Blur-Specific No-Reference Image Quality Assessment: A Classification and Review of Representative Methods

Dingquan Li and Tingting Jiang

1 Introduction

Images may suffer from a variety of distortions during image acquisition, compression, transmission, display, etc. To monitor and improve the quality of images, image quality assessment (IQA) becomes a fundamental technique for modern multimedia systems. Since human is the end-user in most multimedia devices, the most accurate image quality evaluation is achieved by subjective ratings. However, subjective evaluation is difficult to carry out in real-time applications due to its drawbacks of inconvenience, high price, and inefficiency. These drawbacks lead to the need of efficient and effective objective IQA methods that can automatically predict image quality. Objective IQA can be categorized into full-reference IQA (FR-IQA, e.g., PSNR, MSSIM [27]), reduced-reference IQA (RR-IQA, e.g., FEDM [30]), and no-reference IQA (NR-IQA, e.g., BRISQUE [17]). Due to the unavailability of the reference images in most practical applications, NR-IQA is preferable but also more challenging.

In this work, we focus on blur-specific NR-IQA. Blur is one of the most common distortions and unintentional blur impairs image quality. Image blur often occurs in the following situations: (1) out of focus, (2) object motion and camera shake, (3) nonideal imaging systems, (4) atmospheric turbulence or aerosol

D. Li
LMAM, School of Mathematical Sciences & IDM, Peking University, Beijing, People's Republic of China
e-mail: dingquanli@pku.edu.cn

T. Jiang (✉)
School of Electrical Engineering and Computer Sciences & IDM, Peking University, Beijing, People's Republic of China
e-mail: ttjiang@pku.edu.cn

scattering/absorption, and (5) image compression and image denoising [6, 9, 19]. Researches on blur-specific NR-IQA can bring new perspectives for related applications, e.g., autofocusing and image deblurring [12].

Various blur-specific NR-IQA methods have been proposed over the last two decades. In 2009, Ferzli and Karam [6] summarized several traditional methods, most of which are designed for "autofocus" applications. In 2012, Vu and Chandler [25] gave an overview and classified existing methods into edge-based methods (based on measuring the edge spread), pixel-based methods (operated in the spatial domain but with no assumption on edges), and transform-based methods (operated in the spectral domain). Many novel blur-specific NR-IQA methods have been recently developed, e.g., [7, 13–15, 29]. This work aims to give an overall classification of existing methods and systematically introduces 18 representative methods, especially the recently developed ones, so as to provide an integrated and valuable reference for blur-specific NR-IQA research and help researchers keep abreast of the recent progress.

Most existing methods only test their performance on Gaussian blur images. It is thus unclear how these methods would behave in the presence of noise or in more complex realistic situations. Stability and practicability are very important for objective methods, so we conduct comparative experiments for 13 representative methods with available codes on two sorts of images, including Gaussian blur images from TID2013 [20] and realistic blur images from BID [4]. By comparing the experimental results, we can see that most existing methods have satisfactory performance on Gaussian blur images, but they fail to accurately estimate the image quality of realistic blur images. Therefore, further study is needed in this field. At last, we discuss the issues of realistic blur, on which practical blur-specific NR-IQA methods focus.

The rest of this paper is organized as follows. In Sect. 2, existing blur-specific NR-IQA methods are classified, and several representative methods are reviewed. Then experimental settings, results, analysis, and discussions are presented in Sect. 3. Finally, concluding remarks are made in Sect. 4.

2 Blur-Specific No-Reference Image Quality Assessment: Classification and Review

Blur-specific NR-IQA methods based on traditional signal/image processing technologies have been investigated over 20 years. However, traditional signal processing technologies cannot accurately express the diversity of blur process and the complexity of human visual system (HVS), so researchers have turned to machine learning technologies for estimating image quality of blur images. Though blur-specific NR-IQA methods based on machine learning technologies have only been studied in recent years, they are in a rapid growth. In terms of this natural fact, we divide the existing blur-specific NR-IQA methods into two categories: learning-free
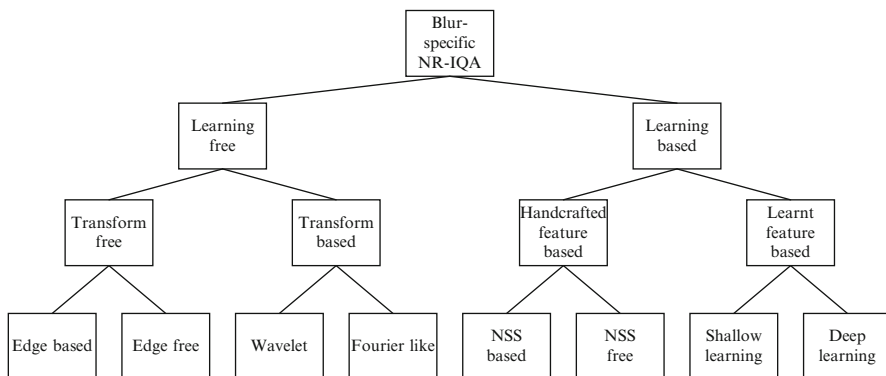
Fig. 1 An overall classification of blur-specific no-reference image quality assessment methods, including learning-free methods and learning-based methods. NSS is the abbreviation of "natural scene statistics"

methods and learning-based methods. An overall classification is shown in Fig. 1, where the representative methods in each category will be discussed in detail later. It should be noted that, for Gaussian blur images, image sharpness can be used as the synonym of image quality, while image blurriness can be used as the antonym of image quality. In previous literatures, researchers mainly considered Gaussian blur images; therefore, we not only review blur-specific NR-IQA methods but also review both sharpness and blurriness estimators. However, we should also note that the abovementioned relationships among quality, sharpness, and blurriness are not necessarily true in realistic situations.

2.1 Learning-Free Methods

Among learning-free methods, some use the characteristics of blur in the spatial domain (e.g., the spread of edges and the smoothing effects), while others further use the characteristics of blur in the transform domain (e.g., the reduction of high-frequency components and the loss of phase coherence). In this regard, learning-free methods can be classified as transform-free methods and transform-based methods.

Transform-Free Methods

Transform-free methods only make use of the spatial information. They can be further divided into edge-based methods and edge-free methods. The former makes assumptions on edges while the latter does not.

Edge-Based Methods

Given that blur affects the edge's property (e.g., blur tends to make the edge spread), a lot of blur-specific NR-IQA methods are developed based on analyzing the edges. MDWE [16], JNB [6], CPBD [19], PSI [5], and EMBM [8] are the representative edge-based methods.

A. MDWE

Marziliano et al. [16] proposed a blur-specific NR-IQA method, called MDWE, based on measuring the average edge width. The basic assumption is that blur makes the edges spread, so image blur can be directly estimated based on the edge width. The framework of MDWE is shown in Fig. 2. First, the RGB color image is transformed into the gray image. Then, the vertical Sobel filter is used as the edge detector to find vertical edges in the image. Third, for each edge point, the edge width is used as a local blur measure, where the edge width is measured by the distance between the local maximum and minimum points closest to the edge
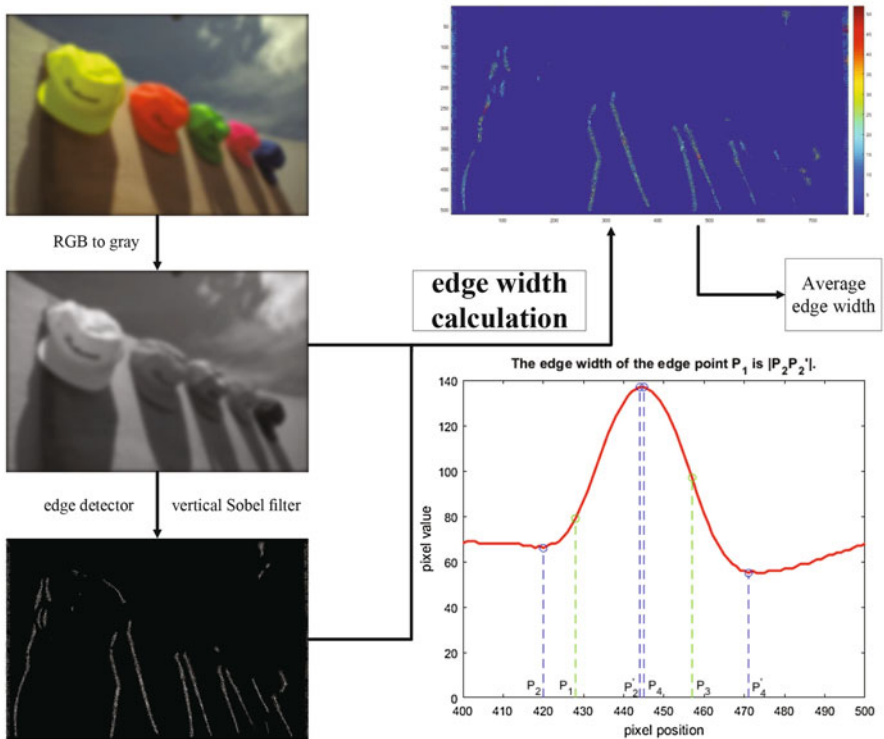

Fig. 2 The framework of MDWE, which is based on measuring the average edge width. The bottom right part is an illustration of how to compute the edge width on one row of the image

(along the horizontal direction). The bottom right part of Fig. 2 illustrates how to compute the edge width. P_2' and P_2 are the local maximum and minimum points closest to the edge point P_1, so P_1's edge width is $|P_2'P_2|$. Similarly, P_3's edge width is $|P_4'P_4|$. Finally, the global blur measure is obtained by averaging the edge width over all edge points. The smaller the global blur value is, the better image quality is.

B. JNB

Ferzli and Karam [6] estimated image quality by integrating the concept of just noticeable blur (JNB) and edge width into a probability summation model. Taking the characteristics of HVS into account, JNB is proposed to deal with the failure of estimating blur among images with different contents. It considers the minimum amount of perceived blurriness at distinct contrast levels. JNB width is the minimum edge width that people can perceive the blur, and subjective experiments were performed to obtain the JNB width. The results showed that:

$$W_{\mathrm{JNB}}(C) = \begin{cases} 5, & C \leq 50 \\ 3, & C \geq 51 \end{cases} \tag{1}$$

where W_{JNB} and C are the JNB width and the image contrast, respectively.

The probability $P(e_k)$ of detecting a blur distortion at edge e_k is determined by a psychometric function.

$$P(e_k) = 1 - \exp\left(-|\frac{W(e_k)}{W_{\mathrm{JNB}}(C_{e_k})}|^{\beta}\right) \tag{2}$$

where $W(e_k)$ is the measured edge width using MDWE [16], $W_{\mathrm{JNB}}(C_{e_k})$ is the JNB width defined in Eq. (1), and C_{e_k} is the local contrast near the edge e_k. β ranges from 3.4 to 3.8, and it is simply fixed to its median value 3.6.

Adopting a probability summation hypothesis, the localized detection probabilities over a region of interest R can be pooled as:

$$P_{\mathrm{blur}}(R) = 1 - \prod_{e_k \in R} (1 - P(e_k)) = 1 - \exp(-D_R^{\beta}) \tag{3}$$

where

$$D_R = \left(\sum_{e_k \in R} |\frac{W(e_k)}{W_{\mathrm{JNB}}(C_{e_k})}|^{\beta}\right)^{\frac{1}{\beta}}$$

Considering the size of the foveal region, the image I is divided into blocks with a block size of 64×64. The edge blocks are the blocks with more than 0.2%

edge pixels. For each edge block R_b, the probability of detecting blur in R_b can be computed by Eq. (3). Therefore, the probability of detecting blur in the image I is given by:

$$P_{\text{blur}}(I) = 1 - \prod_{R_b \in I}(1 - P_{\text{blur}}(R_b)) = 1 - \exp(-D^\beta) \qquad (4)$$

where

$$D = \left(\sum_{R_b \in I} |D_{R_b}|^\beta\right)^{\frac{1}{\beta}} \qquad (5)$$

Finally, the image quality score s is determined by:

$$s = \frac{L}{D} \qquad (6)$$

where L is the total number of processed blocks and D is given by Eq. (5).

C. CPBD

JNB method [6] is based on the assumption that the blur impairment increases when P_{blur} increases; however, it ignores that blur is not likely to be perceived when the edge width is small enough (below the JNB width). When $W(e_k)$ equals $W_{\text{JNB}}(C_{e_k})$ in Eq. (2), the probability of blur detection $P_{\text{blur}}(e_k)$ is the just noticeable blur detection probability $P_{\text{JNB}} = 63\%$. That is, when the probability of blur detection P_{blur} at edge e_k is below $P_{\text{JNB}} = 63\%$, the blur is not likely to be detected. Based on this assumption, Narvekar and Karam [19] used the concept of JNB together with a cumulative probability of blur detection. The flowchart of CPBD is shown in Fig. 3a, which is the same as JNB method except for the last pooling step. The last pooling step in CPBD is obtained from the normalized histogram of the probability of blur detection P_{blur} in the image, and the image quality score equals to the percentage of edges whose blur cannot be detected, which can be calculated as:

$$\text{CPBD} = P(P_{\text{blur}} \leq P_{\text{JNB}}) = \sum_{P_{\text{blur}}=0}^{P_{\text{JNB}}} P(P_{\text{blur}}) \qquad (7)$$

D. PSI

Feichtenhofer et al. [5] proposed a perceptual sharpness index (PSI) based on the statistical analysis of local edge gradients. First, an adaptive edge selection procedure based on a threshold and thinning process is applied to select the most
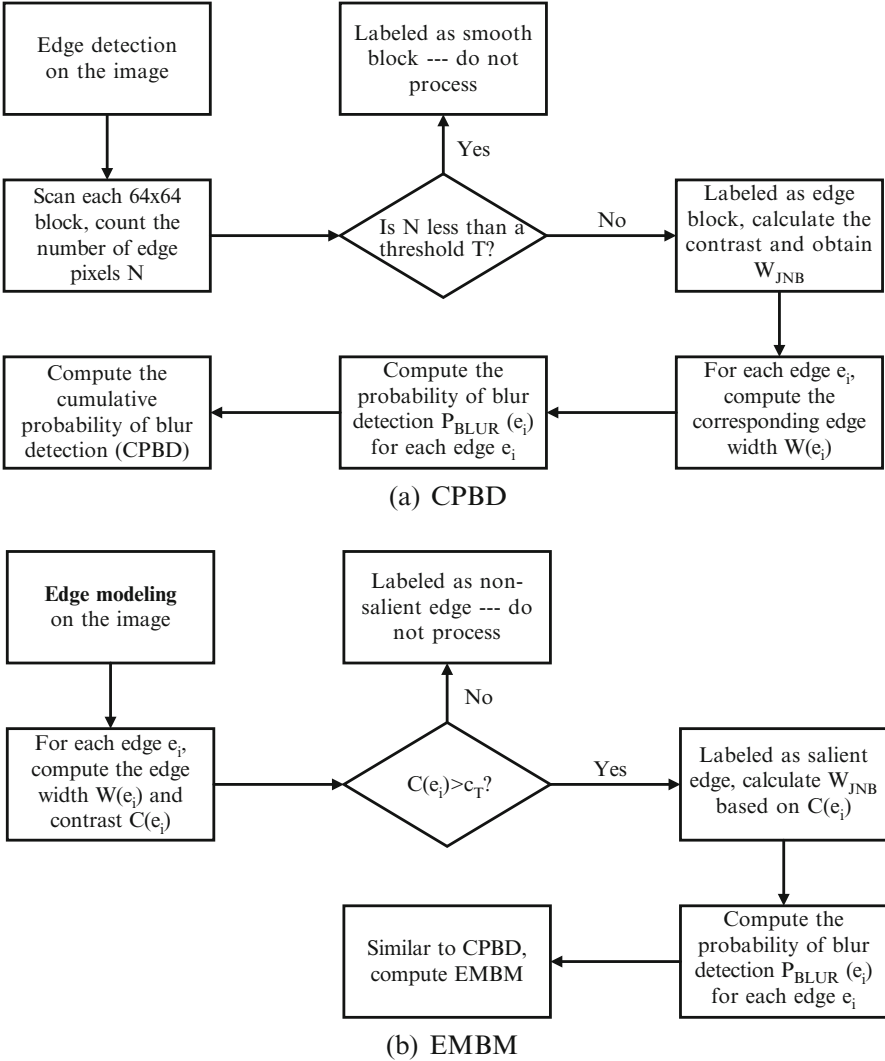
(a) CPBD

```
Edge detection          Labeled as smooth
on the image            block --- do not
                        process
   │                        ▲
   │                        │ Yes
   ▼                        │
Scan each 64x64     Is N less than a   No   Labeled as edge
block, count the    threshold T?      ───►  block, calculate the
number of edge                              contrast and obtain
pixels N                                    W_JNB
                                                │
                                                ▼
Compute the         Compute the         For each edge e_i,
cumulative          probability of blur  compute the
probability of blur  detection P_BLUR (e_i)  corresponding edge
detection (CPBD)    for each edge e_i    width W(e_i)
```

(a) CPBD

(b) EMBM

```
Edge modeling           Labeled as non-
on the image            salient edge --- do
                        not process
   │                        ▲
   │                        │ No
   ▼                        │
For each edge e_i,   C(e_i)>c_T?    Yes   Labeled as salient
compute the edge                    ───►  edge, calculate W_JNB
width W(e_i) and                          based on C(e_i)
contrast C(e_i)                                │
                                                ▼
Similar to CPBD,      Compute the
compute EMBM          probability of blur
                      detection P_BLUR (e_i)
                      for each edge e_i
```

(b) EMBM

Fig. 3 The flowcharts of CPBD and EMBM

significant edges in the image. Second, the edge widths of the selected edges are computed by an improved edge width measurement based on diagonal edge gradients. Third, according to the human perception of acutance, edge widths above the JNB width are subtracted by the edge slopes. Finally, the local sharpness map is deduced by applying the above three steps in a block-wise way. Since the sharpest regions in an image are most related to human sharpness perception, the global image quality score is determined by the highest Qth percentile average of the local sharpness values.

E. EMBM

Guan et al. [8] proposed a blur-specific NR-IQA method EMBM by integrating the concept of edge modeling into JNB, where the edge is modeled as the step function convolved with a Gaussian function. The flowchart of EMBM is shown in Fig. 3b. There are some differences between EMBM and CPBD. First, unlike CPBD, the parametric edge model in EMBM is used for edge description and detection, from which the edge width and the local contrast for each edge can be computed simultaneously. Second, all edges are depicted parametrically, so EMBM needs not to be performed in a block-wise way. Third, EMBM only considers salient edges (with large contrast) that grab most attention from human visual perception.

Edge-Free Methods

Operating in the spatial domain but with no assumption on edges, edge-free methods are based on the intrinsic characteristics of the image, the comparison between an image and its re-blurred version, or the comparison between pixels and their adjacent pixels. SVC [21], BIBS [3], MLV [1], ARISMc [7], and CATV [2] are the representative edge-free methods.

A. SVC

Sang et al. [21] observed that the singular values in the singular value curve (SVC) decay exponentially, and they decay even faster with larger degree of blur. So the degree of attenuation can be used to capture the image blur. Since the shape of SVC closely resembles an inverse power function, Sang et al. fitted the singular value curve by the equation $y = x^{-q}$. After taking logarithms of the equation, the fitting process can be achieved by linear regression. Denote the kth singular value as $s_k, k = 1, 2, \cdots, r$, then the estimated q follows the formula:

$$q = \frac{\sum_{k=1}^{r} \ln k \ln \frac{1}{s_k}}{\sum_{k=1}^{r} \ln k \ln k} \tag{8}$$

Since the tails of the singular value curve are almost indistinguishable, they are not helpful for the estimation of q. Therefore, in practice, the truncated sum is considered:

$$q = \frac{\sum_{s_k > c} \ln k \ln \frac{1}{s_k}}{\sum_{s_k > c} \ln k \ln k} \tag{9}$$

where c is a threshold value and it is set to 50.

$$J=I*G(\sigma)$$

$$\boxed{J=I} \;\; \boxed{J \neq I} \;\; \boxed{J=J(\sigma_{\max})}$$

$$\sigma = 0 \qquad \sigma = \sigma_{\min} \qquad \sigma = \sigma_{\max}$$

Fig. 4 The re-blurred process of image I in BIBS. $G(\sigma)$ is the 3×3 Gaussian blur kernel with standard deviation σ

B. BIBS

Image quality can be measured in a "FR-like" fashion by comparing the blurred image with its re-blurred version which is generated by applying a Gaussian filter. This is based on the observation that the blurred image changes less than the pristine image after the re-blurring process. Bong et al. [3] predicted blind image blur score (BIBS) by applying a re-blurring process (see Fig. 4), where two specific states in the re-blurring process are selected: the state ($\sigma = \sigma_{\min}$) that the re-blurred image starts to change its pixel values, and the state ($\sigma = \sigma_{\max}$) that re-blurred image never changes anymore. Then, image quality is measured based on the shape difference of local histogram between the image and its re-blurred versions.

C. MLV and CATV

Bahrami and Kot [1] proposed a novel blur-specific NR-IQA method based on the content-based weighting distribution of the maximum local variation (MLV). MLV of a pixel $I_{i,j}$ is defined as the maximum variation between the intensity of $I_{i,j}$ with respect to its 8-neighbor pixels:

$$\psi(I_{i,j}) = \max\{|\, I_{i,j} - I_{x,y}\,| \mid x = i - 1, i, i + 1; \, y = j - 1, j, j + 1.\} \quad (10)$$

And the MLV map $\Psi(I)$ of an image I is constructed by:

$$\Psi(I) = \begin{pmatrix} \psi(I_{1,1}) & \cdots & \psi(I_{1,N}) \\ \vdots & \ddots & \vdots \\ \psi(I_{M,1}) & \cdots & \psi(I_{M,N}) \end{pmatrix} \quad (11)$$

where M and N are the numbers of row and column in the image I, respectively.

Since variations in the pixel values can be an indication of image quality, the statistics of MLV distribution can be used for quality assessment. Bahrami and Kot [1] modeled the MLV distribution by the generalized Gaussian distribution (GGD), which is given by:

$$f(\Psi(I); \mu, \gamma, \sigma) = \left(\frac{\gamma}{2\sigma \Gamma(\frac{1}{\gamma}) \sqrt{\frac{\Gamma(\frac{1}{\gamma})}{\Gamma(\frac{3}{\gamma})}}} \right) e^{-\left(\left(\frac{\Psi(I)-\mu}{\sigma \sqrt{\frac{\Gamma(\frac{1}{\gamma})}{\Gamma(\frac{3}{\gamma})}}} \right) \right)^{\gamma}} \tag{12}$$

where μ, σ, and γ are the mean, standard variation, and shape parameter, respectively. $\Gamma(\cdot)$ is the Γ function. The estimated standard deviation σ is an indicator of image blurriness, where σ increases by decreasing image blurriness. To take the human sharpness perception into account, high variation regions should be emphasized. This can be achieved by assigning higher weights to the larger MLV pixels, which results in a weighted MLV map $\Psi_w(I)$

$$\Psi_w(I) = \begin{pmatrix} w_{1,1}\psi(I_{1,1}) & \cdots & w_{1,N}\psi(I_{1,N}) \\ \vdots & \ddots & \vdots \\ w_{M,1}\psi(I_{M,1}) & \cdots & w_{M,N}\psi(I_{M,N}) \end{pmatrix} \tag{13}$$

where weight is defined as $w_{i,j} = e^{\eta_{i,j}}$ and $\eta_{i,j}$ is the normalized rank (ranging from 0 to 1) of $\psi(I_{i,j})$ when $\Psi(I)$ is sorted in ascending order.

Instead of considering the MLV distribution, later Bahrami and Kot [2] proposed a content aware total variation (CATV) method by parameterizing the image total variation (TV) distribution using GGD. Image quality is defined as the standard deviation σ modified by the shape parameter γ to account for the image content variation, i.e, image quality score s is given by:

$$s = \frac{\sigma}{\gamma^{\frac{1-\gamma}{2}}} \tag{14}$$

D. ARISMc

Gu et al. [7] proposed ARISMc to estimate image quality based on the analysis of the locally estimated coefficients in the autoregressive (AR) parameter space. The framework of the proposed ARISMc method is shown in Fig. 5.

To take chrominance information into account, the image is first transferred to YIQ color space. Then, for each channel, the AR parameters at each pixel are estimated based on an 8th-order AR model.

Since image blurring can increase the similarity of locally estimated AR parameters, image quality can be assessed by energy difference and contrast difference of locally estimated AR parameters. The energy difference $E_{i,j}$ and contrast difference $C_{i,j}$ of AR parameters at pixel (i, j) are given by:

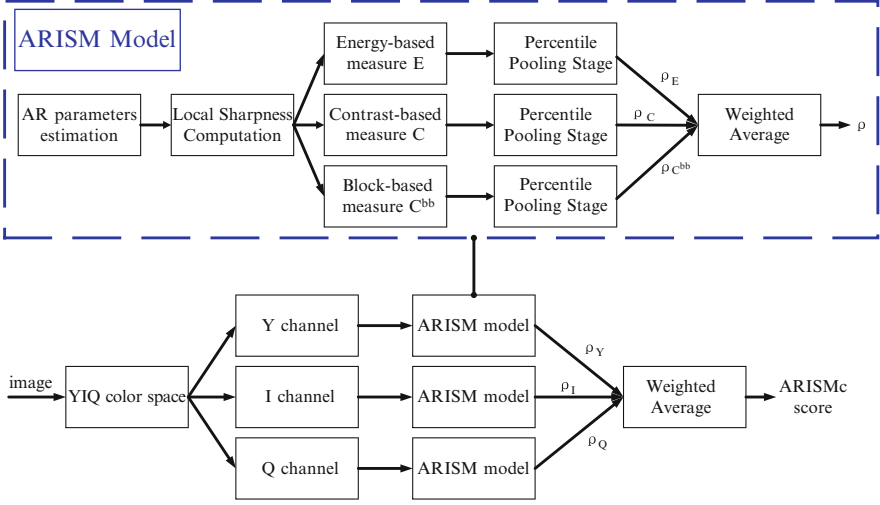$$E_{i,j} = | \max_{(x,y)\in\Omega_{i,j}} W_{x,y} - \min_{(x,y)\in\Omega_{i,j}} W_{x,y} |^2 \tag{15}$$

Fig. 5 The framework of ARISMc

$$C_{i,j} = \frac{|\max_{(x,y)\in\Omega_{i,j}} W_{x,y} - \min_{(x,y)\in\Omega_{i,j}} W_{x,y}|^2}{\max_{(x,y)\in\Omega_{i,j}} W_{x,y}^2 + \min_{(x,y)\in\Omega_{i,j}} W_{x,y}^2} \tag{16}$$

where $\Omega_{i,j} = \{(x,y)|x \in [i-1, i+1], y \in [j-1, j+1], (x,y) \neq (i,j)\}$ and $\{W_{x,y}, (x,y) \in \Omega_{i,j}\}$ denotes the estimated AR parameters at pixel (i,j).

The contrast difference can be further modified into block-based version:

$$C_{u,v}^{bb} = \frac{1}{B}\sqrt{\sum_{(i,j)\in\Phi_{u,v}} C_{i,j}} \tag{17}$$

where $\Phi_{u,v} = \{(i,j)|i \in [(u-1)B, uB], j \in [(v-1)B, vB]\}, 1 \leq u \leq \lfloor M/B \rfloor, 1 \leq v \leq \lfloor N/B \rfloor$, B is the block size and M and N are the row and column of the image.

At percentile pooling stage, the largest $Q_k\%$ values in the $k(k \in \{E, C, C^{bb}\})$ are averaged to obtain ρ_k. Then, the overall score for an image channel is given by a weighted average of $\rho_k(k \in \{E, C, C^{bb}\})$. Finally, the ARISMc score for estimating the image quality is given by a weighted average of the three overall scores in YIQ channels.

Transform-Based Methods

In the transform domain, blur has some quality-relevant characteristics such as the reduction of high-frequency components and the loss of phase coherence.

Transform-based methods utilize wavelet transform or "Fourier-like" transform (e.g., Fourier transform and cosine transform), so we categorize them into wavelet-based methods and Fourier-like methods.

Wavelet Methods

A. FISHbb

Blur leads to the reduction of high-frequency components. Vu and Chandler [24] proposed a fast wavelet-based method FISHbb by analyzing the energies of the high-frequency coefficients. The image is broken into 16×16 blocks with 50% overlapping in advance.

For each block, the DWT coefficients are obtained by Cohen-Daubechies-Fauraue 9/7 filters with three levels of decomposition. The three high-frequency sub-bands are denoted as S_{LH_n}, S_{HL_n}, and S_{HH_n}, $(n = 1, 2, 3)$. Then, the overall log energy at each DWT level is computed as the weighted average log energy of the three high-frequency sub-bands, which is given by:

$$E_n = (1 - \alpha) \frac{E_{LH_n} + E_{HL_n}}{2} + \alpha E_{HH_n}, \alpha = 0.8. \tag{18}$$

$$E_{XY_n} = \log_{10} \left[1 + \frac{1}{M_n N_n} \sum_{i,j} S^2_{XY_n}(i, j) \right], XY \in \{LH, HL, HH\}. \tag{19}$$

where M_n and N_n are the size of the nth sub-band.

Next, the overall sharpness score FISH of a block is given by the weighted sum of the overall log energy in the three levels, which is obtained by:

$$\text{FISH} = \sum_{n=1}^{3} 2^{3-n} E_n \tag{20}$$

Finally, a sharpness map is derived. And to consider the human sharpness perception, the single FISH_bb score is computed by the root mean square of the 1% largest values of the local sharpness map.

$$\text{FISH}_{bb} = \sqrt{\frac{1}{K} \sum_{k=1}^{K} \text{FISH}^2_k} \tag{21}$$

where K denotes the number of blocks which received the 1% largest FISH scores of the sharpness map and $\text{FISH}_k, k = 1, \cdots, K$ denote the FISH scores of these blocks.

B. LPC

Wang and Simoncelli [26] showed that step edges result in strong local phase coherence (LPC) structures across scales and space in the complex wavelet transform domain and blur causes the loss of such phase coherence. This gives a different perspective for understanding blur perception. Following this idea, Hassen et al. [9] proposed a blur-specific NR-IQA method based on the strength of the LPC near edges and lines. Figure 6 shows a simple flowchart of LPC. First, the image is passed through 3-scale 8-orientation log-Gabor filters, and the complex coefficient at the ath scale, the bth orientation, and the kth spatial location is denoted as c_{abk}. Then the LPC strength at bth orientation and kth spatial location is computed by:

$$
S_{\text{LPC}}^{\{b,k\}} = \cos\left(\sum_{a=1}^{3} w_a \Phi\{c_{abk}\}\right) = \cos\left(\Phi\left\{\prod_{a=1}^{3} c_{abk}^{w_a}\right\}\right) = \frac{\mathscr{R}\left\{\prod_{a=1}^{3} c_{abk}^{w_a}\right\}}{|\prod_{a=1}^{3} c_{abk}^{w_a}|} \quad (22)
$$

where $\Phi\{\cdot\}$ and $\mathscr{R}\{\cdot\}$ are the phase function and the real part of a complex number. $[w_1, w_2, w_3] = [1, -3, 2]$ denotes the weights during LPC evaluation.

Then the LPC strength measure at all orientations and kth spatial location is pooled by a weighted average. To give higher importance for the orientations with more energy, the weights are determined by the magnitude of the first (finest) scale coefficient c_{1bk}. So the LPC strength measure at kth spatial location is given by:
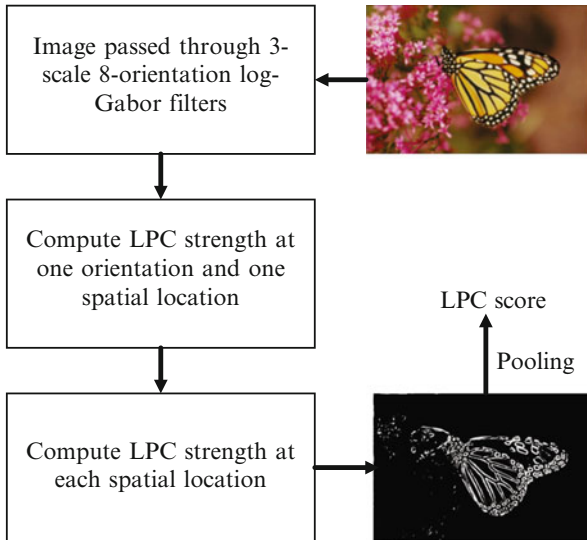

Fig. 6 A simple flowchart of LPC. It consists of three steps: computation of complex coefficients, LPC strength calculation, and LPC strength pooling

$$S_{\mathrm{LPC}}^{\{k\}} = \frac{\sum_{b=1}^{8} |c_{1bk}| S_{\mathrm{LPC}}^{\{b,k\}}}{\sum_{b=1}^{M} |c_{1bk}| + c_0} \tag{23}$$

where $c_0 = 2$ is a constant to avoid instability when $\sum_{b=1}^{M} |c_{1bk}|$ is close to zero.

Finally, a spatial LPC map is obtained. Let $S_{\mathrm{LPC}}^{\{(k)\}}, k = 1, \cdots, K$ denote the sorted LPC strength values in descending order such that $S_{\mathrm{LPC}}^{\{(1)\}} \geq S_{\mathrm{LPC}}^{\{(2)\}} \geq \cdots \geq S_{\mathrm{LPC}}^{\{(K)\}}$. To emphasize the importance of the sharpest regions in human visual perception, the overall image quality score S_{LPC} is obtained by a weighted average, where the weights are assigned based on the ranks of LPC values.

$$S_{\mathrm{LPC}} = \frac{\sum_{k=1}^{K} u_k S_{\mathrm{LPC}}^{\{(k)\}}}{\sum_{k=1}^{K} u_k} \tag{24}$$

where u_k is the weight assigned to the k-th ranked spatial LPC value and is calculated by:

$$u_k = \exp\left[-\left(\frac{k-1}{K-1}\right)/\beta_k\right], \beta_k = 1e - 4. \tag{25}$$

Fourier-Like Methods

A. S3

Vu and Chandler [25] proposed a blur-specific NR-IQA method S3 based on the combination of spectral and spatial measures. The flowchart of S3 is shown in Fig. 7. According to the reduction of high-frequency components in blur images, the spectral measure $S_1(\mathbf{x})$ of a block $\mathbf{x}$ is initially defined as the slope of the local magnitude spectrum $\alpha_{\mathbf{x}}$, then rectified by a sigmoid function to account for HVS, i.e.,

$$S_1(\mathbf{x}) = 1 - \frac{1}{1 + e^{\beta_1(\alpha_{\mathbf{x}} - \beta_2)}}, \beta_1 = -3, \beta_2 = 2. \tag{26}$$

To further consider the contrast effect, the spatial measure $S_2(\mathbf{x})$ of a block $\mathbf{x}$ is calculated based on the local total variation, which is given by:

$$S_2(\mathbf{x}) = \frac{1}{4} \max_{\xi \in \mathbf{x}} \mathrm{TV}(\xi) \tag{27}$$

where ξ is a 2×2 block of $\mathbf{x}$ and $\mathrm{TV}(\xi)$ is the total variation of ξ.

Then, the overall sharpness map S_3 of the image I is obtained by a geometric mean of spectral and spatial measures in a block-wise way:

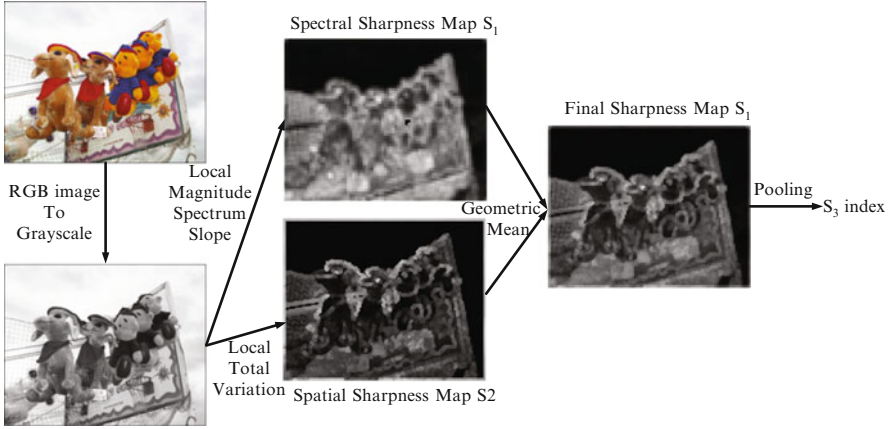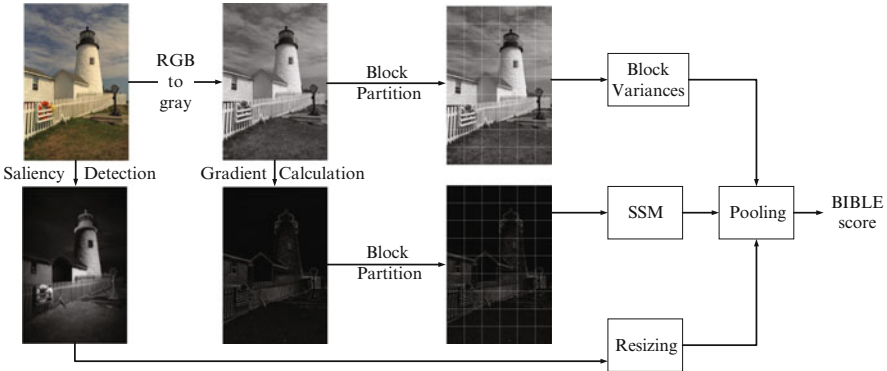**Fig. 7** The flowchart of S3

Fig. 8 The flowchart of BIBLE

$$S_3(\mathbf{x}) = \sqrt{S_1(\mathbf{x})S_2(\mathbf{x})}, \mathbf{x} \in I. \tag{28}$$

Finally, to consider the human sharpness perception, the overall sharpness score is calculated as the average of the largest 1% values of the overall sharpness map.

B. BIBLE

Having observed that blur affects the moment energy, Li et al. [13] presented a blind image blur evaluator (BIBLE) to assess image quality based on the variance-normalized moment energy. The flowchart of BIBLE is shown in Fig. 8. The gradient image is divided into equal-sized blocks, and the Tchebichef moments [18] of all blocks are computed. Then the block's energy is calculated by summing up

the squared non-DC moments. Finally, image quality is measured by the variance-normalized moment energy together with a visual saliency model to adapt to the HVS characteristics.

2.2 Learning-Based Methods

Recently, researchers turn to machine learning technologies for blur-specific NR-IQA. Learning-based methods comprise two steps: feature extraction and quality prediction. The most important thing is to extract features that can reflect image quality. Once it is done, quality prediction can be achieved by support vector regression (SVR), neural network, probabilistic prediction model, etc. To emphasize the importance of feature extraction, the learning-based methods are classified as handcrafted feature-based methods and learnt feature-based methods.

Handcrafted Feature-Based Methods

Handcrafted features are generally extracted from the nature scene statistic (NSS) models. Meanwhile, they can also be obtained by some NSS-free low-level image features (e.g., contrast, brightness, etc.) So we divide the handcrafted feature-based methods into NSS-based methods and NSS-free methods.

A Representative NSS-Based Method BIBE

It is assumed that natural scenes contain certain statistical properties that could be altered by the existence of distortions. Therefore, by modeling the statistical distributions of image coefficients, image quality can be estimated by deviations of these statistics. Wang et al. [28] proposed a blur-specific NR-IQA method BIBE based on the NSS of gradient distribution, where the flowchart of BIBE is shown in Fig. 9. First, the blurred image is passed through the horizontal and vertical Prewitt filters to get the gradient map. Then, the gradient-related distributions represented by histograms are modeled using the generalized Gaussian distribution (GGD) or asymmetric GGD. Finally, the NSS features (parameters of the models) are fed into the extreme learning machine [10] to predict the image quality.
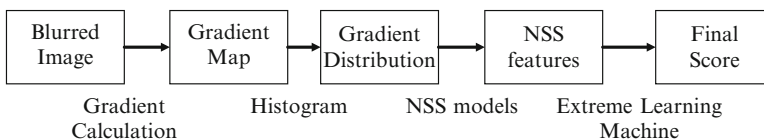

Fig. 9 The flowchart of BIBE

A Representative NSS-Free Method RISE

Inspired by the multi-scale characteristics of HVS when perceiving visual scenes, Li et al. [15] proposed a blur-specific NR-IQA method RISE based on multi-scale features extracted in both the spatial and spectral domains. The flowchart of RISE is shown in Fig. 10.

For an image $I(x, y)$, the scale space $L(x, y, \sigma)$ can be first built by convoluting it with a series of Gaussian filters $G(x, y, \sigma)$:

$$L(x, y, \sigma) = I(x, y) * G(x, y, \sigma), G(x, y, \sigma) = \frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/2\sigma^2} \qquad (29)$$

where σ is the scale and $*$ denotes the convolution.

Second, multi-scale gradient similarity maps $\mathbf{GS}_k, k = 1, 2, 3, 4$ can be obtained by:

$$\mathbf{GS}_k = \frac{\mathbf{D}_k\mathbf{D}_0 + c_1}{\mathbf{D}_k^2 + \mathbf{D}_0^2 + c_1} \qquad (30)$$

where $\mathbf{D}_k$ is the gradient map of the k-th scale image ($k = 1, 2, 3, 4$), $\mathbf{D}_0$ is the gradient map of the original image. c_1 is a small constant to ensure numerical stability.

The gradient similarity features are defined as:

$$f_k^G = \frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N}\mathbf{GS}_k(x, y), k = 1, 2, 3, 4 \qquad (31)$$
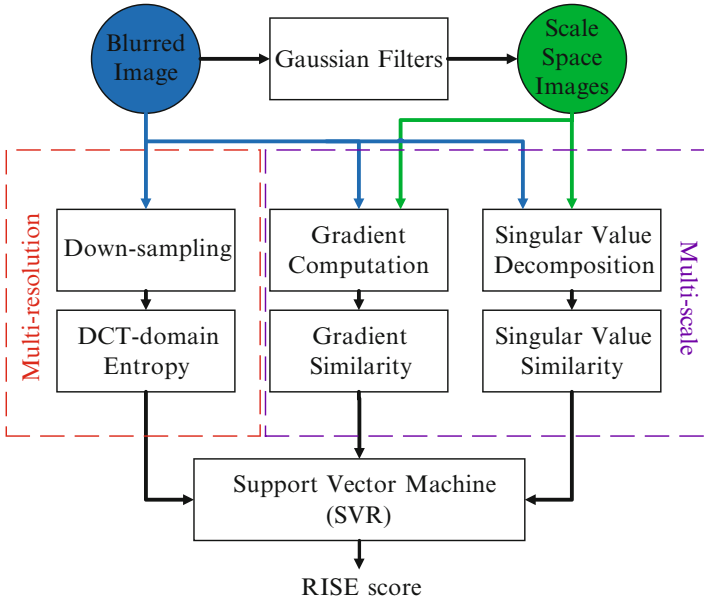

Fig. 10 The flowchart of RISE

where M and N are the number of rows and columns in the image.

Third, multi-scale singular value similarity f_k^S, $k = 1, 2, 3, 4$ is defined as:

$$f_k^S = \frac{\mathbf{s}_k^T \mathbf{s}_0 + c_2}{\mathbf{s}_k^2 + \mathbf{s}_0^2 + c_2} \tag{32}$$

where $\mathbf{s}_k$ is the singular values of the k-th scale image ($k = 1, 2, 3, 4$) and $\mathbf{s}_0$ is the singular values of the original image. c_2 is a small constant to ensure numerical stability.

To take the impact of the viewing distance into account, the blurred image is also down-sampled to get multi-resolution images, and the DCT domain entropies of all multi-resolution images are calculated as the third type of features.

Finally, all the three types of features are concatenated and fed into an SVR model with RBF kernel to get the quality score.

Learnt Feature-Based Methods

Learnt feature-based methods utilize machine learning methods to learn powerful features that can strongly reflect image quality. These methods can be divided into two categories: shallow learning methods and deep learning methods, in terms of whether features are extracted from shallow learning architectures or deep learning architectures.

A Representative Shallow Learning Method SPARISH

Having observed that over-complete dictionaries learned from natural images can capture edge patterns, Li et al. [14] proposed a blur-specific NR-IQA method SPARISH based on the sparse representation. Figure 11 shows the flowchart of SPARISH. An over-complete dictionary is learnt to construct a sparse coding model for the image gradient blocks, then the variance-normalized block energy over high-variance image blocks is used as the quality score, where the block energy is obtained from the sparse coefficients.
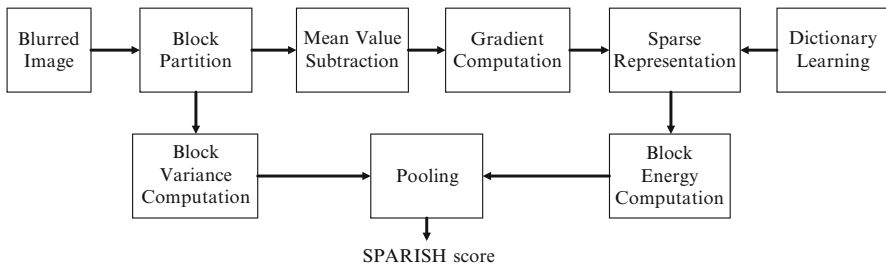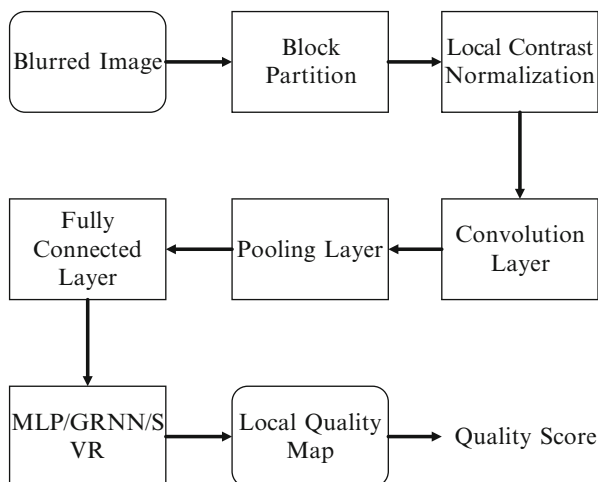

Fig. 11 The flowchart of SPARISH

Fig. 12 The flowchart of Yu's CNN. MLP is the abbreviation of multilayer perceptron, GRNN indicates the general regression neural network and SVR means support vector regression

A Representative Deep Learning Method Yu's CNN

Human visual mechanism is very complicated and cannot be accurately expressed by shallow learning architectures. Recently, deep learning techniques have been applied for general purpose IQA [11]. Yu et al. [29] made an attempt on applying deep learning architectures to blur image quality assessment. The flowchart of Yu's CNN is shown in Fig. 12. The image patches pre-processed by the local contrast normalization are passed through a convolutional layer, a down-sampling layer and a fully connected layer to extract patch features and then the features are mapped to patch quality scores by a regression model (MLP, GRNN or SVR). Finally, the average of patch quality scores is used as the overall image quality score.

3 Experiments

3.1 Experimental Settings

Evaluated Methods

We choose the 13 representative methods with available codes for comparative experiments, i.e., 6 transform-free methods (MDWE, CPBD, PSI, EMBM, MLV, ARISMc), 4 transform-based methods (FISHbb, LPC, S3, BIBLE), 1 handcrafted feature-based method RISE, and 2 learnt feature-based methods (SPARISH and Yu's

CNN). The quality prediction models for learning-based methods are trained on the LIVE blur database [22].

Evaluation Criteria

Video Quality Experts Group (VQEG) [23] suggests to map the objective score o to the subjective score s using a four-parameter logistic function:

$$F(o) = \frac{\tau_1 - \tau_2}{1 + e^{\frac{o - \tau_3}{\tau_4}}} + \tau_2 \tag{33}$$

where τ_1, τ_2, τ_3, and τ_4 are free parameters to be determined during the nonlinear mapping process, with initial values as $\tau_1 = \max(s)$, $\tau_2 = \min(s)$, $\tau_3 = \mathrm{mean}(o)$, and $\tau_4 = \mathrm{std}(o)/4$.

Three evaluation criteria are chosen to evaluate the method's performance: Spearman's rank-order correlation coefficient (SRCC), Pearson's linear correlation coefficient (PLCC), and root-mean-square error (RMSE). SRCC indicates how well the relationship between subjective and objective scores can be described using a monotonic function. PLCC is a measure of the linear correlation between the subjective and objective scores after the nonlinear mapping. RMSE is used to measure the differences between the subjective and objective scores after the nonlinear mapping. For a good method, the values of SRCC and PLCC are close to 1, while the value of RMSE is close to 0.

Testing Databases

We consider blurred images from TID2013 [20] and BID [4]. **Gaussian blur images** from TID2013 are obtained using Gaussian filters, which are to approximate the out-of-focus blur. There are 125 blurred images generated from 25 reference images and 5 blur kernels. **Realistic blur images** from BID are taken from real world along with a variety of scenes, camera apertures, and exposure time. There are 586 images, most of which suffer from realistic out-of-focus blur or motion blur. The subjective scores are in the form of mean opinion score (MOS) with a range [0, 9] on TID2013 and [0, 5] in BID.

3.2 Experimental Results

The performance comparison is shown in Table 1, where the best three values are in boldface. On Gaussian blur images from TID2013, most existing methods correlate well with subject ratings in terms of SRCC and PLCC (SRCC, PLCC > 0.8). On realistic blur images from BID, FISHbb outperforms the others. However, the SRCC

Table 1 Performance comparison and average computational time (seconds/image) on TID2013 [20] and BID [4], where the best three values in each column are marked in boldface

Method	TID2013 [20]				BID [4]			
	SRCC↑	PLCC↑	RMSE↓	Time(s)↓	SRCC↑	PLCC↑	RMSE↓	Time(s)↓
MDWE [16]	0.816	0.835	0.686	0.184	0.307	0.320	1.186	3.333
CPBD [6]	0.852	0.855	0.647	**0.068**	0.018	0.004	1.252	**1.826**
PSI [5]	0.868	0.879	0.594	**0.015**	0.069	0.198	1.228	**0.210**
EMBM [8]	0.865	0.875	0.604	0.212	0.299	0.314	1.189	2.913
MLV [1]	0.879	0.883	0.587	**0.037**	0.320	0.363	1.167	**0.481**
ARISMc [7]	**0.901**	0.898	0.549	6.466	-0.013	0.155	1.237	87.966
FISHbb [24]	0.858	0.876	0.603	0.167	**0.477**	**0.473**	**1.103**	2.283
LPC [9]	0.889	0.892	0.565	0.565	0.316	**0.385**	**1.155**	8.378
S3 [25]	0.861	0.881	0.590	4.927	**0.413**	0.423	**1.134**	52.656
BIBLE [13]	**0.899**	**0.905**	**0.531**	0.916	**0.361**	0.370	1.163	10.982
RISE [15]	**0.932**	**0.923**	**0.481**	0.345	0.120	0.135	1.252	7.138
SPARISH [14]	0.893	**0.900**	**0.543**	1.485	0.307	0.328	1.183	25.217
Yu's CNN [29]	0.843	0.860	0.639	3.100	0.030	0.160	1.236	40.857

The negative of SRCC indicates that the prediction trend is contrary to what it is supposed to be

of FISHbb is less than 0.5 (far less than 1), which indicates that there is still a large space for designing robust and effective blur-specific NR-IQA methods for realistic blur images.

Computational time is also an important aspect for evaluating the performance of NR-IQA methods since many practical applications need to run in real time. Images in TID2013 are 512×384 pixels, while the size of images in BID is larger, ranging from 1280×960 to 2272×1704. All tests are carried out on a desktop computer with Intel Core i7 6700 K CPU at 4 GHz, 32 GB RAM, Windows 10, and Matlab R2016a (Yu'CNN is implemented using Python 2.7.6 and tested on Ubuntu 14.04 using the CPU of the same desktop computer). We used the default settings of the codes and did not optimize them. In Table 1, we also report the average computational time (seconds/image) on TID2013 and BID. We can observe that (1) the fast three methods (PSI, MLV, CPBD) are operated in the spatial domain and (2) most methods run fast on TID2013. However, as the image size increases, methods such as ARISMc, S3 get quite slow on BID, which cannot meet the requirement of time-sensitive applications.

Discussion on Realistic Blur

It is hard to model all the influence factors in the real world. Besides the Gaussian and out-of-focus blur, there are other crucial factors to be considered, e.g., motion blur, ghosting, macrophotography, and image content variation in Fig. 13.

- Motion blur: there are few NR-IQA methods for assessing the quality of motion blur images, though its related problem "motion deblurring" is a hot topic.
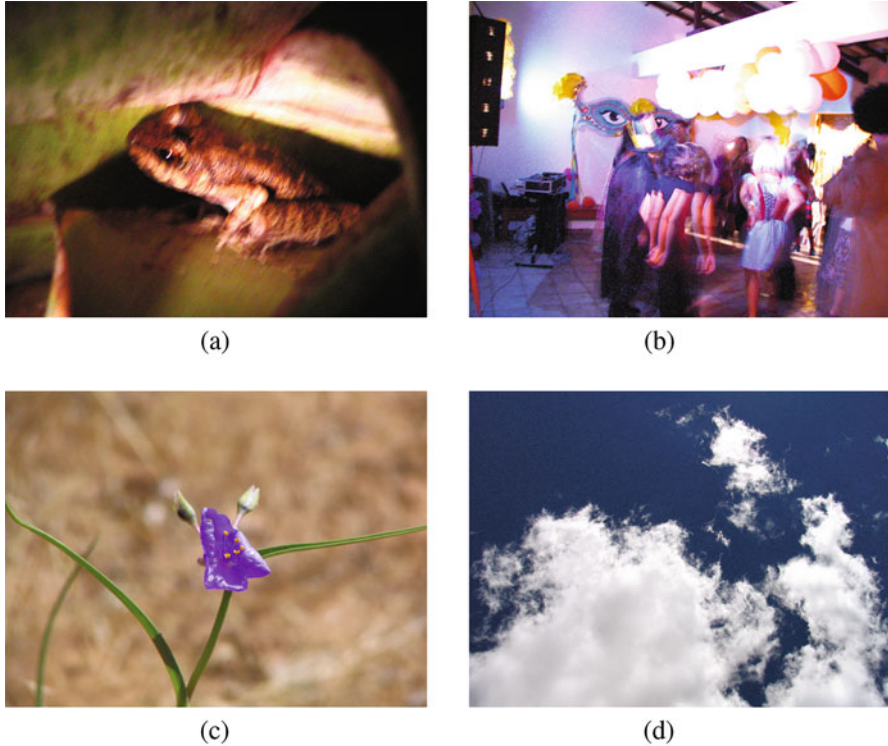
Fig. 13 Crucial factors (besides the Gaussian and out of focus blur) that influence the quality of realistic blur images. (**a**) Motion blur. (**b**) Ghosting. (**c**) Macrophotography. (**d**) Image content variation

Motion blur has directionality while Gaussian blur is isotropic. In terms of this specific characteristic of motion blur, one may further consider the directionality and the directional features for quality estimation on motion blur images. We believe that a large realistic motion blur image database with subjective ratings will facilitate the works.

- Ghosting: ghosting effect arises when the motion degree is very high, which differs from the ordinary motion blur.
- Macrophotography: the blur in Bokeh is to strengthen the photo's expressiveness. In view of this, to evaluate the quality of macrophotography images, aesthetic factors may need to be taken into account.
- Image content variation: due to the image content variation, NR-IQA methods may produce quite different objective scores for images with very similar subjective quality. At the meantime, NR-IQA methods may produce inconsistent predictions on image pairs with quite different subjective quality. To ease the impact of image content variation on blur-specific NR-IQA methods, the image content variation and blur distortion level should be jointly considered.

4 Conclusion

In this paper, we have classified previous works and have reviewed 18 representative blur-specific NR-IQA methods. Remarkable progress has been made in the past decades, evidenced by a number of state-of-the-art methods correlating well with subjective evaluations on Gaussian blur images. However, experimental results have also shown that most of the existing methods fail to estimate image quality of realistic blur images. It is the evidence that the blur-specific NR-IQA problem is far from being solved. We have also discussed on realistic blur, especially the issue on image content variation that should be considered in the development of blur-specific NR-IQA methods.

References

1. Bahrami K, Kot AC (2014) A fast approach for no-reference image sharpness assessment based on maximum local variation. IEEE Sig Process Lett 21(6):751–755
2. Bahrami K, Kot AC (2016) Efficient image sharpness assessment based on content aware total variation. IEEE Trans Multimed 18(8):1568–1578
3. Bong DBL, Khoo BE (2014) Blind image blur assessment by using valid reblur range and histogram shape difference. Sig Process Image Commun 29(6):699–710
4. Ciancio A, da Costa ALNT, da Silva EAB, Said A, Samadani R, Obrador P (2011) No-reference blur assessment of digital pictures based on multifeature classifiers. IEEE Trans Image Process 20(1):64–75
5. Feichtenhofer C, Fassold H, Schallauer P (2013) A perceptual image sharpness metric based on local edge gradient analysis. IEEE Sig Process Lett 20(4):379–382
6. Ferzli R, Karam LJ (2009) A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB). IEEE Trans Image Process 18(4):717–728
7. Gu K, Zhai G, Lin W, Yang X, Zhang W (2015) No-reference image sharpness assessment in autoregressive parameter space. IEEE Trans Image Process 24(10):3218–3231
8. Guan J, Zhang W, Gu J, Ren H (2015) No-reference blur assessment based on edge modeling. J Vis Commun Image Represent 29:1–7
9. Hassen R, Wang Z, Salama MMA (2013) Image sharpness assessment based on local phase coherence. IEEE Trans Image Process 22(7):2798–2810
10. Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1):489–501
11. Kang L, Ye P, Li Y, Doermann D (2014) Convolutional neural networks for no-reference image quality assessment. In: IEEE conference on computer vision and pattern recognition, pp 1733–1740. IEEE
12. Leclaire A, Moisan L (2015) No-reference image quality assessment and blind deblurring with sharpness metrics exploiting Fourier phase information. J Math Imaging Vis 52(1):145–172
13. Li L, Lin W, Wang X, Yang G, Bahrami K, Kot AC (2016) No-reference image blur assessment based on discrete orthogonal moments. IEEE Trans Cybern 46(1):39–50

14. Li L, Wu D, Wu J, Li H, Lin W, Kot AC (2016) Image sharpness assessment by sparse representation. IEEE Trans Multimed 18(6):1085–1097
15. Li L, Xia W, Lin W, Fang Y, Wang S (2017) No-reference and robust image sharpness evaluation based on multi-scale spatial and spectral features. IEEE Trans Multimed 19(5):1030–1040
16. Marziliano P, Dufaux F, Winkler S, Ebrahimi T (2004) Perceptual blur and ringing metrics: application to JPEG2000. Sig Process Image Commun 19(2):163–172
17. Mittal A, Moorthy AK, Bovik AC (2012) No-reference image quality assessment in the spatial domain. IEEE Trans Image Process 21(12):4695–4708
18. Mukundan R, Ong S, Lee PA (2001) Image analysis by Tchebichef moments. IEEE Trans Image Process 10(9):1357–1364
19. Narvekar ND, Karam LJ (2011) A no-reference image blur metric based on the cumulative probability of blur detection (CPBD). IEEE Trans Image Process 20(9):2678–2683
20. Ponomarenko N, Jin L, Ieremeiev O, Lukin V, Egiazarian K, Astola J, Vozel B, Chehdi K, Carli M, Battisti F, Kuo CCJ (2015) Image database TID2013: peculiarities, results and perspectives. Sig Process Image Commun 30:57–77
21. Sang Q, Qi H, Wu X, Li C, Bovik AC (2014) No-reference image blur index based on singular value curve. J Vis Commun Image Represent 25(7):1625–1630
22. Sheikh HR, Sabir MF, Bovik AC (2006) A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Trans Image Process 15(11):3440–3451
23. VQEG (2000) Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment. Video Quality Experts Group. http://vqeg.org/
24. Vu PV, Chandler DM (2012) A fast wavelet-based algorithm for global and local image sharpness estimation. IEEE Signal Process Lett 19(7):423–426
25. Vu CT, Phan TD, Chandler DM (2012) S_3: a spectral and spatial measure of local perceived sharpness in natural images. IEEE Trans Image Process 21(3):934–945
26. Wang Z, Simoncelli EP (2003) Local phase coherence and the perception of blur. In: Advances in neural information processing systems, pp 1435–1442
27. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612
28. Wang S, Deng C, Zhao B, Huang GB, Wang B (2016) Gradient-based no-reference image blur assessment using extreme learning machine. Neurocomputing 174:310–321
29. Yu S, Wu S, Wang L, Jiang F, Xie Y, Li L (2017) A shallow convolutional neural network for blind image sharpness assessment. PloS One 12(5):e0176632
30. Zhai G, Wu X, Yang X, Lin W, Zhang W (2012) A psychovisual quality metric in free-energy principle. IEEE Trans Image Process 21(1):41–52