig. 4. As can be seen from the figure, the entropy is monotonic to blurriness. Therefore, we take the entropy of bins, f_e as another feature. Finally, the feature vector f_{LBP} is composed of $f_{LBP_{8,1},i}$ $f_{LBP_{8,2},i}$ and f_e.

After calculating the feature vectors, a regression algorithm is applied to learn a mapping from the feature vectors to quality score space. We choose LIBSVM [25] to implement the ϵ-SVR. Given a training dataset $d = \{(x_1,y_1),(x_2,y_3),\ldots (x_z,y_z)\}$, where x_i and $y_i, i \in \{1,\ldots z\}$ are the feature vector of image i and the associated MOS. The standard form of SVR is formulated as:

$$\begin{aligned} &\min_{\mathbf{w},\xi,\hat{\xi}} \ \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{z}(\xi_i + \hat{\xi}_i) \\ &s.t. \ \ \mathbf{w}^T\phi(x_i) - y_i \leqslant \epsilon + \xi_i, \\ &\quad y_i - \mathbf{w}^T\phi(x_i) \leqslant \epsilon + \hat{\xi}, \\ &\quad \xi, \hat{\xi} \geqslant 0, \quad i = 1,2,\ldots z \end{aligned} \tag{6}$$

where $K(x_i,x_j) \equiv \phi(x_i)^T\phi(x_i)$ is the kernel function that performs the

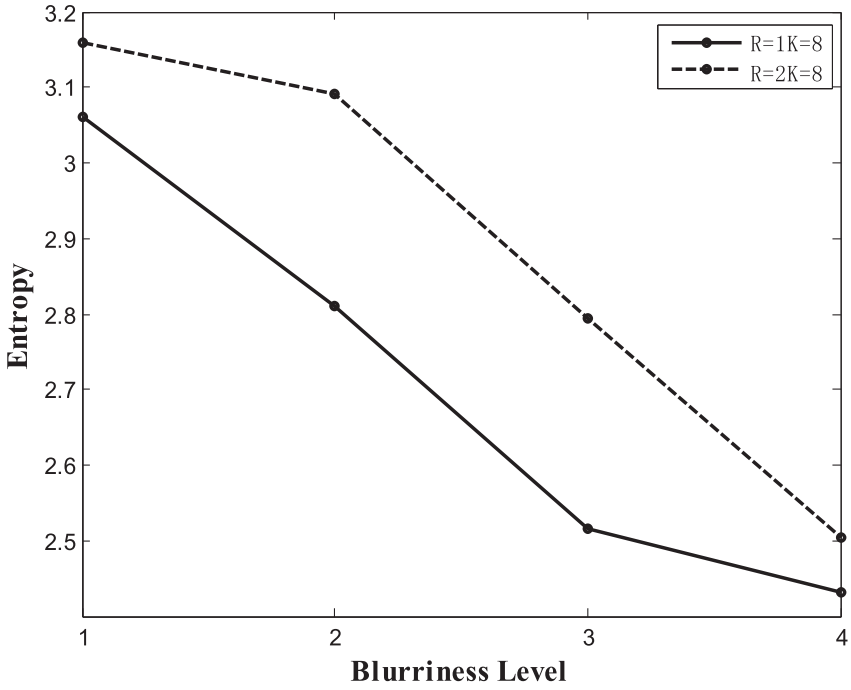

Fig. 4. Entropy plot of different blurriness levels.

nonlinear transformation. ξ and $\hat{\xi}$ are slack variables. z is the number of elements in the training sample subset. In solving the SVR, ϵ-insensitive loss function is used to ignores errors that are smaller than a certain threshold. Meanwhile, the penalty parameter C is used to control the complexity of the prediction function. Since radial basis function (RBF) has fast convergence characteristic and approximates to any nonlinear function, we chose it to measure the similarity between two samples in high dimension. The RBF can be expressed as:

$$K(x_i,x_j) = e^{-\gamma\|x_i-x_j\|^2} \tag{7}$$

where γ is the variance of the kernel function. All the parameters were determined by grid search to obtain the best performance.

3. Results and analysis

3.1. Experiment setup

To verify the performance of the proposed method, we utilized the reference and its blurred images of four available 2D IQA databases including CSIQ [26], LIVE II [27], TID2008 [24] and TID2013 [28]. All the blurred images are obtained using Gaussian low-pass filtering in these databases. The number of them are 150, 174, 100 and 125, respectively, in the four databases. The subjective scores are reported in form of the different of mean opinion scores (DMOS) in LIVE II and CSIQ, while they are represented as mean opinion scores (MOS) in TID2008 and TID2013. Besides, we also explore the effectiveness of the proposed method on 3D image blurriness assessment. Two publicly available 3D IQA databases, namely LIVE 3D IQA Database Phase I [29] and LIVE 3D IQA Database Phase II [30] are employed in this paper. In LIVE 3D IQA Database Phase I, both left and right image are blurred symmetrically, while they are blurred symmetrically and asymmetrically in LIVE 3D IQA Database Phase II.

Three commonly used criteria suggested by VQEG were employed to compare the proposed method with existing sharpness IQA metrics: (1) Pearsons linear correlation coefficient (PLCC), which indicates the prediction accuracy; (2) root-mean-squared error (RMSE); (3) Spearman rank-order correlation coefficient (SRCC), which express the monotonicity by ignoring the relative distance between the data. The five parameter logistic regression function is applied before the computation of PLCC and RMSE to reduce the nonlinearity. It is formulated as:

$$Q_m(q) = \nu_1\left(\frac{1}{2}-\frac{1}{e^{\nu_2(q-\nu_3)}+1}\right) + \nu_4 q + \nu_5 \tag{8}$$

where $Q_m(q)$ is the fitted quality score, and $\nu_l, l = 1,...,5$, are the parameters to be fitted. Theoretically, a good performance can be reflected by a perfect match of quality score and human perception score. And it can be represented as SRCC = 1, PLCC = 1 and RMSE = 0.

3.2. Performance comparison on 2D databases

In this section, we compare the proposed method with several state-of-the-art general-purposed IQA metrics, including GMSD [31], NIQE [5], IL-NIQE [6], ADD-SSIM [32], ADD-GSIM [32], NFERM [3] and SISBLIM [33]. The experimental performances are obtained through the implementation of the released demos. As the proposed method is based on SVR algorithm, thus, the dataset requires to be divided into training and testing sets. In the experiment, 80% datasets are used for training, while the remaining datasets are used for testing. The random training–testing split is repeated 1000 times to reduce the performance bias. Table 1 lists the results in terms of PLCC, SRCC and RMSE on the four 2D databases. The results are reported in forms of median. To emphasize the metrics with good performance, we bold the top two results. As expected, the proposed method achieves considerable results. Specifically, the proposed method is slightly inferior to GMSD, ADD-GSIM and produces the third best performance in terms of RMSE and SRCC in Live II database. It also produces the similar results on the rest of databases. In spite of these, our algorithm always occupies the top three positions across all databases, while other algorithms produce large fluctuation. Overall, according to the weight average values, the proposed method achieves the best performance across the four databases in terms of PLCC, and achieves the second best performance in terms of SRCC. Note that, the SRCC of the proposed method is slightly inferior to ADD-SSIM with 0.2% decrement, while PLCC value is superior to it with 2% improvement. In summary, the proposed method produces very good performance and well-matched with the state-of-the-art general-purposed IQA metrics.

In order to further test and verify the effectiveness, we also compare

Table 1
Summary of experimental results of the proposed method and compared general-purposed IQA metrics.

Database	Crietria	General purpose IQA metric							
		GMSD [31]	NIQE [5]	IL-NIQE [6]	ADD-SSIM [32]	ADD-GSIM [32]	NFERM [3]	SISBLIM [33]	Pro. method
Live II (174)	SRCC	0.9751	0.9449	0.8975	**0.9796**	**0.9773**	0.9555	0.9195	0.9704
	PLCC	**0.9788**	0.9284	0.8732	0.9615	0.9682	0.9577	0.9138	**0.9769**
	RMSE	**4.4535**	8.0815	10.6008	4.9739	4.4398	**4.5263**	8.8341	4.6143
CSIQ (150)	SRCC	0.8864	0.8945	0.8576	0.8935	0.8885	**0.8964**	0.8833	**0.9320**
	PLCC	0.8990	**0.9215**	0.7986	0.8887	0.8967	0.9201	0.8640	**0.9507**
	RMSE	0.1255	**0.1113**	0.1725	0.1314	0.1268	0.1122	0.1443	**0.0876**
TID2008 (100)	SRCC	0.8968	0.8165	0.8099	**0.9356**	**0.9157**	0.8075	0.7858	0.9038
	PLCC	0.8956	0.8376	0.8265	**0.9264**	0.9022	0.8031	0.7817	**0.9132**
	RMSE	0.5221	0.6413	0.6607	**0.4419**	0.5022	0.6992	0.6491	**0.4720**
TID2013 (125)	SRCC	0.9113	0.7968	0.8155	**0.9452**	**0.9344**	0.8498	0.8051	0.9286
	PLCC	0.9095	0.7995	0.8316	**0.9351**	0.9227	0.8505	0.8069	**0.9389**
	RMSE	0.5188	0.7495	0.6930	**0.4421**	0.4810	0.6563	0.7370	**0.4259**
Weight average	SRCC	0.9221	0.8740	0.8520	**0.9402**	0.9320	0.8883	0.8592	**0.9383**
	PLCC	0.9261	0.8806	0.8348	**0.9292**	0.9263	0.8998	0.8641	**0.9495**

Table 2
Performance comparisons with state-of-the-art NR-IQA metrics.

Metrics	TID 2008			TID 2013		
	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
ARISM [2]	0.8430	0.8505	0.6312	0.8954	0.8980	0.5556
LTP [7]	–	–	–	–	**0.9423**	–
NR-LBPS^{riu2} [8]	–	–	–	–	–	–
BIBLE [17]	0.8936	0.8915	0.5268	0.9057	0.8988	0.5290
FISHbb [15]	0.8524	0.8378	0.6135	0.8764	0.8584	0.6009
LPC-SI [18]	0.8586	0.8561	0.6017	0.8490	0.8888	0.5591
MLV [13]	0.8593	0.8548	0.6001	0.8830	0.8787	0.5858
JNB [11]	0.6932	0.6667	0.8458	0.7113	0.6787	0.8772
CPBD [12]	0.8331	0.8412	0.6429	0.8620	0.8518	0.6325
Pro. method	**0.9059**	**0.8962**	**0.4840**	**0.9310**	0.9192	**0.4470**
Metrics	LIVE II			CSIQ		
	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
ARISM [2]	0.9628	0.9679	5.8796	0.9456	0.9255	0.0932
LTP [7]	0.9485	0.9423	–	0.9212	0.9017	–
NR-LBPS^{riu2} [8]	0.9436	0.9426	–	–	–	–
BIBLE [17]	0.9728	0.9649	5.7081	0.9397	0.9132	0.0980
FISHbb [15]	0.9566	0.9597	6.3350	0.9164	0.9177	0.1147
LPC-SI [18]	0.9477	0.9594	6.9404	0.9097	0.9071	0.1192
MLV [13]	0.9619	0.9566	5.9437	0.9489	0.9247	0.0904
JNB [11]	0.8398	0.8373	11.8083	0.8339	0.7729	0.1555
CPBD [12]	0.9128	0.9429	8.8827	0.9155	0.8847	0.1153
Pro. method	**0.9771**	**0.9704**	**4.6255**	**0.9490**	**0.9315**	**0.0880**

the proposed method with several state-of-the-art blurriness-specified NR-IQA metrics as well as two LBP-based methods, including ARISM [2], BIBLE [17], FISHbb [15], LPC-SI [18], MLV [13], JNB [11], CPBD [12], LTP [7] and NR-LBPS^{riu2} [8]. To ensure the fairness of the comparison, all the results except LTP [7] and NR-LBPS^{riu2} [8] (their codes are not released by authors) are obtained by using demos released by authors. For LTP and NR-LBPS^{riu2}, we directly copy the experiment results from the associated papers. Table 2 tabulates the comparison results among the selected metrics on four public databases (the bold values indicate the best performance). In the table, "–" denotes empty. From the table, it can be intuitively observed that the proposed method always obtains the highest performance across all the selected metrics on all databases except CSIQ. The LTP [7] obtains the best performance on CSIQ, however, it is inferior to the proposed method on the other databases. BIBLE [17] is the second best method from comprehensive view. It produces better performance on LIVE II and CSIQ, while yields to general performance on TID2008 and TID2013. On the contrary, the proposed method achieves the best performance on most databases and improves the performance over BIBLE by 2.53% in PLCC on TID2013. Overall, the proposed method produces the highest scores and correlates with human perception well in terms of prediction accuracy and monotonicity.

In order to further confirm the superiority of the proposed method, we calculate the statistical significance between it and all the completing methods. For this purpose, the two sample t-test was employed to measure whether the mean values of two independent PLCC samples were equivalent or not. The test is conducted at 5% significance level. Table 3 lists the experiment results on four databases. For simple representation, "1" ("−1") indicates that the proposed method is superior (inferior) to the compared method, while "0" means that both methods are statistically equivalent. As expected, the proposed method significantly outperforms most competing methods on four databases. To be more specific, it is statistically equivalent to MLV, BIBLE, GMSD and ADD-GSIM once on CSIQ 2008 and CSIQ 2013 and superior to all competing methods on Live II and CSIQ.

3.3. Robustness analysis

To validate that the method is not heavily depended on the training set, we utilized the following analysis. The training set is varied from

Table 3
Results of two sample t-test performed between PLCC obtained by competing IQA methods on four databases.

Metric	Live II	CSIQ	TID 2008	TID 2013
ARISM [2]	1	1	1	1
NFERM [3]	1	1	1	1
BIBLE [17]	1	1	0	1
FISHbb [15]	1	1	1	1
LPC-SI [18]	1	1	1	1
MLV [13]	1	0	1	1
JNB [11]	1	1	1	1
CPBD [12]	1	1	1	1
SISBLIM [33]	1	1	1	1
GMSD [31]	1	1	0	1
NIQE [5]	1	1	1	1
IL-NIQE [6]	1	1	1	1
ADD-SSIM [32]	1	1	1	1
ADD-GSIM [32]	1	1	0	1

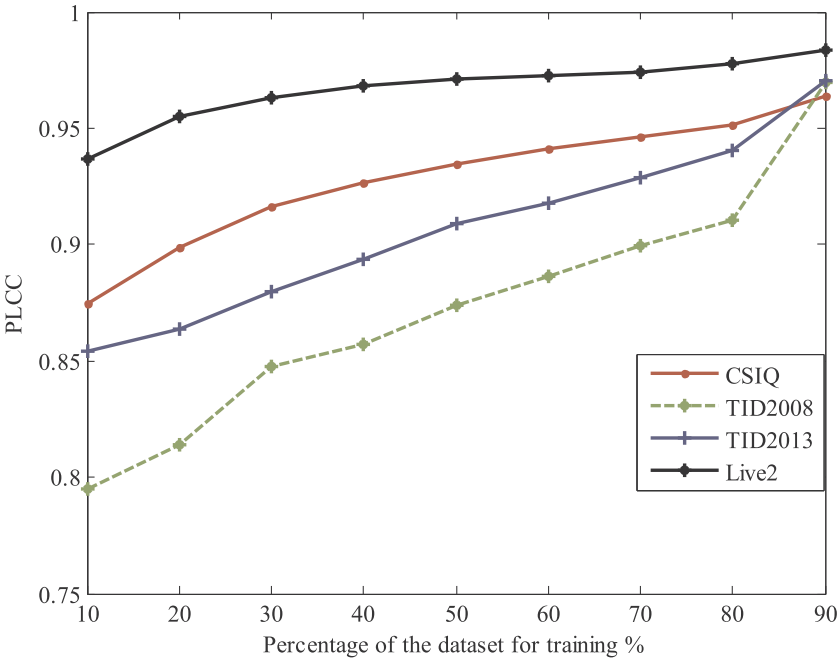

Fig. 5. Plot of median PLCC as a function of training's proportion.

90% of the content to only 10%, while the remaining dataset is used for testing. For each case, the dataset was randomly operated 1000 times and the values of three criterions were reported in form of median. Fig. 5 depicts the relationship between PLCC performance and training ratio. It can be intuitively observed that the PLCC value increases in all databases as training ratio increasing. Moreover, the performance is still high even though the training set is less than 30%. This shows that the proposed method is robust.

3.4. Parameters sensitivity

Because LBP possesses two parameters, i.e., radius R and sampling point K, it is meaningful to explore their impacts on the overall performance. In this section, we conduct an extra experiment to unveil the mystery. For this purpose, we directly utilize f_{LBP} to form the feature vectors for avoiding the interference of f_e. Generally speaking, the LBP descriptor usually utilizes 8 sampling points. Taking this as a breakthrough point, we first fix the sampling point as 8 and change the radius from 1 to 4. Second, we fix the radius as 1 and set the sampling point as {4, 8, 16}. We manually extract the histograms bins, which are consistent with the degree of blurriness, in cases of various radiuses and sampling points. Table 4 gives the experiment results performed on TID 2008. In the table, "$R1$-$2K8$" denotes the model by combining the extracted features when the radiuses are 1 and 2, sampling point is 8. From the table, we can observe that: (1) we only obtain general performance when utilizes single radius and sampling point (e.g., $R1K8$,$R2K8$,$R3K8$ and $R3K8$); (2) although $R1K16$ performs encouraging results, it requires more computational complexity than $R1K8$ and $R1K4$; (3) the performance increases when combining various radiuses together. However, there is no obvious difference among $R1$-$2K8$,$R1$-$3K8$ and $R1$-$4K8$. Therefore, considering the trad-off between computational complexity and performance, we reasonably choose $R1$-$2K8$ in this article.

Table 4
Results of the proposed method with different parameters on TID 2008 database.

Model	PLCC	SRCC	RMSE	Model	PLCC	SRCC	RMSE
R1K8	0.8446	0.8211	0.6123	R1-2K8	0.8923	0.8742	0.5244
R2K8	0.8728	0.8526	0.5602	R1-3K8	0.9051	0.8925	0.4928
R3K8	0.8765	0.8574	0.5585	R1-4K8	0.8948	0.8797	0.5206
R4K8	0.8159	0.7986	0.6675	R1K4	0.8814	0.8722	0.5476
R1K16	0.9007	0.8897	0.5013				

3.5. Performance on JPEG2000 and unsharp masking distortions

In this part, we first make an extra experiment to validate the proposed method's performance on JPEG2000 distortion. It is observed from Fig. 2 that, part of histogram bins change regularly with the increase of blurriness. Therefore, the proposed method might also work well on other distortions that cause blurriness in an image. To reduce the amount of data, JPEG2000 concentrates most of important visual information on a small number of wavelet coefficients, and the remaining coefficients are roughly quantified or directly converted to zero. To some certain, JPEG2000 leads to loss of high frequency components, leading to the generation of blurriness. As depicted in Fig. 6, the edge of image spreads out when suffered from serious JPEG2000 compression. In view of this, we explored the effectiveness of the proposed method on four public databases. Experimental results are tabulated in Table 5 in form of median. As shown by the data in Table 5 that the proposed method also performs notable performance with respect to JPEG2000 distortion evaluation.

It is widely known that the visual appearance of an image may be significantly improved by emphasizing its high frequency contents. As high frequency contents are emphasized, the edge and detail information is enhanced. However, excessive manipulation can produce side effects. On the one hand, the perceptual quality of an image might be declined by excessively increasing high frequency. On the other hand, it may also neglect the low frequency contents. Therefore, in this paragraph, we make an attempt to examine whether the proposed method suits for the evaluation of image sharpness. To solve this problem, we choose three representative images from TID2008 and process them via linear unsharp masking algorithm. The linear unsharp masking algorithm was employed to generate the enhanced image I':

$$I' = I + \lambda \cdot |I - I_f| \tag{9}$$

where I is the original image; I_f is the output of low pass filter; "||" is the absolute value operator; λ is the positive scaling factor that controls the level of high frequency enhancement. In this article, the image was filtered by Gaussian filter with template 5×5; the λ ranges from 0.1 to

(a) Reference image (b) JPEG 2000

Fig. 6. The image "woman" and its associated distorted images: (a) Reference image, and (b) JPEG 2000.

Table 5
Experimental results on JPEG2000 distortion.

Database	PLCC	SRCC	RMSE
LIVE II	0.9609	0.9521	6.7435
CSIQ	0.9157	0.8892	0.1270
TID2008	0.9483	0.9180	0.6019
TID2013	0.9461	0.9028	0.5450

2 in steps of 0.3. To the end, 21 images are obtained in total. 15 observers were recruited to give the opinion scores of the generated images. Then, the MOS were obtained by averaging all the subjective scores of each image. Finally, we extract the image feature vectors and feed them into the trained SVR module on TID 2008. In Fig. 7, there are three lines of words. The first line is the positive scaling information, while the second and third lines are the MOS and predicted scores via the proposed method. As expected, the proposed method obtains encouraging performance with PLCC of 0.71.

3.6. Performance comparison on 3D databases

With the rapidly advances in stereoscopic display, 3D multimedia comes into people's life due to its incomparable advantage on reappearing vivid scene. Recently, many movies, TV programs and computer games are produced using 3D technologies. Indeed, 3D images are ineluctable to meet distortions like 2D images. Hence, it is significant to design a NR IQA method that is competent for both 2D and 3D distortion assessment. In this part, we examine the performance of the proposed method on 3D IQA problem. Firstly, both left and right image of 3D image pair are analyzed separately to extract LBP features (using the same procedure as 2D image). Then the overall feature vector is composed of LBP features extracted from both left and right images. Finally, a model is trained using SVR to map the feature space to quality space. Table 6 tabulates the comparison results among the proposed method and several state-of-the-art 3D IQA methods. Since the source codes are not released by authors, we extract the experimental results directly from the corresponding papers. The best results are highlighted in bold. We can make the following observations from the presented results: (1) the proposed method delivers very promising on both 3D IQA databases. It achieves the best performance on PLCC and RMSE across all databases; (2) compared to other metrics, the proposed method produces very high performance on both databases. In particular, the proposed method is suitable for both symmetric and asymmetric distortion types, while the performance of some metrics declines fast when it meets asymmetric distortion (e.g., Chen [34] and STRIQE [35] achieves promising results on symmetric Jp2k distortion, but produces general results on asymmetric Jp2k distortion). Through these observations, we have sufficient reason to prove that our algorithm can be well qualified for both the 2D and 3D image blurriness evaluation tasks.

3.7. Runtime analysis

To evaluate the runtime, a comparison was conducted between the proposed method and the selected methods on estimating the images of

Fig. 7. The image "woman" and its associated unsharp masking images: from (a) to (b) with different λ.

Table 6
Performance comparisons with state-of-the-art 3D IQA metrics.

Database	Metric	Type	Blur			Jp2k		
			PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
LIVE3D Phase I	Shao [23]	NR	0.9032	**0.8989**	–	0.9275	**0.9088**	–
	Chen [30]	NR	0.917	0.878	5.898	0.907	0.863	5.402
	Chen [34]	FR	0.942	0.925	4.822	0.912	0.888	5.320
	STRIQE [35]	FR	0.9630	0.8428	–	0.9353	0.9066	–
	Pro. method	NR	**0.9657**	0.8571	**3.5128**	**0.9485**	0.8824	**3.9908**
LIVE3D Phase II	Shao [23]	NR	0.8904	0.8817	–	0.9013	0.8833	–
	Chen [30]	NR	0.941	0.900	4.725	0.899	0.867	4.298
	Chen [34]	FR	0.963	0.908	3.747	0.834	0.814	5.562
	STRIQE [35]	FR	0.9757	**0.9305**	–	0.8765	0.8732	–
	Pro. method	NR	**0.9840**	0.9050	**2.2586**	**0.9230**	**0.8882**	**3.4406**

Table 7
Comparison of mean computational time (in seconds).

Metrics	512 × 384	768 × 512	1920 × 1080	3840 × 2160
NFERM [3]	30.30	59.93	313.87	1271.43
ARISM [2]	11.16	18.86	103.32	334.61
IL-NIQE [6]	5.00	5.20	5.21	5.46
SISBLIM [33]	1.53	2.72	14.12	56.22
BIBLE [17]	1.35	2.57	13.57	52.87
FISHbb [15]	0.63	0.95	4.69	18.59
LPC-SI [18]	0.57	1.57	6.26	26.61
JNB [11]	0.81	1.16	8.03	34.72
CPBD [12]	0.29	0.53	3.36	22.71
NIQE [5]	0.19	0.35	1.64	6.71
MLV [13]	0.07	0.11	0.54	2.02
Pro. method	0.19	0.37	1.13	4.19

size 512 × 384, 768 × 512, 1920 × 1080 and 3840 × 2160 pixels. Table 7 provides details of comparison results (in seconds). All algorithms are implemented in MATLAB 2013b and executed on a 2.6 GHz processor with 8 GB RAM, Windows 7 Pro 64-bit laptop.

As shown in Table 7, the computational time increases as with the increment of image size. Even though the image size is 3840 × 2160 pixel, the run time is still short using the proposed method. Overall, the proposed method is slightly slower than MLV but superior to other methods to a great extent. This phenomenon might be attributed to the following reasons: (1) compared to other metrics, the proposed method is computed directly and doesn't require feature space transformation; (2) the proposed method only calculates the LBP features, which needs less computational complexity.

4. Conclusion

In this paper, an effective and efficient NR blurriness assessment metric is proposed based on LBP features. Through comprehensive statistic, we reveal that the entropy of LBP histogram bins and part of LBP histogram bins change monotonous to the degree of blurriness. Therefore, the special bins and entropy are chosen as the feature vectors. The predicted quality score is obtained by using a pre-trained regression module. Results obtained through comprehensive experiments prove that the proposed method outperforms the state-of-the-art FR-QA and NR-IQA metrics on LIVE II, CSIQ, TID2008 and TID2013 databases in terms of high consistency with human perception and low computational complexity. Besides, the proposed method can also be competent for 3D image blurriness assessment problem without alterations. Experimental results show that the proposed method do well in 3D blurriness assessment under both symmetric and asymmetric blurred distortions.

Acknowledgment

Support for this program is provided by the National Natural Science Foundation of China under Grants 61471262 and 61731003, the International (Regional) Cooperation and Exchange under Grant 61520106002, and the Doctoral Fund of Ministry of Education of China under Grant 20130032110010.

References

[1] Weisi Lin, C.C. Jay Kuo, Perceptual visual quality metrics: a survey, J. Visual Commun. Image Rep. 22 (4) (2011) 297–312.
[2] Ke Gu, Guangtao Zhai, Weisi Lin, Xiaokang Yang, Wenjun Zhang, No-reference image sharpness assessment in autoregressive parameter space, IEEE Trans. Image Process. 24 (10) (2015) 3218–3231.
[3] Ke Gu, Guangtao Zhai, Xiaokang Yang, Wenjun Zhang, Using free energy principle for blind image quality assessment, IEEE Trans. Multimedia 17 (1) (2015) 50–63.
[4] K. Gu, G. Zhai, X. Yang, W. Zhang, C.W. Chen, Automatic contrast enhancement technology with saliency preservation, IEEE Trans. Circ. Syst. Video Technol. 25 (9) (2015) 1480–1494.
[5] A. Mittal, R. Soundararajan, A.C. Bovik, Making a "completely blind" image quality analyzer, IEEE Signal Process. Lett. 20 (3) (2013) 209–212.
[6] Lin Zhang, Lei Zhang, Alan C. Bovik, A feature-enriched completely blind image quality evaluator, IEEE Trans. Image Process. 24 (8) (2015) 2579–2591.
[7] P.G. Freitas, W.Y.L. Akamine, M.C.Q. Farias, No-reference image quality assessment based on statistics of local ternary pattern, in: 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), June 2016, pp. 1–6.
[8] Min Zhang, Jin Xie, Xiangrong Zhou, Hiroshi Fujita, No reference image quality assessment based on local binary pattern statistics, Visual Communications and Image Processing (VCIP), 2013, IEEE, 2013, pp. 1–6.
[9] K. Gu, D. Tao, J.F. Qiao, W. Lin, Learning a no-reference quality assessment model of enhanced images with big data, IEEE Trans. Neural Networks Learn. Syst. (2017), http://dx.doi.org/10.1109/TNNLS.2017.2649101.
[10] Huan Yang, Yuming Fang, Weisi Lin, Perceptual quality assessment of screen content images, IEEE Trans. Image Process. 24 (11) (2015) 4408–4421.
[11] R. Ferzli, L.J. Karam, A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB), IEEE Trans. Image Process. 18 (4) (2009) 717–728.
[12] N.D. Narvekar, L.J. Karam, A no-reference image blur metric based on the cumulative probability of blur detection (CPBD), IEEE Trans. Image Process. 20 (9) (2011) 2678–2683.
[13] Khosro Bahrami, Alex C. Kot, A fast approach for no-reference image sharpness assessment based on maximum local variation, IEEE Signal Process. Lett. 21 (6) (2014) 751–755.
[14] Christoph Feichtenhofer, H. Fassold, P. Schallauer, A perceptual image sharpness metric based on local edge gradient analysis, IEEE Signal Process. Lett. 20 (4) (2013) 379–382.
[15] P.V. Vu, D.M. Chandler, A fast wavelet-based algorithm for global and local image sharpness estimation, IEEE Signal Process. Lett. 19 (7) (2012) 423–426.
[16] Chong Yaw Wee, Raveendran Paramesran, R. Mukundan, Xudong Jiang, Image quality assessment by discrete orthogonal moments, Pattern Recogn. 43 (12) (2010) 4055–4068.
[17] L. Li, W. Lin, X. Wang, G. Yang, K. Bahrami, A.C. Kot, No-reference image blur assessment based on discrete orthogonal moments, IEEE Trans. Cybernet. 46 (1) (2016) 39–50.
[18] R. Hassen, Z. Wang, M.M. Salama, Image sharpness assessment based on local phase coherence, IEEE Trans. Image Process. 22 (7) (2013) 2798–2810.
[19] Timo Ojala, Inendavid Harwood, A comparative study of texture measures with classification based on feature distributions, Pattern Recogn. 29 (1) (1996) 51–59.
[20] B. Zhang, Y. Gao, S. Zhao, J. Liu, Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor, IEEE Trans. Image

Process. 19 (2) (2010) 533–544.
[21] I. Choi, D. Kim, Local transform features and hybridization for accurate face and human detection, IEEE Trans. Software Eng. 35 (6) (2013) 1423–1436.
[22] Timo Ojala, Matti Pietikainen, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987.
[23] F. Shao, K. Li, W. Lin, G. Jiang, Q. Dai, Learning blind quality evaluator for stereoscopic images using joint sparse representation, IEEE Trans. Multimedia 18 (10) (2016) 2104–2114.
[24] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, F. Battisti, Tid2008 – a database for evaluation of full-reference visual quality assessment metrics, Adv. Modern Radioelectron. 10 (2009) 30–45.
[25] ChihChung Chang, ChihJen Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (3) (2011) 389–396.
[26] Eric Cooper Larson, Damon Michael Chandler, Most apparent distortion: full-reference image quality assessment and the role of strategy, J. Electron. Imaging 19 (1) (2010) 143–153.
[27] H.R. Sheikh, M.F. Sabir, A.C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, IEEE Trans. Image Process. 15 (11) (2006) 3440–3451.
[28] Nikolay Ponomarenko, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Lukui Jin, Jaakko Astola, Benoit Vozel, Kacem Chehdi, M. Carli, F. Battisti, Color image database tid2013: peculiarities and preliminary results, in: European Workshop on Visual Information Processing, 2013, pp. 106–111.
[29] Anush Krishna Moorthy, Che Chun Su, Anish Mittal, Alan Conrad Bovik, Subjective evaluation of stereoscopic image quality, Signal Process. Image Commun. 28 (8) (2013) 870–883.
[30] M.J. Chen, L.K. Cormack, A.C. Bovik, No-reference quality assessment of natural stereopairs, IEEE Trans. Image Process. 22 (9) (2013) 3379–3391.
[31] Wufeng Xue, Lei Zhang, Xuanqin Mou, Alan C. Bovik, Gradient magnitude similarity deviation: a highly efficient perceptual image quality index, IEEE Trans. Image Process. 23 (2) (2014) 684–695.
[32] K. Gu, S. Wang, G. Zhai, W. Lin, Analysis of distortion distribution for pooling in image quality prediction, IEEE Trans. Broadcast. 62 (2) (2016) 446–456.
[33] Ke Gu, Guangtao Zhai, Xiaokang Yang, Wenjun Zhang, Hybrid no-reference quality metric for singly and multiply distorted images, IEEE Trans. Broadcast. 60 (3) (2014) 555–567.
[34] Ming Jun Chen, Che Chun Su, Do Kyoung Kwon, Lawrence K. Cormack, Alan C. Bovik, Full-reference quality assessment of stereopairs accounting for rivalry, Signal Process. Image Commun. 28 (9) (2013) 1143–1155.
[35] Sameeulla Khan Md, Balasubramanyam Appina, Sumohana S. Channappayya, Full-reference stereo image quality assessment using natural stereo scene statistics, IEEE Signal Process. Lett. 22 (11) (2015) 1985–1989.