REGULAR PAPER

A simple spatial domain method for quality evaluation of blurred images

Md Amir Baig[1] · Athar A. Moinuddin[2] · E. Khan[2]

Abstract

In this paper, a simple yet highly accurate algorithm for No-Reference quality estimation of blurred images is proposed. The proposed work is motivated by the fact that when blurring occurs, regions with large pixel variations are likely to be affected more than the regions with small pixel variations. Since the human visual system is also more attentive to distortions in the regions with large pixel variations; therefore, it will be advantageous to exploit them for blurriness estimation. Moreover, blurring also causes loss of details, thereby resulting in the decrease in mean pixel variation and the maximum pixel variation. It is also observed that the ratio of mean to maximum pixel variation increases with blurriness. Motivated by these facts, the maximum pixel variation and mean pixel variation are utilized to estimate the quality of the image affected by the blurriness. The proposed algorithm is highly competitive and outperforms most of the state-of-the-art algorithms both in time-complexity as well as accuracy over various standard databases.

1 Introduction

A number of factors, including low light photography, out-of-focus image capturing, processing, and compression, among others, can degrade images. Degraded photos may be evaluated for quality using subjective or objective approaches. However, objective methods are employed for automatic evaluation. It can be classified as full-reference (FR), reduced-reference (RR), or no-reference (NR) depending on whether the reference image is available. While only partial information about the reference images is given in the RR approach, it is presumed that the reference image is available in the FR method. However, in many instances, such as when a defocused camera is used to take an image, generating blurriness, or when low light is present, causing additive white noise (AWGN) distortion, neither the reference image itself nor any of its information is available for comparison. Consequently, the natural solution is the NR approach [1].

The methods for evaluating quality can also be divided into general-purpose and distortion-specific categories. The quality of an image that has been altered by any particular distortion, including blockiness, blurriness, noise, and quantization, can be predicted using general-purpose algorithms. Even though general-purpose methods are preferred, not all distortion types are assessed with high accuracy. Also, due to their versatile nature, the computational complexity of general-purpose methods is high. Contrarily, distortion-specific algorithms are optimized for one or more chosen known distortion types, resulting in highly accurate and simple algorithms [2]. Therefore, distortion-specific approaches are typically favored if the kind and source of the image's

✉ Md Amir Baig
amir.jvwu@gmail.com

Athar A. Moinuddin
atharamoin@gmail.com

E. Khan
ekhan67@gmail.com

1 University Women's Polytechnic, Zakir Husain College of Engineering & Technology, Aligarh Muslim University, Aligarh, Uttar Pradesh 202002, India

2 Department of Electronics Engineering, Zakir Husain College of Engineering & Technology, Aligarh Muslim University, Aligarh, Uttar Pradesh 202002, India

distortion are known. The possible cases of specific distortion of blurriness are discussed as follows. Gaussian blurring is one of the common and dominant types of distortion perceived in images when captured by low end cameras like the ones used in mobile phones. The blurriness is also introduced when unwanted details and the noise are removed from the image by smoothing. Sometimes, to meet the bandwidth or storage constraints, the high degree of compression is used despite the fact that some degree of blurriness may develop. These are the situations when the distortion type in the images is known. Therefore, suitable and efficient blurriness specific quality evaluation method is required to be developed. A number of blur specific NR methods to measure the quality of blurred images have been developed in the past and it is still an active research area. Some of the blur specific algorithms [3–17] developed in the last decade are briefly reviewed as follows.

In [3], first, the sharpness has been measured in both spectral and spatial domain at block level, and then, their corresponding geometric means are computed. Top 1% of the means are selected to formulate the metric. As, this method uses spectral domain as well as spatial domain, therefore, results into higher computational complexity. Furthermore, the accuracy is not consistent across the different databases. A metric named as FISH/FISH$_{bb}$ was proposed in [4]. This method uses Log-energies of each discrete wavelet transform (DWT) subbands to compute block level sharpness and then pooled to get quality of entire image. However, this method is not accurate over unseen images as it relies on empirical weight assignment by experimenting on images of seen databases. A complex wavelet transform-based method is developed in [5]. In this method, phases of complex wavelet coefficients are exploited to measure sharpness of an image.

To quantify the blurriness, Generalized Gaussian distribution (GGD) and asymmetric GGD (AGGD) parameters are computed over the estimated maximum local variation (MLV) in 3×3 neighborhood [6]. By measuring these two parameters, MLVs are weighted throughout the image with higher weights are assigned to higher MLVs values. The sharpness metric is then formulated by computing their standard deviation. In [7, 8], reblurring of images has been used to develop the quality metric. In [8], a step-by-step approach is employed to determine the minimum standard deviation needed to generate valid reblur images. Using the difference between test and reblur images, a blur metric is formulated. Whereas in [7], the metric is developed by the information of shape of local histogram by taking into the account of difference between reblur and test images. The contraction of histogram distribution of the blurred images with respect to natural images has been exploited to formulate the metric in [9]. A learning-based sharpness metric which used Zernike moments and gradient magnitude (GM) in the framework was proposed in [12]. Zernike moments

estimate irregularities and distortions caused by blurring, whereas GM quantify the loss of sharpness and fine details. Dictionary learning-based sharpness metric was proposed in [13]. As the atoms are usually edge patterns in the dictionaries and image blur is defined by edge spread, an over-complete dictionary was used to calculate the image sharpness. Its accuracy is high but computational complexity is high.

Free energy principle and auto-regressive based models were proposed in [11, 14]. In [11], free energy principle and the NFEQM model [18] were combined to develop a sharpness metric. It is high in accuracy, though computationally intensive. In [14], noise energy was computed at block level (known in the paper as 'stem noise'), to quantify image blurriness. The characteristic of estimated energy reflects the blurriness of the image. This method found to be moderately accurate across different databases. One of the recent sharpness assessment metrics based on local contrast map (spatial domain) and DWT (transform domain) is proposed in [15]. This method is also not suitable for all the databases. In [16], a discrete Fourier Transform (DFT)-based method was proposed. In this metric development, the average of the ratio of the magnitude of AC coefficients before and after the addition of constant was computed over entire DFT coefficients. It was first computed over local level and then pooled to get overall quality score. This method was fast but moderate in accuracy. In [17], the difference between sharpest and blurriest spot of an image is exploited to measure the sharpness. It is fast and highly accurate.

Nowadays, due to the availability of computational power and large sets of labeled training images, a lot of researches are going in the field of image quality assessment which utilizes Deep Learning methods. Some of the methods that may be useful to assess the quality of blurred images can be found in [10, 19–24].

Though the accuracy of blur specific algorithms had improved in recent past and many highly accurate algorithms were developed. Most of these are either based on Machine Learning/Deep learning or parameter tuning while developing the model. However, due to the training methods, the trained model may not always be guaranteed for accurate quality estimation of images of unseen/future
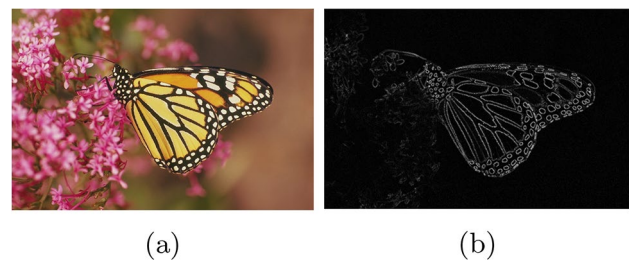

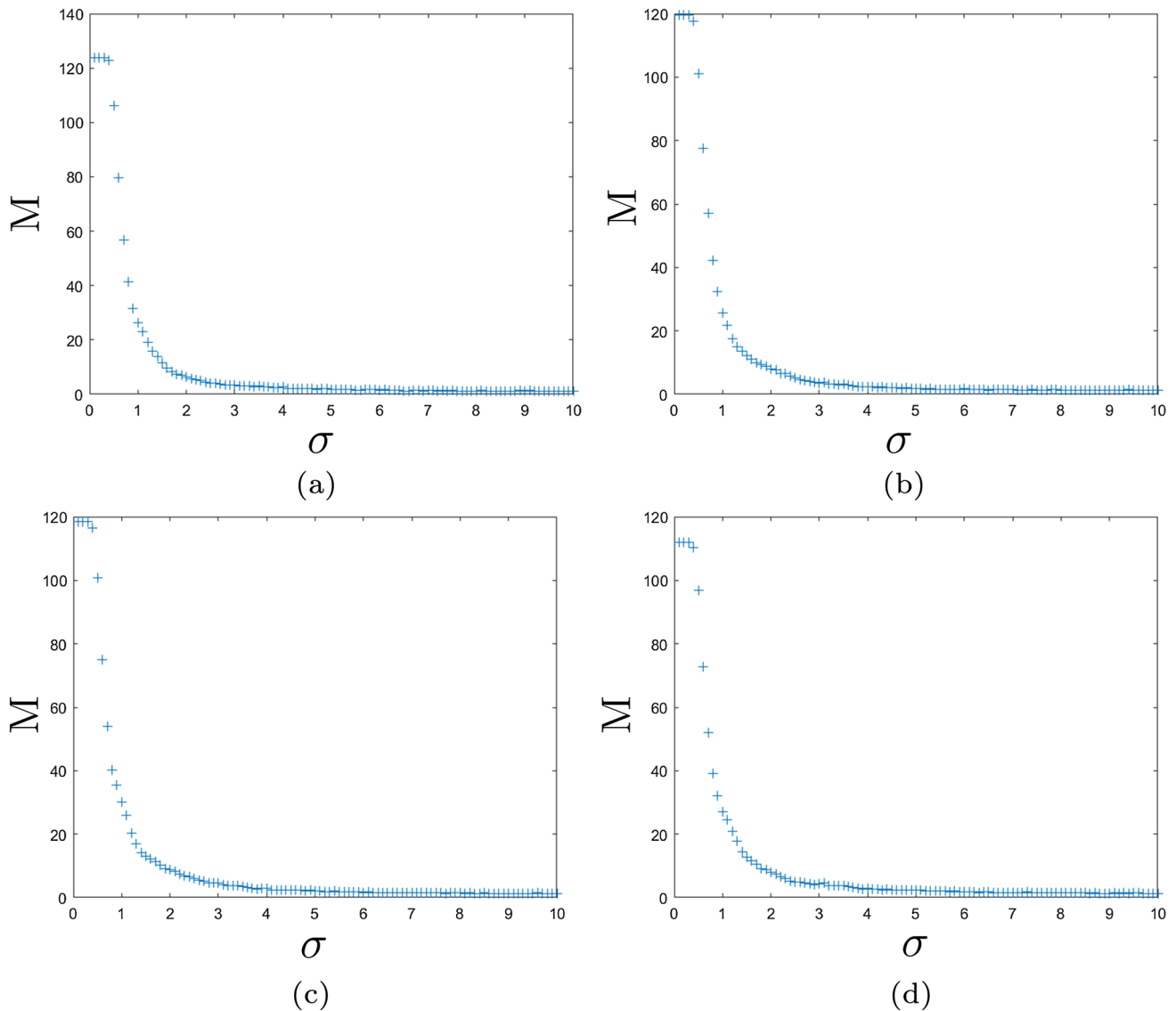**Fig. 1** (a) Monarch image. (b) The pixel variations map

Fig. 2 Plot of maximum pixel variation (M) vs. blurriness level (σ)

databases. In this paper, a highly accurate method independent of Machine/Deep Learning algorithms and databases is proposed. Moreover, its time complexity is lower than the existing state-of-the-art algorithms. Fast blur image quality assessment offers quick and efficient evaluation of image quality, making it suitable for real-time applications, large datasets, and resource-constrained environments. It provides prompt feedback and scalability, and maintains a balance between computational efficiency and accuracy. The proposed work is motivated by the fact that when blurring occurs, regions with different pixel variations are affected differently. As human visual system (HVS) is also more attentive to distortions in the regions with large pixel variations; therefore, it is meritorious to exploit them for blurriness estimation. Moreover, blurring also causes loss of overall details, thereby resulting in the decrease in mean pixel variation and the maximum pixel variation. The ratio of mean to maximum pixel variation is observed as an indicator of the degree of blurriness. Motivated by these facts, the maximum pixel variation and mean pixel variation are utilized to estimate the quality of the image affected by the blurriness.

The rest of this paper is organized as follows. It starts with the detailed discussion of the proposed methodology in Sect. 2. It is then followed by performance validation and comparison with the existing methods for various databases in Sect. 3. Finally, the paper is concluded in Sect. 4.
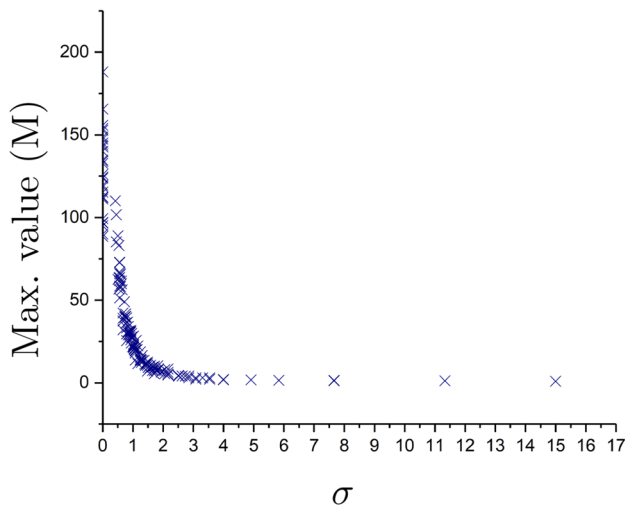
Fig. 3 Scatterplot of maximum pixel variation (*M*) vs. blurriness level (σ)

2 Proposed methodology

As discussed earlier, variation in pixel values is an indication of image sharpness/blurriness. The pixel variations of an image can be evaluated by a high-pass Laplacian filter coefficients [2, 25] as given by H in Eq. 1

$$H = \frac{1}{4}\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \tag{1}$$

For example, to observe the pixel variations of image Monarch (Fig. 1a), it is convoluted with the high-pass Laplacian filter H. The resulting filtered image representing its pixel variations map is shown in Fig. 1b.

On comparing the two images, it is observed that small pixel variations (smooth regions) are reflected by darker pixels, whereas large pixel variations (sharp regions) are reflected by brighter pixels.

It is well known that an image with Gaussian blur distortion can be mathematically represented by convolution of the image with a Gaussian low-pass filter. Let the blurred image be denoted by $I(i, j)$. As the value of standard deviation (σ) of the filter increases, blurriness in the image also increases. As HVS is most attentive to the region where pixel variation is maximum, therefore, it is important to see how the blurriness affects the maximum of absolute pixel variation(M).

To measure the value of *M*, first pixel-wise blurred image $I(i, j)$ is convolved($\circledast$) with H to obtain the high-pass filtered output $I(i, j)'$ as given in Eq. 2

$$I'(i, j) = (I \circledast H)(i, j) \tag{2}$$

M is then obtained by finding the maximum absolute value in $I(i, j)$' as given in Eq. 3

$$M = max[|I(i, j)'|] \tag{3}$$

The plot of *M* as a function of blurriness level (σ) for four different natural images (from LIVE database) is shown in Fig. 2. It is clear that maximum pixel variation M decreases with the increase in blurriness level (σ). For further validation, scatter plot between the maximum pixel variation M and σ is plotted for the images of LIVE [26] database and is shown in Fig. 3. From the scatter plot, it is also clear that maximum pixel variation M is large for less blurred images (small σ) and small for more blurred images (large σ). On computing the correlation (SROCC) between M and σ, it is found to be high (i.e., SROCC = 0.992). Moreover, M and σ are inversely related, i.e., maximum pixel variation decreases with increase in standard deviation/blurriness. In NR quality evaluation, σ is usually not available; therefore, maximum pixel variation M can be exploited as a feature to measure the blurriness of the image and is given by Eq. 4

$$Q_1 = \frac{1}{M} \tag{4}$$

When the maximum value is high, indicating sharp features, the inverse (Q_1) will be low. Conversely, when the maximum value is low (indicating blurriness), the (Q_1) will be high. This feature essentially emphasizes the importance of having sharp elements in the image for higher quality scores. Further, blurring also causes loss of details thereby resulting in the decrease in the mean of absolute pixel variation $\bar{m}$ (Fig. 4) and the maximum pixel variation M (Fig. 2). On comparing Figs. 2 and 4, it is seen that the value of M decreases more rapidly in comparison to $\bar{m}$ as the blurriness level increases. It is observed that the ratio of mean pixel variation to maximum pixel variation, $\bar{m}/M$, increases with increase in blurriness as evident from Fig. 5. When this ratio is high, it indicates that, on average, the edges and transitions in the image are relatively intense compared to the sharpest features. Conversely, a low ratio suggests that there are pronounced sharp features in the image compared to the overall average. This metric effectively distinguishes between images with uniform sharpness and those with localized areas of high sharpness. Thereby, $\bar{m}/M$ may also be used as an indicator of level of blurriness. Therefore, the other quality indicator is given by Eq. 5

$$Q_2 = \frac{\bar{m}}{M} \tag{5}$$

Figure 6 shows an image with different level of blurriness and the corresponding quality indicator scores Q_1 and Q_2. From these values, it is clear that they are the reflector of blurriness in the image. The values of Q_1 and Q_2 are merged
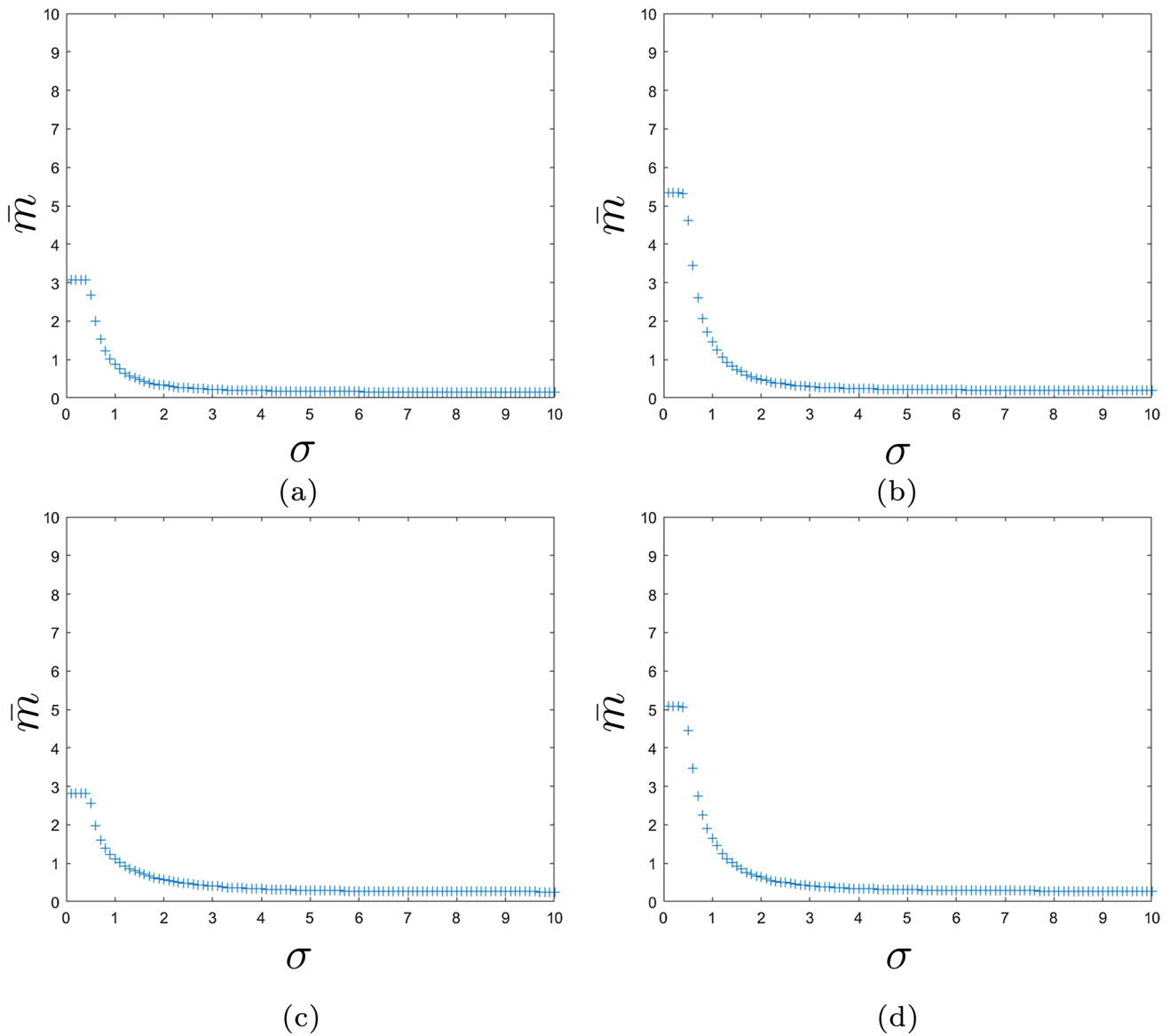
Fig. 4 Plot of mean pixel variation ($\bar{m}$) vs. blurriness level (σ)

by weighted geometric mean to get unified quality score, Q, as given in Eq. 6

$$Q = Q_1^{\gamma} \times Q_2^{1-\gamma} \tag{6}$$

The value of γ is determined empirically, and from extensive simulation, the value of $\gamma = 0.5$ is found to give the best result. This value of γ signifies that both the values of Q_1 and Q_2 are equal in weightage. The effect of different values of γ over the quality score is discussed in Sect. 3.1 of the next section. The process of computing quality metric Q is shown in the block diagram given in Fig. 7. To measure the pixel variations, first, an image is high-pass-filtered by H followed by computing the maximum pixel variation (M) and mean

($\bar{m}$) pixel variation. Then, using Eqs. 4 and 5, Q_1 and Q_2 are computed. Finally, the weighted geometric mean of Q_1 and Q_2 are used to compute overall quality score Q.

3 Results and discussion

In this section, the effectiveness of the proposed algorithms is thoroughly assessed and compared with that of other algorithms.
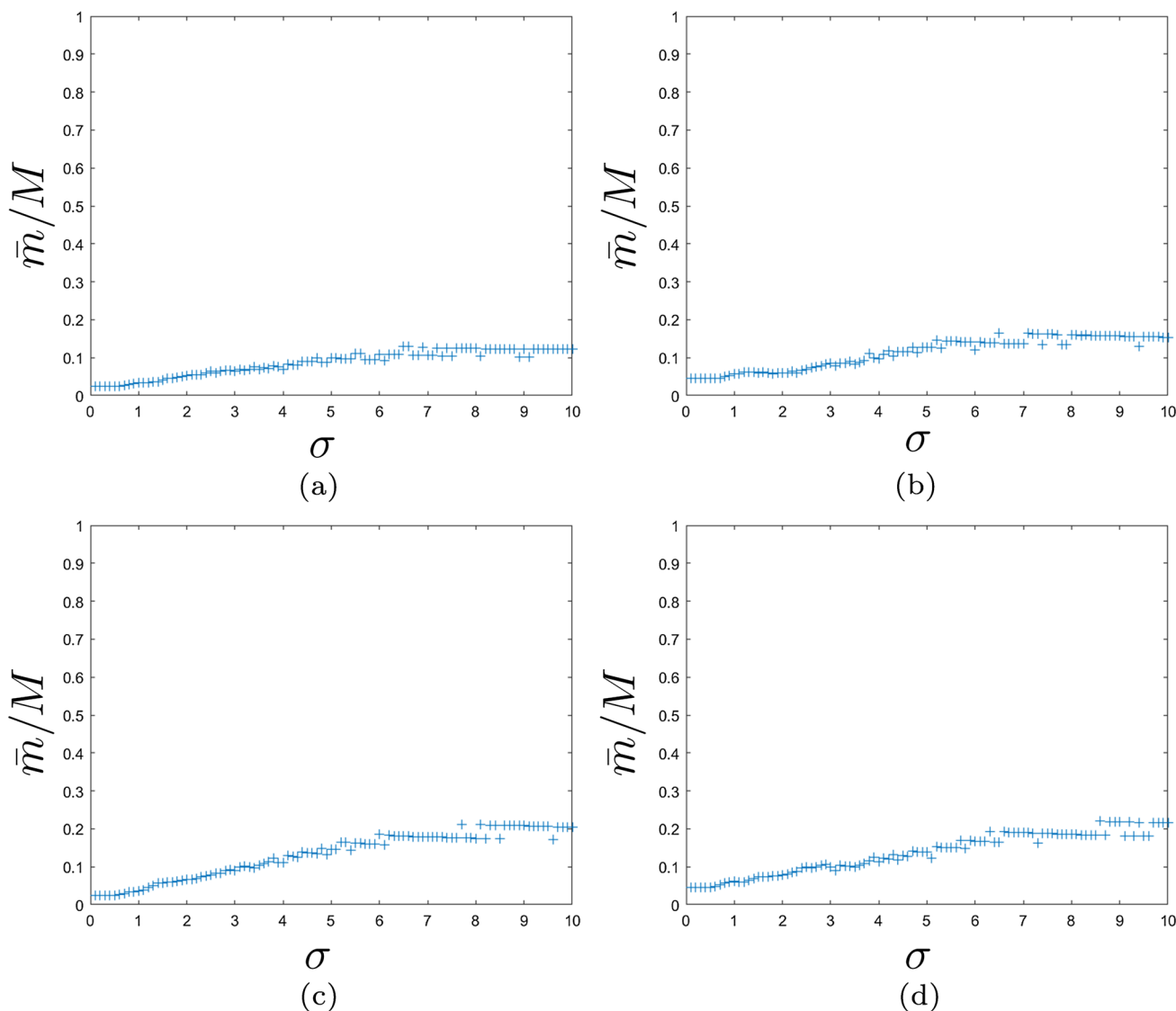
Fig. 5 Plot of ratio of mean pixel variation to maximum pixel variation ($\bar{m}/M$) vs. blurriness level (σ)

3.1 Experimental settings

The performance of the proposed method is evaluated on the standard databases, namely LIVE [26], VCL [27], TID2008 [28], TID2013 [29], and CSIQ [30]. While mean opinion score (MOS) of the blurred images is provided in TID2008, TID2013, and VCL databases, blurred images with differential mean opinion score (DMOS) scores are provided in LIVE and CSIQ. The accuracy is measured by computing Spearman's-rank-order-correlation-coefficient (SROCC), Pearson's-linear-correlation-coefficient (PLCC), and root-mean-square-error (RMSE) using the predicted subjective scores and the true subjective scores (MOS/DMOS) [15]. PLCC and RMSE measure the linearity and error, respectively, between the predicted and true scores.

The magnitudes of PLCC and SROCC are in the range [0, 1], where the greatest value of 1 denotes a perfect linear and monotonic relationship, respectively, and the minimum value of 0 denotes a complete lack of correlation. When the subjective scores are predicted perfectly, the RMSE value is 0. It increases with increase in difference of predicted and true subjective scores. Also, the range of subjective scores across different databases varies; therefore, range of RMSE values differ accordingly.

To predict the perceived subjective quality scores (MOS/DMOS), the quality metric Q is mapped to MOS/DMOS for entire databases using the logistic fitting function [15, 31] given in Eq. 7
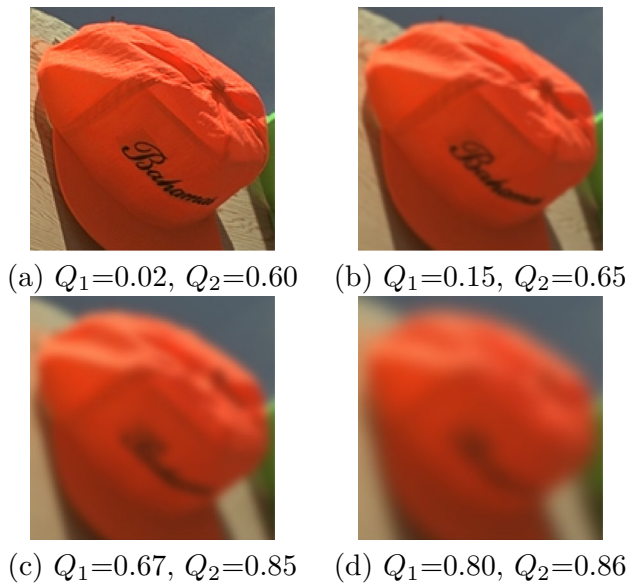
(a) $Q_1=0.02$, $Q_2=0.60$ (b) $Q_1=0.15$, $Q_2=0.65$

(c) $Q_1=0.67$, $Q_2=0.85$ (d) $Q_1=0.80$, $Q_2=0.86$

Fig. 6 Quality score Q_1 and Q_2 of an image with different levels of blurriness

$$DMOS = \psi_1\left(\frac{1}{2} - \frac{1}{1 + e^{\psi_2(Q - \psi_3)}}\right) + \psi_4 Q + \psi_5 \qquad (7)$$

In Eq. 7, $\psi_1, \psi_2, \psi_3, \psi_4$, and ψ_5 are the model parameters. It is worth to note that the value of the model parameters varies with the databases as they differ in methodology of subjective rating. Furthermore, logistic fitting function with four model parameters [4, 17]is also used which may give slightly different PLCC and RMSE; however, the SROCC remains same. Moreover, to evaluate the performance of training free methods, computing the monotonicity (SROCC) between Q and subjective scores from the given databases is sufficient enough as the test images may come from unknown/future databases and independent of number of parameters used in logistic function [14]. However, whenever required, both PLCC and SROCC are included in the discussion.

Table 1 SROCC for different values of γ over different databases

γ	LIVE	VCL	TID2008	TID2013	CSIQ
0.0	0.658	0.766	0.739	0.771	0.759
0.1	0.790	0.845	0.834	0.865	0.833
0.2	0.874	0.887	0.885	0.910	0.883
0.3	0.924	0.910	0.903	0.9316	0.909
0.4	0.952	0.917	**0.910**	**0.940**	0.921
0.5	0.963	0.923	0.903	0.933	**0.927**
0.6	0.969	0.927	0.894	0.925	0.925
0.7	**0.970**	**0.928**	0.880	0.913	0.920
0.8	0.969	0.926	0.864	0.897	0.916
0.9	0.967	0.925	0.841	0.880	0.912
1.0	0.964	0.921	0.827	0.868	0.907

The best values are given in bold

Table 2 PLCC for different values of γ over different databases

| γ	LIVE	VCL	TID2008	TID2013	CSIQ
0.0	0.704	0.750	0.755	0.771	0.806
0.1	0.813	0.836	0.836	0.860	0.871
0.2	0.881	0.882	0.885	0.906	0.904
0.3	0.920	0.911	**0.905**	0.925	0.920
0.4	0.942	0.927	0.902	**0.928**	0.927
0.5	0.954	0.935	0.896	0.924	0.929
0.6	0.961	0.938	0.887	0.916	**0.928**
0.7	**0.963**	0.939	0.876	0.906	0.926
0.8	**0.963**	**0.940**	0.864	0.896	0.923
0.9	**0.963**	0.936	0.853	0.886	0.920
1.0	0.961	0.933	0.843	0.877	0.9171

The best values are given in bold

The proposed algorithm is implemented using MATLAB 2018a on a computer equipped with an Intel R Xeon 2.13 GHz CPU, 20 GB of RAM, and a 500 GB hard drive.

As discussed in the last section (see Eq. 6), the quality metric Q depends on empirically computed γ. The SROCC between Q and MOS/DMOS with changes in γ is computed for all the image databases as given in Table 1.

Fig. 7 Block diagram of the proposed method

Table 3 RMSE for different values of γ over different databases

| γ	LIVE	VCL	TID2008	TID2013	CSIQ
0.0	17.023	16.124	0.769	0.7912	0.170
0.1	13.950	14.186	0.645	0.637	0.141
0.2	11.360	11.467	0.547	0.529	0.122
0.3	9.394	10.032	**0.498**	0.680	0.117
0.4	8.024	9.129	0.508	**0.463**	0.108
0.5	7.157	8.633	0.520	0.477	**0.106**
0.6	6.674	8.420	0.683	0.501	0.107
0.7	6.463	8.520	0.566	0.528	0.108
0.8	**6.428**	**8.314**	0.590	0.555	0.110
0.9	6.500	8.596	0.654	0.579	0.112
1.0	6.632	8.760	0.632	0.601	0.114

The best values are given in bold
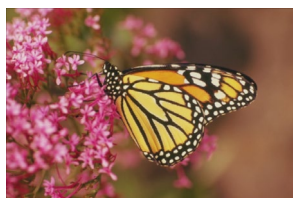
(a) $DMOS$=21.42
Q=0.04
$DMOS_P$=26.65
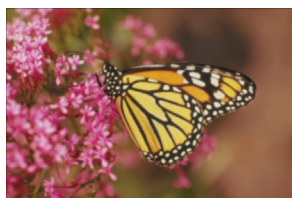
(b) $DMOS$=41.62
Q=0.08
$DMOS_P$=47.50
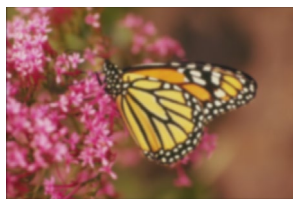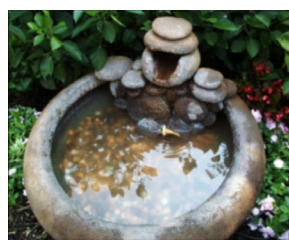
(c) $DMOS$=55.80
Q=0.15
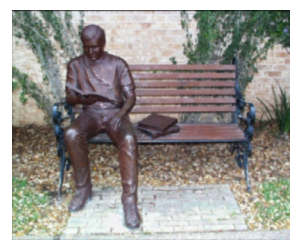$DMOS_P$=64.40
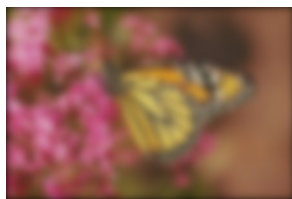
(d) $DMOS$=83.27
Q=0.40
$DMOS_P$=90.98

Fig. 8 Subjective score (DMOS) and quality metric (Q) for images with same content and different degree of blurriness

(a) $DMOS$=47.53
Q=0.08
$DMOS_P$=46.43

(b) $DMOS$=49.57
Q=0.09
$DMOS_P = 50.04$

(c) $DMOS$=45.11
Q=0.07
$DMOS_P$=44.20

(d) $DMOS$=46.51
Q=0.08
$DMOS_P$=48.42

Fig. 9 Subjective score (DMOS) and quality metric (Q) for images with different content and similar degree of blurriness
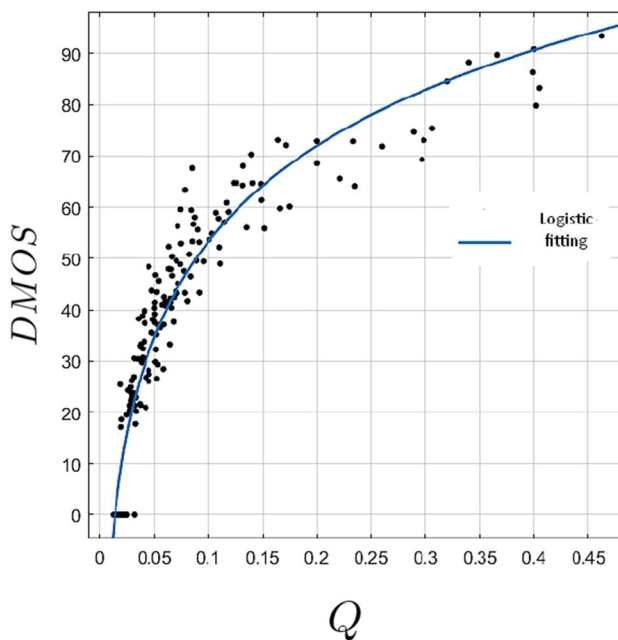
Fig. 10 Scatter plot of subjective score (DMOS) vs. quality metric (Q)

The best SROCC for each database is also highlighted in the table. It may be observed that the best SROCC varies for γ in the range of [0.4, 0.7]. By analyzing, it is evident that the value of γ=0.5 may be the best choice if SROCC across the databases is considered and, therefore, is used throughout the paper. Moreover, the PLCC and RMSE given in Table 2 and Table 3, respectively, also support the chosen value of γ.
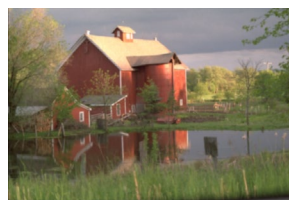
Moreover, Table 1 also gives the individual effect of Q_1 and Q_2 on the quality metric Q. When the value of γ=0, it implies that the value of metric depends on Q_2 only (as evident from Eq. 6). Similarly, for γ=1, the quality metric depends on Q_1 only. In both these cases, the SROCC is low. Therefore, it may be deduced that combining both Q_1 and Q_2 is worthwhile as it gives better accuracy.
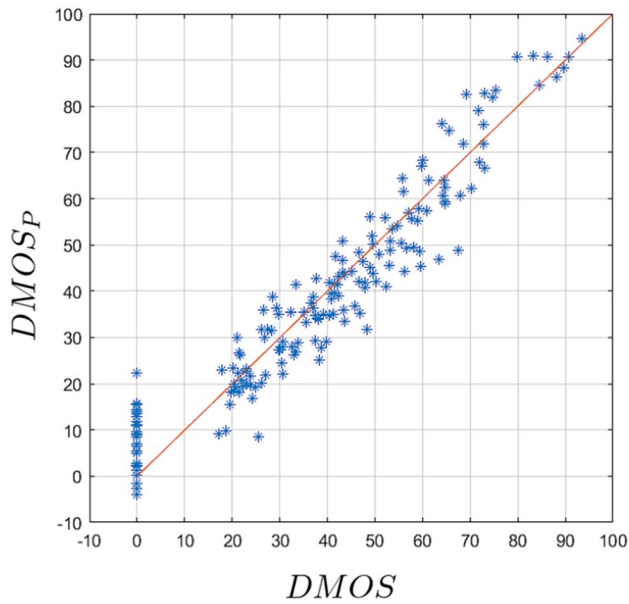
Fig. 11 Scatter plot of true and predicted *DMOS*

3.2 Accuracy over a set of images

First, the quality score Q is computed for an image with four different levels of blurriness, as shown in Fig. 8. It is evident that the objective quality score Q is consistent with the subjective DMOS score. A higher value of Q reflects higher value of DMOS, whereas a lower value of Q reflects lower value of DMOS. Furthermore, scores Q can clearly discriminate image with different levels of blurriness. Next, the quality scores Q is computed for four different images with similar level of blurriness, as shown in Fig. 9. Since the images are having similar level of blurriness, their DMOS

values are almost similar and so are the predicted quality scores Q. Here also the scores Q are consistent with the DMOS scores indicating the good prediction accuracy of the proposed algorithms.

For further validation of the proposed algorithm, the quality scores Q is computed for all the blurred images of LIVE database. Figure 10 shows the scatter plot with logistic fitting curve between the objective quality score Q and the subjective DMOS score. From the figure, it is observed that with increasing DMOS (blurriness in the image), the quality scores Q also increases and are highly correlated. Using the logistic fitting function of Eq. 7, the quality score Q (logarithmic value in base 10) is used to obtain the predicted DMOS ($DMOS_P$). The values of parameters obtained by logistic fitting are $\psi_1 = -44.31, \psi_2 = 6.171, \psi_3 = 13.66, \psi_4 = 62.1, \psi_5 = 93.22$. The values of $DMOS_P$ for the images of Figs. 8 and 9 are indicated under respective images. For the entire images of LIVE database, the scatter plot of $DMOS_P$ and true DMOS is shown in Fig. 11. Form these results, it is apparent that the $DMOS_P$ is close to the true DMOS. It is worth to note that in LIVE database, the range of DMOS is around [0-100] for pristine images and its 5 different level of blurriness.

3.3 Accuracy comparison over standard image databases

The effectiveness of the proposed metric is also evaluated in comparison to other state-of-the-art algorithms for images from standard databases, including LIVE, VCL, TID2008, TID2013, and CSIQ. Except for the algorithms that have a dash (−) in front of them, the MATLAB code for these algorithms is publicly available. Table 4 gives the SROCC (between Q and DMOS) comparison with the other algorithms. The top three performers in each database are

Table 4 SROCC comparison of the proposed method with state-of-the-art NR blur specific metrics

	S_3 [3]	FISH$_{bb}$ [4]	LPC-SI [5]	MLV [6]	ARISM$_c$ [11]	EZMBM [12]
LIVE	0.963	0.960	0.922	0.956	**0.970**	Training
VCL	0.591	0.912	0.915	0.527	**0.931**	–
TID2008	0.850	0.825	**0.896**	0.854	0.862	0.839
TID2013	0.865	0.851	**0.920**	0.879	0.899	–
CSIQ	0.886	0.893	0.902	0.907	0.906	0.899
	SPARISH [13]	STEM-NOISE [14]	BISHARP [15]	[16]	[17]	Proposed
LIVE	**0.973**	0.923	0.960	0.959	0.962	**0.963**
VCL	**0.930**	–	0.898	0.904	0.921	**0.923**
TID2008	0.884	0.755	0.885	0.823	**0.940**	0.903
TID2013	0.892	–	0.909	0.850	**0.952**	0.933
CSIQ	0.887	–	**0.913**	0.889	**0.932**	0.927

The best values are given in bold

Table 5 PLCC comparison of the proposed method with state-of-the-art NR blur specific metrics

	S_3 [3]	$FISH_{bb}$ [4]	LPC-SI [5]	MLV [6]	$ARISM_c$ [11]
LIVE	0.945	0.956	0.907	0.946	**0.961**
VCL	0.905	0.922	0.913	0.768	**0.940**
TID2008	0.800	0.838	**0.892**	0.856	0.845
TID2013	0.848	0.866	**0.916**	0.880	0.896
CSIQ	0.770	0.921	**0.923**	0.876	0.895

	EZMBM [12]	SPARISH [13]	STEM-NOISE [14]	BISHARP [15]	[16]
LIVE	-	**0.968**	–	**0.961**	0.960
VCL	–	**0.936**	–	0.907	0.920
TID2008	–	0.887	–	0.891	0.816
TID2013	–	0.890	–	0.909	0.857
CSIQ	–	0.860	–	0.918	0.902

	[17]	Proposed
LIVE	0.926	0.954
VCL	0.860	**0.935**
TID2008	**0.918**	**0.896**
TID2013	**0.925**	**0.924**
CSIQ	**0.940**	**0.929**

The best values are given in bold

Table 6 RMSE comparison of the proposed method with state-of-the-art NR blur specific metrics

	S_3 [3]	FISH$_{bb}$ [4]	LPC-SI [5]	MLV [6]	ARISM$_c$ [11]	EZMBM [12]
LIVE	7.864	**6.40**	8.640	7.767	**5.815**	–
VCL	10.374	9.450	9.771	11.22	**7.890**	–
TID2008	0.705	0.641	**0.530**	0.612	0.623	–
TID2013	0.661	0.691	0.661	0.601	0.554	–
CSIQ	0.183	0.129	**0.112**	0.141	0.122	–
	SPARISH [13]	STEM-NOISE [14]	BISHARP [15]	[16]	[17]	Proposed
LIVE	8.386	–	**6.001**	6.733	9.053	7.156
VCL	**8.450**	–	10.501	9.562	12.434	**8.633**
TID2008	0.540	–	0.540	0.679	**0.465**	**0.520**
TID2013	0.551	–	**0.519**	0.643	**0.473**	0.477
CSIQ	0.173	–	0.117	0.124	**0.098**	**0.114**

The best values are given in bold

Table 7 Statistical significance test (F test)

	LIVE	VCL	TID2008	TID2013	CSIQ
LPC-SI	1	1	1	1	1
ARISM$_c$	1	1	1	1	1
SPARISH	0	1	0	1	1
BISHARP	1	1	0	1	0
[16]	0	1	1	1	1
[17]	1	1	0	0	1

Table 8 SROCC comparison of the proposed method with Deep learning metrics

	LIVE	TID2008	TID2013	CSIQ
Kang's CNN [19]	**0.983**	0.850	**0.922**	0.781
Yu's CNN [20]	**0.965**	**0.919**	0.914	**0.925**
MEON [21]	0.948	–	0.880	**0.905**
SGDNeT [22]	0.940	–	0.914	0.860
NSSADNN [23]	**0.986**	–	0.844	0.893
MSFF [24]	0.962	**0.928**	0.928	–
Proposed	0.963	**0.903**	**0.933**	**0.927**

The best values are given in bold

highlighted in bold font. It may be observed that the proposed metric is among one of them. While other algorithms as discussed follows are not consistent with accuracy for each of the databases. For example, the algorithms such as SPARISH [13] and ARISM/ARISM$_c$ [11] are among highly accurate algorithms for LIVE database; however, the accuracy suffers for TID2008, TID2013, and CSIQ databases.

Among the tabulated algorithms, [17] is most consistent. The proposed method is highly competitive with [17] except for TID2008 database.

By comparing the PLCC and RMSE, as given in Table 5 and Table 6, it may be observed that the proposed method is among the top three performers except for LIVE database. Therefore, overall, the proposed method is highly competitive in accuracy.

3.4 Statistical significance test

To establish the efficiency of the proposed method, statistical significance test is done using the F test (one-tailed, 5% significance level) against some competitive metrics, as given in Table 7. The *F* test method from [1] is followed in this paper. In the table, the value of '1' reflects that the proposed metric is more significant than the metrics in the corresponding rows, while '0' indicates that the tabulated metric is equally significant as the proposed one. From the table, it may be observed that the proposed metric outperforms the listed algorithms for most of the databases, reflecting the effectiveness of the proposed method.

3.5 Comparison with deep learning methods

The proposed method is also compared with state-of-the-art Deep Learning-based quality assessment methods, as given in Table 8. Most of the Deep Learning-based methods are inconsistent in accuracy across the different databases. Even though the proposed method is training free, it is better and robust than the learning-based tabulated methods.

Table 9 SROCC comparison of the proposed method with state-of-the-art NR general-purpose metrics

	BRISQUE [32]	NFEQM [18]	NFERM [33]	GMLOG [34]	PIQUE [35]	BIQES [36]	BNB [2]	PSQA [37]	Proposed
LIVE	Training	0.886	Training	Training	0.920	0.928	0.943	**0.961**	**0.963**
VCL	0.900	–	0.828	0.887	0.894	0.877	0.791	**0.934**	**0.923**
TID2008	0.875	0.741	0.844	0.789	0.817	**0.910**	0.739	–	**0.903**
TID2013	0.880	–	0.885	0.829	0.848	**0.922**	0.799	0.911	**0.933**
CSIQ	0.874	**0.894**	0.870	0.868	0.843	0.835	0.880	0.848	**0.927**

The best values are given in bold

Table 10 Time complexity (in seconds) of different algorithms for an image of size 768×512 pixels

Algorithm	Time(s)
S_3 [3]	15.10
$FISH_{bb}$ [4]	7.3
LPC-SI [5]	0.51
MLV [6]	0.19
$ARISM_c$ [11]	17.5
SPARISH [13]	10.5
BISHARP [15]	0.18
[16]	0.06
[17]	0.04
Proposed	**0.02**

The best values are given in bold

3.6 Comparison with NR general-purpose algorithms

The proposed algorithm is also compared with state-of-the-art general-purpose algorithms as tabulated in Table 9. The best two algorithm for each database is shown in bold font. From the table, it is observed that the proposed method is highly competitive over databases of VCL and TID2008 and outperforms all other algorithms over LIVE, TID2013, and CSIQ databases. Also, it may be noted that the performance of the proposed metric upholds the necessity of distortion-specific metrics.

3.7 Time complexity comparison

Other than accuracy, complexity is also an important factor of performance evaluation in real-time applications. The complexity of different algorithms are measured in terms of time-complexity and is given in Table 10. From this table, it is evident that the proposed method is simpler as it has relatively very low time-complexity in comparison to other algorithms. LPC-SI is 9 times and BISHARP is 25 times slower than the proposed method. The method proposed in this paper is significantly better in terms of time-complexity (3×faster) than [16]. It is also simpler than the competitive metric given in [17].

4 Conclusion

The variation in neighboring pixels of an image is one of the indicators of blurriness. In this paper, the maximum and mean pixel variation of an image has been exploited to estimate the quality of a blurred image. The proposed method has been found highly accurate and competitive across the standard databases for distortion specific as well as general-purpose algorithms. Moreover, it also outperforms the most

of the existing algorithms in terms of time complexity and hence may be used in real-time applications.

Availability of data and materials The database used in this work is publicly available and the source is cited in the paper.

Code availability Code can be made available on the request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest/competing interests.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication The authors gives the consent for publication.

References

1. Athar, S., Wang, Z.: A comprehensive performance evaluation of image quality assessment algorithms. IEEE Access **7**, 140030–140070 (2019)
2. Fang, R., Al-Bayaty, R., Wu, D.: Bnb method for no-reference image quality assessment. IEEE Trans. Circuits Syst. Video Technol. **27**(7), 1381–1391 (2016)
3. Vu, C.T., Phan, T.D., Chandler, D.M.: S3: A spectral and spatial measure of local perceived sharpness in natural images. IEEE Trans. Image Process. **21**(3), 934–945 (2011)
4. Vu, P.V., Chandler, D.M.: A fast wavelet-based algorithm for global and local image sharpness estimation. IEEE Signal Process. Lett. **19**(7), 423–426 (2012)
5. Hassen, R., Wang, Z., Salama, M.M.: Image sharpness assessment based on local phase coherence. IEEE Trans. Image Process. **22**(7), 2798–2810 (2013)
6. Bahrami, K., Kot, A.C.: A fast approach for no-reference image sharpness assessment based on maximum local variation. IEEE Signal Process. Lett. **21**(6), 751–755 (2014)
7. Bong, D.B., Khoo, B.E.: An efficient and training-free blind image blur assessment in the spatial domain. IEICE Trans. Inf. Syst. **97**(7), 1864–1871 (2014)
8. Bong, D.B.L., Khoo, B.E.: Blind image blur assessment by using valid reblur range and histogram shape difference. Signal Process: Image Commun **29**(6), 699–710 (2014)
9. Bong, D.B.L., Khoo, B.E.: Objective blur assessment based on contraction errors of local contrast maps. Multimed Tools Appl **74**, 7355–7378 (2015)
10. Yan, R., Shao, L.: Blind image blur estimation via deep learning. IEEE Trans. Image Process. **25**(4), 1910–1921 (2016)
11. Gu, K., Zhai, G., Lin, W., Yang, X., Zhang, W.: No-reference image sharpness assessment in autoregressive parameter space. IEEE Trans. Image Process. **24**(10), 3218–3231 (2015)
12. Lim, C., Paramesran, R., Jassim, W.A., Yu, Y.-P., Ngan, K.N.: Blind image quality assessment for gaussian blur images using exact zernike moments and gradient magnitude. J. Franklin Inst. **353**(17), 4715–4733 (2016)
13. Li, L., Wu, D., Wu, J., Li, H., Lin, W., Kot, A.C.: Image sharpness assessment by sparse representation. IEEE Trans. Multimed **18**(6), 1085–1097 (2016)
14. Nakhaei, A.A., Helfroush, M.S., Danyali, H., Ghanbari, M.: Subjectively correlated estimation of noise due to blurriness distortion based on auto-regressive model using the yule-walker equations. IET Image Proc. **12**(10), 1788–1796 (2018)
15. Gvozden, G., Grgic, S., Grgic, M.: Blind image sharpness assessment based on local contrast map statistics. J. Vis. Commun. Image Represent. **50**, 145–158 (2018)
16. Baig, M.A., Moinuddin, A.A., Khan, E., Ghanbari, M.: DFT-based no-reference quality assessment of blurred images. Multimed Tools Appl **81**(6), 7895–7916 (2022)
17. SHI Chenyang, L.Y.: No reference image sharpness assessment based on global color difference variation. Chin. J. Electron. **33**(E220058), 1 (2024)
18. Zhai, G., Wu, X., Yang, X., Lin, W., Zhang, W.: A psychovisual quality metric in free-energy principle. IEEE Trans. Image Process. **21**(1), 41–52 (2011)
19. Kang, L., Ye, P., Li, Y., Doermann, D.: Convolutional neural networks for no-reference image quality assessment. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1733–1740 (2014)
20. Yu, S., Wu, S., Wang, L., Jiang, F., Xie, Y., Li, L.: A shallow convolutional neural network for blind image sharpness assessment. PLoS ONE **12**(5), 0176632 (2017)
21. Ma, K., Liu, W., Zhang, K., Duanmu, Z., Wang, Z., Zuo, W.: End-to-end blind image quality assessment using deep neural networks. IEEE Trans. Image Process. **27**(3), 1202–1213 (2017)
22. Yang, S., Jiang, Q., Lin, W., Wang, Y.: SGDNet: An end-to-end saliency-guided deep neural network for no-reference image quality assessment. In: Proceedings of the 27th ACM International Conference on Multimedia. pp. 1383–1391 (2019)
23. Yan, B., Bare, B., Tan, W.: Naturalness-aware deep no-reference image quality assessment. IEEE Trans. Multimed **21**(10), 2603–2615 (2019)
24. He, S., Liu, Z.: Image quality assessment based on adaptive multiple Skyline query. Signal Process: Image Commun **80**, 115676 (2020)
25. Pertuz, S., Puig, D., Garcia, M.A.: Analysis of focus measure operators for shape-from-focus. Pattern Recogn. **46**(5), 1415–1432 (2013)
26. Sheikh, H.: LIVE image quality assessment database release 2. http://live.ece.utexas.edu/research/quality (2005). Accessed 13 Jan 2024
27. Zarić, A., Tatalović, N., Brajković, N., Hlevnjak, H., Lončarić, M., Dumić, E., Grgić, S.: Vcl@ fer image quality assessment database. AUTOMATIKA: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije. 53(4): 344–354. (2012)
28. Ponomarenko, N., Carli, M., Lukin, V., Egiazarian, K., Astola, J., Battisti, F.: Tampere image database. [online]. http://www.ponomarenko.info/tid2008.htm (2008)
29. Ponomarenko, N., Jin, L., Ieremeiev, O., Lukin, V., Egiazarian, K., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F., et al.: Image database tid2013: peculiarities, results and perspectives. Signal Process: Image Commun **30**, 57–77 (2015)
30. Larson, E., Chandler, D.: Categorical image quality (CSIQ) database. https://s2.smu.edu/~eclarson/csiq.html

31. Sheikh, H.R., Sabir, M.F., Bovik, A.C.: A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Trans. Image Process. **15**(11), 3440–3451 (2006)

32. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. **21**(12), 4695–4708 (2012)

33. Gu, K., Zhai, G., Yang, X., Zhang, W.: Using free energy principle for blind image quality assessment. IEEE Trans. Multimed **17**(1), 50–63 (2014)

34. Xue, W., Mou, X., Zhang, L., Bovik, A.C., Feng, X.: Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features. IEEE Trans. Image Process. **23**(11), 4850–4862 (2014)

35. Venkatanath, N., Praneeth, D., Bh, M.C., Channappayya, S.S., Medasani, S.S.: Blind image quality evaluation using perception based features. Twenty First National Conference on Communications (NCC). pp. 1–6 (2015)

36. Saha, A., Wu, Q.M.J.: Utilizing image scales towards totally training free blind image quality assessment. IEEE Trans. Image Process. **24**(6), 1879–1892 (2015)

37. Liu, L., Wang, T., Huang, H.: Pre-attention and spatial dependency driven no-reference image quality assessment. IEEE Transactions on Multimedia. **21**(9), 2305–2318 (2019). https://doi.org/10.1109/TMM.2019.2900941